

А.А. БОРОВКОВ

ТЕОРИЯ ВЕРОЯТНОСТЕЙ

Издание третье,
существенно переработанное
и дополненное



1999

Издательство «Эдиториал УРСС»
Москва

Издательство института математики
Новосибирск



Настоящее издание осуществлено при финансовой поддержке Российского фонда фундаментальных исследований (проект № 98-01-14170)

Боровков Александр Алексеевич

ТЕОРИЯ ВЕРОЯТНОСТЕЙ

М.: Эдиториал УРСС, 1999. — 472 с.

ISBN 5-901006-66-6

Книга охватывает широкий круг вопросов, начиная с оснований теории вероятностей и заканчивая основными элементами теории случайных процессов. Сюда входят: достаточно полный аппарат современной теории вероятностей; разного рода предельные законы для сумм независимых случайных величин; теоремы о поведении траекторий, порожденных этими суммами, включая относящиеся сюда так называемые факторизационные тождества; элементы теории восстановления и различные ее приложения; цепи Маркова и эргодические теоремы для них; элементы теории информации; теория мартингалов и стохастически рекурсивных последовательностей; основы теории случайных процессов; теоремы об основных свойствах винеровских и пуассоновских процессов; функциональные предельные теоремы; элементы теории марковских, стационарных и гауссовских процессов и др.

Группа подготовки издания:

Директор — Доминго Марин Рикой

Заместители директора — Наталья Финогенова, Ирина Макеева

Компьютерный дизайн — Виктор Романов, Василий Подобед

Верстка — Наталия Бекетова

Редакционно-корректорские работы — Виталий Волков, Лариса Кирдяшкина

Обработка графики — Елена Ефремова

Обработка текста — Евгений Макаров, Анна Тюрина, Андрей Стулов

Техническое обеспечение — Елена Логвинова, Марина Круцко, Наталья Аринчева

Издательство «Эдиториал УРСС». 113208, г. Москва, ул. Чертановская, д. 2/11, ком. прав.

Лицензия ЛР № 064418 от 14.05.96 г. Подписано к печати 21.04.99 г.

Формат 70 × 100/16. Тираж 1000 экз. Авт. л. 38. Печ. л. 29,5. Зак. № 587.

Отпечатано в АОТ «Политех-4». 129110, г. Москва, ул. Б. Переславская, 46.



ISBN 5-901006-66-6

© Эдиториал УРСС, 1999

Содержание

Предисловие	7
Введение	10
Глава 1. Дискретное пространство элементарных событий	13
§ 1. Вероятностное пространство	13
§ 2. Классическая схема	15
§ 3. Схема Бернулли	18
§ 4. Вероятность объединения событий. Примеры	20
Глава 2. Произвольное пространство элементарных событий	23
§ 1. Аксиомы теории вероятностей. Вероятностное пространство	23
§ 2. Свойства вероятности	29
§ 3. Условная вероятность. Независимость событий и испытаний	30
§ 4. Формула полной вероятности и формула Байеса	32
Глава 3. Случайные величины и функции распределения	37
§ 1. Определения и примеры	37
§ 2. Свойства функций распределения и примеры	39
§ 3. Многомерные случайные величины	47
§ 4. Независимость случайных величин и классов событий	50
§ 5*. О бесконечных последовательностях случайных величин	57
§ 6. Интегралы	57
Глава 4. Числовые характеристики случайных величин	64
§ 1. Математическое ожидание	64
§ 2. Условные функции распределения и условные математические ожидания	68
§ 3. Математические ожидания функций независимых случайных величин	72
§ 4. Математическое ожидание случайных величин, не зависящих от будущего	72
§ 5. Дисперсия	76
§ 6. Коэффициент корреляции и другие числовые характеристики	78
§ 7. Неравенства	80
§ 8. Обобщение понятия условного математического ожидания	82
§ 9. Условные распределения	89
Глава 5. Последовательность независимых испытаний с двумя исходами	96
§ 1. Законы больших чисел	96
§ 2. Локальная предельная теорема	97
§ 3. Теорема Муавра—Лапласа и ее уточнения	102
§ 4. Теорема Пуассона и ее уточнения	104
§ 5. Неравенства для вероятностей больших отклонений в схеме Бернулли	111
Глава 6. О сходимости случайных величин и распределений	114
§ 1. Сходимость случайных величин	114
§ 2. Сходимость распределений	121
§ 3. Условия слабой сходимости	126

Глава 7. Характеристические функции	130
§ 1. Определение и свойства характеристических функций	130
§ 2. Формула обращения	135
§ 3. Теорема непрерывности (сходимости)	137
§ 4*. Другой подход к доказательству теорем сходимости к известному распределению	138
§ 5. Применение характеристических функций в теореме Пуассона	141
§ 6. Характеристические функции многомерных распределений. Многомерное нормальное распределение	143
§ 7. Другие применения х. ф. Свойства гамма-распределения	146
§ 8. Производящие функции. Применение к изучению ветвящегося процесса. Задача о вырождении	151
Глава 8. Последовательности независимых случайных величин. Предельные теоремы .	155
§ 1. Закон больших чисел	155
§ 2. Центральная предельная теорема для одинаково распределенных случайных величин	156
§ 3. Закон больших чисел для произвольных независимых случайных величин .	157
§ 4. Центральная предельная теорема для сумм произвольных независимых случайных величин	164
§ 5. Другой подход к доказательству предельных теорем. Оценки погрешности .	171
§ 6. Локальная предельная теорема	174
§ 7. Закон больших чисел и центральная предельная теорема в многомерном случае	177
§ 8*. Вероятности больших отклонений	179
§ 9*. Сходимость к другим устойчивым законам	187
Глава 9. Элементы теории восстановления	195
§ 1. Процессы восстановления, функции восстановления	195
§ 2. Основная теорема восстановления в решетчатом случае	199
§ 3. Экссесс и дефект случайного блуждания. Предельное распределение в решетчатом случае	204
§ 4. Теорема восстановления и предельное распределение эксцесса и дефекта в нерешетчатом случае	206
§ 5. Закон больших чисел и центральная предельная теорема для процесса восстановления	210
Глава 10. Последовательности независимых случайных величин. Свойства траектории $(0, S_1, S_2, \dots)$ в целом	213
§ 1. Законы нуля и единицы. Верхние и нижние функции	213
§ 2. Сходимость рядов независимых случайных величин	217
§ 3. Усиленный закон больших чисел	219
§ 4. Усиленный закон больших чисел для произвольных независимых слагаемых	223
Глава 11. Факторизационные тождества	225
§ 1. Факторизационные тождества и их первые следствия	225
§ 2. Факторизационные тождества. Свойства траектории $(0, S_1, S_2, \dots)$	230
§ 3. Распределение $S = \max_{k \geq 0}(0, \zeta) = \max_{k \geq 0} S_k$	233
§ 4. Системы обслуживания	234
§ 5. Факторизационные тождества для распределений, связанных с показательной функцией	235
§ 6. Симметричные непрерывно распределенные случайные величины	237
§ 7. Тождество Поллачека—Спитцера	238
§ 8*. Явные формулы для дискретных блужданий, непрерывных сверху	240

Глава 12. Последовательности зависимых испытаний. Цепи Маркова	245
§ 1. Счетные цепи Маркова. Определения и примеры. Классификация состояний	245
§ 2. Необходимые и достаточные условия возвратности состояний. Теорема об однотипности состояний неразложимой цепи, структура цепи в периодическом случае	249
§ 3. Теоремы о случайных блужданиях по решетке	252
§ 4. Эргодические теоремы	257
§ 5*. Поведение переходных вероятностей для разложимых цепей	264
§ 6. Цепи Маркова с произвольным множеством состояний. Эргодичность цепей, имеющих положительный атом	265
§ 7*. Эргодичность харрисовых цепей Маркова	272
Глава 13. Информация и энтропия	284
§ 1. Определения, свойства информации и энтропии	284
§ 2. Энтропия конечной цепи Маркова. Теорема об асимптотическом поведении информации длинного сообщения, ее приложения	288
Глава 14. Мартингалы	292
§ 1. Определения, простейшие свойства, примеры	292
§ 2. О сохранении свойства быть мартингалом при замене времени на случайное. Тождество Вальда	296
§ 3. Неравенства	308
§ 4. Теоремы сходимости	312
§ 5. Ограниченность моментов стохастических последовательностей	316
Глава 15. Стационарные (в узком смысле) последовательности	321
§ 1. Основные понятия	321
§ 2. Свойства эргодичности (метрической транзитивности), перемешивания и слабой зависимости	325
§ 3. Эргодическая теорема	328
Глава 16. Стохастически рекурсивные последовательности	332
§ 1. Основные понятия	332
§ 2. Эргодичность при наличии обновляющих событий. Условия ограниченности	333
§ 3. Условия эргодичности, связанные с монотонностью f	339
§ 4. Условия эргодичности для сжимающих в среднем преобразований, удовлетворяющих условию Липшица	340
Глава 17. Случайные процессы с непрерывным временем	347
§ 1. Общие определения	347
§ 2. Условия регулярности процессов	351
Глава 18. Процессы с независимыми приращениями	357
§ 1. Общие свойства	357
§ 2. Винеровские процессы, свойства траекторий	359
§ 3. Законы повторного логарифма	360
§ 4. Пуассоновские процессы	364
§ 5. Описание распределений всего класса процессов с независимыми приращениями	367
Глава 19. Функциональные предельные теоремы	371
§ 1. Сходимость к винеровскому процессу (принцип инвариантности)	371
§ 2. Закон повторного логарифма	378
§ 3. Сходимость к пуассоновскому процессу	381

Глава 20. Марковские процессы	385
§ 1. Определения и общие свойства	385
§ 2. Марковские процессы со счетным множеством состояний. Примеры	388
§ 3. Ветвящиеся процессы	394
§ 4. Полумарковские и регенерирующие процессы	397
§ 5. Диффузионные процессы	404
Глава 21. Процессы с конечными моментами второго порядка, гауссовские процессы	410
§ 1. Процессы с конечными моментами второго порядка	410
§ 2. Гауссовские процессы	413
§ 3. Задача о прогнозе	414
Приложения	417
Приложение 1. Теорема о продолжении вероятностной меры	419
Приложение 2. Теорема Колмогорова о согласованных распределениях	424
Приложение 3. Интегрирование	426
§ 1. Пространство с мерой	426
§ 2. Интеграл по вероятностной мере	427
§ 3. Дальнейшие свойства интегралов	430
§ 4. Интеграл по произвольной мере	434
§ 5. Теорема Лебега о разложении и теорема Радона—Никодима	437
§ 6. Слабая сходимость и сходимость по вариации распределений в произвольных пространствах	441
Приложение 4. Теоремы Хелли и Арцела—Асколи	447
Приложение 5. Доказательство теоремы Берри—Эссена	449
Приложение 6. Теоремы восстановления	453
Приложение 7. Таблицы	460
Список литературы	465
Предметный указатель	467

Предисловие

Настоящая книга написана на основе книги «Теория вероятностей», вышедшей в 1986 году в издательстве «Наука» (г. Москва). Многие разделы подверглись значительной переработке, появились новые главы. Автором руководило при этом стремление к более совершенному и связному изложению, к более простым и наглядным доказательствам. Изданию 1986 года предшествовали еще два издания (1972 и 1976 годов). Первое из них появилось как обработка курсов лекций, которые читались автором на механико-математическом факультете Новосибирского государственного университета. Каждое новое издание учитывало замечания читателей и пополнялось новыми разделами, которые делали изложение более цельным и полным.

Предполагается, что читатели в дополнение к традиционному курсу математического анализа знакомы также с элементами теории меры и, в частности, с понятием интеграла по мере на абстрактном пространстве и с его простейшими свойствами. Однако отсутствие этих дополнительных знаний не мешает успешному освоению книги, если только читатель будет готов некоторые утверждения принять не в их максимально общей форме. Есть и другая возможность. Читатель полностью избежит затруднений, если он при чтении соответствующих разделов ознакомится с краткими приложениями, которые помещены в конце книги и в которых получены все необходимые результаты.

Первые 10 глав книги посвящены основаниям теории вероятностей (включая основные предельные теоремы для сумм растущего числа случайных величин), и их лучше читать подряд. Дальнейшие главы излагают более специальные разделы теории вероятностей, среди которых можно выделить 2 блока: случайные процессы с дискретным временем (случайные последовательности: главы 12, 14–16) и случайные процессы с непрерывным временем (главы 17–21).

В книге есть также главы, которые стоят несколько в стороне от основного русла изложения, намеченного выше. К ним относится глава 11 «Факторизационные тождества», которая не только предлагает ряд очень полезных вероятностных результатов, но и обнаруживает интересные связи задач о случайных блужданиях при наличии границ с граничными задачами теории функций комплексного переменного. Несколько в стороне стоят также главы 13 «Информация и энтропия» и 19 «Функциональные предельные теоремы». Проблемы первой из них тесно связаны с теорией вероятностей, но они появляются в книгах по этой дисциплине крайне редко. Вторая излагает предельные теоремы о сходимости процессов, порожденных последовательными суммами случайных величин, к винеровскому и пуассоновскому процессам; там же устанавливается в качестве следствия закон повторного логарифма.

Книга содержит много методологических усовершенствований. Некоторые разделы публикуются в учебной монографической литературе впервые (такие, например, как глава 16 о стохастически рекурсивных последовательностях, играющих важную роль в приложениях).

Книга может служить основой учебных курсов для студентов, обладающих достаточно хорошей математической подготовкой (скажем, в рамках двух лет математических специальностей университетов), а также для аспирантов. Приведем перечень разделов, которые могут составить (вариантов тут может быть много) содержание семестрового или полуторасеместрового курса теории вероятностей. Вот один из вариантов: главы 1, 2; § 1–4 главы 3; § 1–6 главы 4 (частично); § 2, 4 главы 5 (частично); § 1–3 главы 6 (частично); § 1, 2, 4–6 главы 7; § 1, 2, 4 главы 8 (§ 4 частично); § 1, 3 главы 10; основные результаты из главы 12.

Более полное изложение основ теории вероятностей можно найти в [16, 22, 25], основ теории случайных процессов — в [7, 16].

В процессе работы над изданиями этой книги я получал советы и помощь от многих моих коллег и друзей. Я благодарен Ю. В. Прохорову, В. В. Петрову и Б. А. Рогозину за многие полезные советы, которые помогли улучшить первый вариант книги. Я глубоко признателен А. Н. Колмогорову, чьи замечания и ценные рекомендации, особенно методологического характера, способствовали усовершенствованию второго варианта книги. В связи со вторым и третьим вариантами я вновь обязан В. В. Петрову, сообщившему мне свои замечания, и П. Франкену, с которым я имел много полезных обсуждений при переводе книги на немецкий язык.

В заключение я хотел бы поблагодарить В. В. Юринского, А. И. Саханенко, К. А. Боровкова и других моих коллег, также сообщивших мне свои замечания по рукописи. Мне хотелось бы поблагодарить здесь всех, кто тем или иным образом содействовал подготовке и улучшению книги.

А. А. Боровков

Вниманию читателя!

Значок ■ в тексте означает окончание доказательства. Нумерация формул, лемм, теорем, параграфов в каждой главе самостоятельная. При ссылках на материал из других глав используется двойная нумерация. Например, ссылка (9.11) означает, что имеется в виду формула (11) главы 9; теорема 3.4 означает теорему 4 главы 3; § 6.2 означает § 2 главы 6. Параграфы, отмеченные звездочкой, при первом чтении книги можно опустить.

Введение

1. Возникновение теории вероятностей относят обычно к XVII веку и связывают с комбинаторными задачами азартных игр. Азартные игры трудно считать серьезным занятием. Но именно они привели к задачам, которые не укладывались в рамки существовавших тогда математических моделей и стимулировали тем самым введение новых понятий, подходов и идей. Эти новые элементы можно встретить уже у Ферма, Паскаля, Гюйгенса и, в более развитой форме, чуть позже, у Якоба Бернулли, Лапласа, Гаусса и др. Имена перечисленных ученых, безусловно, украшают родословную теории вероятностей, в какой-то мере связанную, как мы видели, с пороками общества. Впрочем, как выяснилось, иногда именно это обстоятельство может сообщать ей в глазах читателей некоторую дополнительную привлекательность.

Первым руководством по теории вероятностей был трактат Гюйгенса «О расчетах в азартной игре», вышедший в 1657 году. Предмет этого трактата тот же, что и в работах Ферма и Паскаля: игральные кости и карточные игры (см. задачи в рамках § 2 главы 1). Словно предвидя дальнейшее развитие событий, Гюйгенс писал: «Я полагаю, что при внимательном изучении предмета читатель заметит, что имеет дело не только с игрой, но что здесь закладываются основы очень интересной и глубокой теории». Трактат Гюйгенса, известный также тем, что в нем впервые вводится понятие математического ожидания, был включен позднее Якобом Бернулли в знаменитую книгу «Искусства предположений», опубликованную уже после его смерти в 1713 году¹⁾. С этой книгой связаны понятие схемы Бернулли (см. § 3 главы 1) и доказательство первой предельной теоремы теории вероятностей — закона больших чисел. Доказательство Бернулли хотя и громоздко (ср. с § 1 главы 5), но математически безупречно.

В конце прошлого и в начале этого века стали возникать более серьезные проблемы, порожденные нуждами естествознания и приведшие к развитию большого, в значительной мере самостоятельного раздела математики, именуемого сегодня теорией вероятностей. Эта область знаний вплоть до настоящего времени находится в состоянии интенсивного развития. Своей стройностью, современной формой и многими своими достижениями теория вероятностей во многом обязана трудам наших замечательных соотечественников П. Л. Чебышева, А. А. Маркова, А. Н. Колмогорова и др.

То обстоятельство, что увеличение наших знаний о природе предъявляет все новые требования к теории вероятностей, на первый взгляд представляется парадоксом. Ведь основной объект теории вероятностей, как читатель, по-видимому, уже знает, есть случайность или неопределенность, как правило, связанная с незнанием. Именно так обстоит дело в классическом примере с подбрасыванием монеты, где трудно учесть все факторы, влияющие на положение монеты после падения.

Однако отмеченный парадокс является лишь видимым. На самом деле точных,

¹⁾ В переводе на русский язык книга была впервые издана в 1913 году. В 1986 году вышло ее новое издание на русском языке. (Бернулли Я. О законе больших чисел. М.: Наука.)

детерминированных количественных законов в природе почти не существует. Скажем, классический пример таких законов — закон о зависимости давления газа от его температуры — есть на самом деле результат вероятностного характера о числе соударений частиц о стенки сосуда и их скоростях. Просто в области обычных температур и давлений случайные отклонения, которые тут имеют место, с большой вероятностью очень малы и не регистрируются нашими приборами. Иначе обстоит дело при изучении более редких потоков частиц, скажем, космического излучения, хотя качественной разницы между этими двумя примерами нет.

Можно пойти и в несколько ином направлении и указать на принцип неопределенности, в силу которого для любой пары физических характеристик, связанных этим принципом, фиксация одной из них делает невозможным точное определение другой. Тут уже случайность появляется не как следствие недостаточности наших знаний, а как принципиальное явление и отражение природы вещей. Например, время жизни радиоактивного ядра случайно по существу, и эта случайность не может быть устранена увеличением наших познаний.

Таким образом, неопределенность стояла в начале процесса познания, она будет стоять и на всем пути его. Эти замечания носят, конечно, общий характер. Но вопрос о том, когда следует применять методы теории вероятностей и когда нет, по-видимому, всегда будет определяться соотношением между степенью точности, с которой мы хотим изучать данное явление, и сведениями о его природе, которыми мы располагаем.

2. Почти во всех областях человеческой деятельности существуют ситуации, когда те или иные эксперименты или наблюдения могут быть повторены большое число раз в одинаковых условиях. Теорию вероятностей интересуют те эксперименты, результат которых, выраженный каким-либо образом, может меняться от опыта к опыту. События, относящиеся к результату эксперимента, которые при этом могут происходить или не происходить, называют обычно *случайными событиями*.

Мы можем, например, подбрасывать монету. Эксперимент имеет всего два исхода: монета падает либо гербом вверх, либо решеткой. При этом до получения результата эксперимента нельзя сказать, какой именно исход осуществится. Как уже отмечалось, это происходит оттого, что мы практически не в состоянии учесть все факторы, влияющие на положение монеты в момент падения. Примерно то же самое будет происходить, если вы при каждой лотерее будете покупать один билет и будете пытаться предугадать, выиграет он или нет. Или, наблюдая за работой достаточно сложного механизма, будете пытаться заранее определить, выйдет он из строя до назначенного срока или после. В таких ситуациях при рассмотрении результатов *отдельных* экспериментов бывает очень трудно обнаружить какие-либо закономерности. И, стало быть, здесь мало оснований строить какую-либо теорию.

Однако, если обратить внимание на последовательность большого числа такого рода одинаковых экспериментов, то обнаружится интересное явление. Если индивидуальные результаты опытов ведут себя «неправильно», то средние результаты обнаруживают устойчивость. Будем, например, повторять наш эксперимент с подбрасыванием монеты и обозначим через n_{Γ} число выпадений герба после первых n испытаний. Построим следующий график: на оси абсцисс отложим число проведенных экспериментов, а на оси ординат — отношение n_{Γ}/n (рис. 1; график построен для последовательности исходов *ггргггггггггг*, где *г* — герб, *р* — решетка).

Мы заметим, что ломаная, соединяющая точки $(n, n_{\Gamma}/n)$, с ростом n очень быстро прижимается к прямой $n_{\Gamma}/n = 1/2$. Чтобы проверить это обстоятельство,

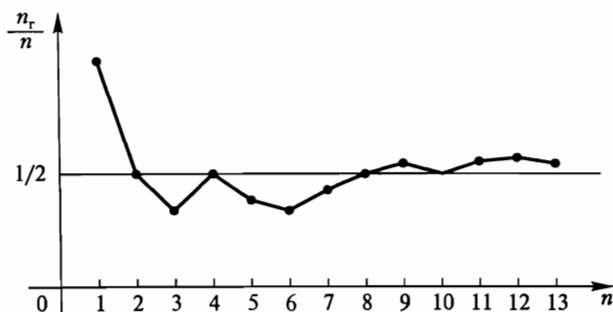


Рис. 1.

Бюффон²⁾ в XVIII веке провел 4040 подбрасываний монеты. Из них герб выпал 2048 раз, так что частота выпадения герба n_{Γ}/n оказалась равной 0,508. Пирсон провел 24000 бросаний монеты; герб выпал 12012 раз, $n_{\Gamma}/n = 0,5005$.

Оказывается, это явление имеет общий характер: *частота осуществления какого-либо исхода в последовательности повторяемых в одинаковых условиях экспериментов приближается к некоторому числу $p \in [0, 1]$ при росте числа экспериментов.* Это объективный закон природы, который лежит в основе теории вероятностей.

Естественно было бы принять за вероятность некоторого исхода в эксперименте как раз то число p , к которому приближается частота появления этого исхода. Но такое определение вероятности (его связывают обычно с именем Мизеса) оказывается неудобным. Прежде всего, последовательность частот $\{n_{\Gamma}/n\}$ появления некоторого исхода при проведении одной серии экспериментов будет, как правило, отличаться от последовательности частот появления того же исхода при другой серии экспериментов. Кроме того, на самом деле мы будем иметь не бесконечную последовательность частот, а только конечное число ее элементов. Получить всю последовательность невозможно.

По-видимому, вероятность нужно определять иначе, но так, чтобы имело место отмеченное желательное свойство частоты: в каком-то смысле отношение n_{Γ}/n должно приближаться при увеличении n к вероятности рассматриваемого события. Отсюда можно извлечь, в частности, что вероятностями могут быть только числа $p \in [0, 1]$, при этом вероятность достоверного события следует считать равной 1.

Мы определим вероятность сначала в несколько упрощенной ситуации — в так называемом *дискретном случае*.

²⁾ Эти данные заимствованы из [8].

Глава 1

Дискретное пространство элементарных событий

§ 1. Вероятностное пространство

Для математического описания экспериментов со случайными исходами нам потребуется прежде всего понятие *пространства элементарных событий* (или *исходов*), соответствующего рассматриваемому эксперименту. Таким пространством будем называть любое множество Ω взаимоисключающих исходов эксперимента такое, что каждый интересующий нас результат эксперимента может быть однозначно описан с помощью элементов этого множества.

В самых простых экспериментах мы обычно имеем дело с *конечными* пространствами элементарных событий. В рассмотренном выше примере с подбрасыванием монеты Ω состоит из двух элементов — герба и решетки. В опыте с подбрасыванием шестигранной игральной кости Ω также конечно и состоит из 6 элементов. Но даже с подбрасыванием монеты (или игральной кости) можно связать такие эксперименты, в которых будет невозможно обойтись конечными пространствами элементарных исходов. Рассмотрим, например, такой эксперимент: монета бросается до тех пор, пока впервые не выпадет герб, после этого эксперимент прекращают. Если p будет означать выпадение решетки, а g — выпадение герба, то «элементарный исход» этого эксперимента можно представить в виде последовательности $(pp \dots pg)$. Таких последовательностей бесконечно много, и все они различны, так что для однозначного описания всех исходов эксперимента нельзя обойтись никаким конечным пространством.

Рассмотрим конечные или счетные пространства элементарных событий Ω . Это так называемые *дискретные* пространства. Элементы пространства Ω мы будем обозначать буквой ω и называть их *элементарными событиями* (или *элементарными исходами*).

Само понятие пространства элементарных событий математически является *неопределяемым* — оно исходно, так же как понятие точки в геометрии. Конкретная природа Ω нас, как правило, интересовать не будет.

Событием мы будем называть любое подмножество $A \subseteq \Omega$ элементов из Ω (событие A произошло, если произошло какое-либо из элементарных событий $\omega \in A$).

Суммой двух событий A и B называется событие $A + B$ ($A \cup B$), состоящее из элементарных событий, принадлежащих хотя бы одному из событий A или B . *Произведением* AB ($A \cap B$) называется событие, состоящее из элементарных событий, принадлежащих одновременно A и B . *Разность* событий A и B соответствует множеству $A - B$, состоящему из элементов A , не принадлежащих B .

Ω называется *достоверным* событием. Пустое множество \emptyset называется *невозможным* событием. $\bar{A} = \Omega - A$ называется *дополнительным* событием к A . События A и B *несовместны*, если $AB = \emptyset$.

Пусть, например, эксперимент состоит в двукратном подбрасывании игральной кости. В качестве пространства элементарных событий здесь можно рассмотреть множество, состоящее из 36 элементов (i, j) , где i и j пробегает значения от 1 до 6 и означают количество выпавших очков соответственно при первом и втором бросаниях. События $A = \{i + j \leq 3\}$ и $B = \{j = 6\}$ несовместны. Произведением событий A и $C = \{j \text{ четно}\}$ является событие $(1, 2)$. Отметим, что если бы нас интересовали события, относящиеся лишь к первому бросанию, то мы могли бы рассматривать и более бедное пространство элементарных событий, состоящее из 6 элементов $i = 1, 2, \dots, 6$.

Говорят, что заданы *вероятности элементарных событий*, если на Ω задана неотрицательная числовая функция P такая, что $\sum_{\omega \in \Omega} P(\omega) = 1$ (говорят также, что *функция P задает на Ω распределение вероятностей*).

Вероятностью события A называется число

$$P(A) = \sum_{\omega \in A} P(\omega).$$

Это определение корректно, так как ряд в правой части абсолютно сходится.

Отметим здесь, что конкретные числовые значения функции P нас также интересовать не будут — это лишь вопрос практической ценности той или иной модели. Ясно, например, что в случае симметричной игральной кости для исходов $1, 2, \dots, 6$ надо положить $P(1) = P(2) = \dots = P(6) = 1/6$; для симметричной же монеты надо выбрать значения $P(z) = P(p) = 1/2$, а не какие-нибудь другие. В эксперименте с подбрасыванием монеты до первого появления герба следует положить $P(z) = 1/2$, $P(pz) = 1/2^2$, $P(ppz) = 1/2^3, \dots$. Поскольку $\sum_{n=1}^{\infty} 2^{-n} = 1$, то функция P , заданная таким образом на исходах вида $(p \dots pz)$, будет определять распределение вероятностей на Ω . Чтобы подсчитать, например, вероятность того, что эксперимент закончится на четном шаге (т.е. вероятность события, составленного из исходов $(pz), (pppz), \dots$), надо рассмотреть сумму соответствующих вероятностей, равную $\sum_{n=1}^{\infty} 2^{-2n} = \frac{1}{4} \cdot \frac{4}{3} = \frac{1}{3}$.

В упоминавшихся во введении экспериментах, где происходило угадывание, когда выйдет из строя механизм — до назначенного срока (событие A) или после — основанием для количественных оценок вероятности $P(A)$ обычно могут служить только результаты самих экспериментов. Методы оценок неизвестных вероятностей по результатам наблюдений исследуются математической статистикой, содержание которой мы поясним несколько позже на примере одной из задач этой главы.

Отметим далее, что далеко не для всех экспериментов можно построить модели с дискретным пространством элементарных событий. Допустим, например, что измеряется энергия частиц, возможные значения которой заполняют отрезок $[0, V]$, $V > 0$. Ведь множество точек этого отрезка (т.е. множество элементарных событий) континуально. Или если, например, результатом эксперимента является электрокардиограмма, снятая с больного. Здесь результатом эксперимента является элемент некоторого функционального пространства. В этих случаях требуются более общие схемы.

Из приведенных выше определений, пользуясь абсолютной сходимостью рядов $\sum_{\omega \in A} P(\omega)$, нетрудно получить следующие свойства вероятности:

$$1) \quad P(\emptyset) = 0, \quad P(\Omega) = 1.$$

$$2) \quad P(A + B) = \sum_{\omega \in A \cup B} P(\omega) = \sum_{\omega \in A} P(\omega) + \sum_{\omega \in B} P(\omega) - \sum_{\omega \in A \cap B} P(\omega) = \\ = P(A) + P(B) - P(AB).$$

$$3) \quad P(\bar{A}) = 1 - P(A).$$

Отсюда следует, в частности, что для *непересекающихся* (несовместных) событий A и B

$$P(A + B) = P(A) + P(B).$$

Это свойство *аддитивности вероятности* сохранится и для совокупности из произвольного числа непересекающихся событий A_1, A_2, \dots : если $A_i A_j = \emptyset$ при $i \neq j$, то

$$P\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} P(A_k).$$

Это следует из равенства

$$P\left(\bigcup_{k=1}^n A_k\right) = \sum_{k=1}^n P(A_k)$$

и того, что $P\left(\bigcup_{k=n+1}^{\infty} A_k\right) \rightarrow 0$ при $n \rightarrow \infty$.

Для произвольных A и B $P(A + B) \leq P(A) + P(B)$. Это неравенство справедливо, очевидно, и для суммы произвольного числа событий:

$$P\left(\bigcup_{k=1}^{\infty} A_k\right) \leq \sum_{k=1}^{\infty} P(A_k).$$

Рассмотрим теперь несколько важных частных случаев.

§ 2. Классическая схема

Пусть Ω состоит из n элементов и все исходы равновероятны, т. е. $P(\omega) = 1/n$ для любого $\omega \in \Omega$. В этом случае вероятность любого события A определяется формулой

$$P(A) = (\text{число элементов } A)/n.$$

Это так называемое *классическое определение вероятности* (используется также термин *равномерное дискретное распределение*).

Пусть задано множество $\{a_1, a_2, \dots, a_n\}$, которое мы будем называть *генеральной совокупностью*. Выборкой объема k из генеральной совокупности называется упорядоченная последовательность $(a_{j_1}, a_{j_2}, \dots, a_{j_k})$. Эту последовательность можно образовывать следующим образом: первый элемент a_{j_1} выберем из всей генеральной совокупности. Следующий элемент a_{j_2} выберем из генеральной совокупности без элемента a_{j_1} ; элемент a_{j_3} выбираем из генеральной совокупности без элементов a_{j_1}

и a_{j_2} и т.д. Полученные таким образом выборки называются *выборками без возвращения*. Ясно, что в этом случае должно быть $k \leq n$. Число таких выборок объема k совпадает с числом размещений из n по k :

$$(n)_k = n(n-1)(n-2) \cdots (n-k+1).$$

Действительно, в соответствии с процессом выбора на первом месте может стоять любой из n элементов генеральной совокупности, на втором — любой из $(n-1)$ элементов и т.д. Более формальное доказательство можно провести индукцией по числу k .

Припишем каждой выборке без возвращения вероятность $1/(n)_k$. Такая выборка будет называться *случайной*. Это есть, очевидно, классическая схема.

Подсчитаем вероятность того, что $a_{j_1} = a_1$, $a_{j_2} = a_2$. Так как остальные $k-2$ места могут быть заняты любым из оставшихся $n-2$ элементов генеральной совокупности, то число выборок без возвращения с элементами a_1 и a_2 на первых местах равно $(n-2)_{k-2}$. Стало быть, вероятность указанного события равна

$$\frac{(n-2)_{k-2}}{(n)_k} = \frac{1}{n(n-1)}.$$

Выборку без возвращения наглядно можно представить как последовательный выбор из совокупности n перенумерованных шаров, находящихся в урне. Вытянутые шары в урну не возвращаются.

Но выборку можно образовывать и другим способом. Из урны вынимается шар и запоминается. Затем он возвращается в урну и снова производится выбор шара; вытянутый шар опять запоминается и возвращается в урну и т.д. Полученная таким способом выборка называется *выборкой с возвращением*. Таких выборов делается k , поэтому всего выборок с возвращением будет n^k . Если приписать каждой выборке вероятность $1/n^k$, то опять будет иметь место классическая схема.

Подсчитаем, например, вероятность того, что в выборке с возвращением объема $k \leq n$ все элементы будут разные. Выборок с неповторяющимися элементами будет столько же, сколько выборок без возвращения, т.е. $(n)_k$. Значит, искомая вероятность равна $(n)_k/n^k$.

Вернемся теперь к выборке без возвращения из генеральной совокупности $\{a_1, a_2, \dots, a_n\}$. Нас будет интересовать число выборок объема $k \leq n$, отличающихся только составом. Число выборок без возвращения объема k , имеющих одинаковый состав и различающихся только порядком элементов, будет $k!$. Поэтому число выборок, различающихся составом, будет равно

$$\frac{(n)_k}{k!} = \binom{n}{k}.$$

Это — *число сочетаний из n по k* при $0 \leq k \leq n$ ¹⁾. Если исходная выборка случайна, то получим опять классическую вероятностную схему, так как вероятность каждой новой выборки равна

$$\frac{k!}{(n)_k} = \frac{1}{\binom{n}{k}}.$$

¹⁾ В дальнейшем для значений $k < 0$ и $k > n$ положим $\binom{n}{k} = 0$.

Пусть в урне n шаров, из них n_1 черных и $n - n_1$ белых. Производится выборка без возвращения объема k . Какова вероятность того, что в выборке будет k_1 черных шаров? Всего выборов, различающихся составом, как было показано, $\binom{n}{k}$. Черных k_1 шаров из всех n_1 черных шаров можно выбрать $\binom{n_1}{k_1}$ способами. Остальные $k - k_1$ белых шаров из всех $n - n_1$ белых шаров можно выбрать $\binom{n - n_1}{k - k_1}$ способами. При этом, очевидно, любой набор черных шаров может сочетаться с любым набором белых. Поэтому всех выборов объема k , различающихся составом и содержащих ровно k_1 черных шаров, будет $\binom{n_1}{k_1} \binom{n - n_1}{k - k_1}$. Значит, искомая вероятность равна

$$P_{n_1, n}(k_1, k) = \binom{n_1}{k_1} \binom{n - n_1}{k - k_1} / \binom{n}{k}.$$

Набор чисел $P_{n_1, n}(0, k)$, $P_{n_1, n}(1, k)$, ..., $P_{n_1, n}(k, k)$ образует так называемое *гипергеометрическое распределение*. Из полученной формулы следует, в частности, что при любом $0 < n_1 < n$

$$\sum_{k_1=0}^k \binom{n_1}{k_1} \binom{n - n_1}{k - k_1} = \binom{n}{k}.$$

Пример 1. В 90-х годах существовала популярная разновидность лотереи, носившая название «спортлото». Участник лотереи из 49 наименований видов спорта (обозначенных просто цифрами) называет шесть. Выигрыш определяется тем, сколько наименований он угадал из шести других наименований, которые в момент розыгрыша лотереи определяются с помощью специального механического устройства, реализующего на глазах у публики случайный выбор. Спрашивается, какова вероятность того, что участник угадает все шесть наименований? Пять наименований и т. д.?

Нетрудно видеть, что это есть не что иное как задача о гипергеометрическом распределении, где в генеральной совокупности, состоящей из 49 предметов, владельцем лотерейного билета выделено шесть «белых». Поэтому вероятность того, что из случайно выбранных шести предметов k_1 предметов окажутся «белыми» (совпадут с отмеченными владельцем билета), равна $P_{6, 49}(k_1, k)$, где объем выборки k равен 6. Например, вероятность угадать все шесть наименований равна

$$P_{6, 49}(6, 6) = \binom{49}{6}^{-1} \approx 7,2 \cdot 10^{-8}.$$

В связи с гипергеометрическим распределением можно сделать одно замечание о природе задач теории вероятностей и математической статистики. Зная состав генеральной совокупности, мы с помощью гипергеометрического распределения можем выяснить, каким может быть состав выборки. Это типичная *прямая* задача теории вероятностей. Но в естественных науках обычно приходится решать *обратные задачи* — по составу выборок определять природу генеральных совокупностей. Такого рода обратные задачи, образно говоря, и составляют содержание математической статистики.

§ 3. Схема Бернулли

Пусть из генеральной совокупности, состоящей из двух элементов $\{0, 1\}$, производится выборка с возвращением объема r . Всего выборок будет 2^r . Пусть p — число из интервала $[0, 1]$. На множестве Ω всех выборок определим неотрицательную функцию P следующим образом: если в выборке содержится ровно k единиц, то $P(\omega) = p^k(1-p)^{r-k}$. Чтобы убедиться, что P есть вероятность, надо доказать равенство

$$P(\Omega) = 1.$$

Легко понять, что k единиц на r местах можно расположить $\binom{r}{k}$ способами. Значит, выборок, содержащих ровно k единиц, будет столько же. Теперь уже можно подсчитать вероятность Ω :

$$P(\Omega) = \sum_{k=0}^r \binom{r}{k} p^k (1-p)^{r-k} = (p + (1-p))^r = 1.$$

Среднее равенство здесь есть формула бинома Ньютона. Заодно мы выяснили, какова вероятность того, что в выборке содержится ровно k единиц:

$$P(k, r) = \binom{r}{k} p^k (1-p)^{r-k}.$$

Это так называемое *биномиальное распределение*. Его можно рассматривать как распределение числа «успехов» в последовательности r испытаний с двумя возможными исходами в каждом испытании: 1 («успех») и 0 («неудача»). Такую последовательность испытаний с вероятностью $P(\omega)$, определенной как $p^k(1-p)^{r-k}$, где k — число успехов в ω , называют *схемой Бернулли*. Оказывается, что испытания в схеме Бернулли обладают свойством независимости, которое будет обсуждаться в следующей главе.

Нетрудно проверить, что вероятность появления 1 на фиксированном месте в выборке (скажем, на месте с номером s) равна p . Действительно, удалив из выборки элемент с номером s , мы получим выборку из той же совокупности, но объема $r-1$. Мы найдем искомую вероятность, если вероятности этих укороченных выборок умножим на p и просуммируем по всем коротким выборкам. Ясно, что получится p . Поэтому число p в схеме Бернулли часто называют вероятностью успеха.

Рассуждая таким же образом, обнаружим, что вероятность появления 1 на k фиксированных местах в выборке равна p^k .

Рассмотрим теперь, как ведут себя вероятности различных исходов $P(k, r)$, когда меняется k . Обратимся к отношению

$$R(k, r) = \frac{P(k, r)}{P(k-1, r)} = \frac{p}{1-p} \frac{r-k+1}{k} = \frac{p}{1-p} \left(\frac{r+1}{k} - 1 \right).$$

Очевидно, что оно монотонно убывает с ростом k , причем при $k/(r+1) < p$ это отношение больше 1, при $k/(r+1) > p$ — меньше. Это означает, что вероятности $P(k, r)$ сначала возрастают с ростом k , а затем при $k > p(r+1)$ убывают.

Сказанное позволяет оценивать с помощью величин $P(k, r)$ вероятности

$$Q(k, r) = \sum_{j=0}^k P(j, r)$$

того, что число успехов в схеме Бернулли не превосходит k . Именно, при $k < p(r+1)$

$$Q(k, r) = P(k, r) \left(1 + \frac{1}{R(k, r)} + \frac{1}{R(k, r)R(k-1, r)} + \dots \right) \leq \\ \leq P(k, r) \frac{R(k, r)}{R(k, r) - 1} = P(k, r) \frac{(r+1-k)p}{(r+1)p - k}.$$

Нетрудно понять, что эта оценка будет весьма точной, если числа k и r велики, а отношение $k/(pr)$ не очень близко к 1. В этом случае сумма

$$1 + \frac{1}{R(k, r)} + \frac{1}{R(k, r)R(k-1, r)} + \dots$$

будет мало отличаться от суммы геометрической прогрессии

$$\sum_{j=0}^{\infty} R^{-j}(k, r) = \frac{R(k, r)}{R(k, r) - 1}$$

и будет справедливо приближенное равенство

$$Q(k, r) \approx P(k, r) \frac{(r+1-k)p}{(r+1)p - k}. \quad (1)$$

Например, при $r = 30$, $p = 0,7$, $k = 16$ имеем $rp = 21$, $P(k, r) \approx 0,023$. Здесь отношение $\frac{(r+1-k)p}{(r+1)p - k}$ равно $15 \cdot 0,7/5,7 \approx 1,84$. Следовательно, правая часть в (1), оценивающая $Q(k, r)$, равна приблизительно $0,023 \cdot 1,84 \approx 0,042$. Истинное значение $Q(k, r)$ при выбранных значениях r , p , k равно 0,040 (с точностью до трех знаков после запятой).

Формула (1) будет использована в примере 5.1 (§ 5.2).

Рассмотрим теперь генеральную совокупность, состоящую из n элементов, из них n_1 элементов первого вида и $n_2 = n - n_1$ второго вида. Производится выборка без возвращения объема r .

Теорема 1. Пусть n и n_1 стремятся к бесконечности так, что $n_1/n \rightarrow p$, где p — число из отрезка $[0, 1]$. Тогда для гипергеометрического распределения справедливо соотношение

$$P_{n_1, n}(r_1, r) \rightarrow P(r_1, r).$$

Доказательство. Разделим в формуле для $P_{n_1, n}(r_1, r)$ (см. § 2) числитель и знаменатель на n^r . Положив $r_2 = r - r_1$, $n_2 = n - n_1$ получим

$$P_{n_1, n}(r_1, r) = \frac{r!(n-r)!}{n!} \frac{n_1!}{r_1!(n_1-r_1)!} \frac{n_2!}{r_2!(n_2-r_2)!} = \\ = \frac{r!}{r_1!r_2!} \frac{\frac{n_1}{n} \left(\frac{n_1}{n} - \frac{1}{n} \right) \left(\frac{n_1}{n} - \frac{2}{n} \right) \dots \left(\frac{n_1}{n} - \frac{r_1-1}{n} \right)}{\frac{n}{n} \left(1 - \frac{1}{n} \right) \dots \left(1 - \frac{r-1}{n} \right)} \frac{n_2}{n} \left(\frac{n_2}{n} - \frac{1}{n} \right) \dots \\ \dots \left(\frac{n_2}{n} - \frac{r_2-1}{n} \right) \rightarrow \binom{r}{r_1} p^{r_1} (1-p)^{r_2} = P(r_1, r)$$

при $n \rightarrow \infty$. ■

При достаточно больших n в силу доказанного предложения $P_{n_1, n}(r_1, r)$ близко к $P(r_1, r)$. Поэтому схему Бернулли можно представлять себе как выборку без возвращения из очень большой генеральной совокупности с соответствующим составом (доля черных шаров равна p).

Рассмотрим в заключение две задачи.

Представим себе n ячеек, в которые случайно размещаются r перенумерованных частиц. Каждая частица может попасть в любую из n ячеек, поэтому всего различных размещений r частиц в n ячейках будет n^r . Размещение частиц в ячейках можно представить как выборку с возвращением объема r из генеральной совокупности, содержащей n элементов. Мы будем предполагать, что имеем дело с классической схемой, когда вероятность каждого исхода равна $1/n^r$.

1) Чему равна вероятность того, что в k -ю ячейку попало ровно r_1 частиц? Остальные $r - r_1$ частиц, которые не попали в ячейку k , распределены по $n - 1$ оставшимся ячейкам. Способов разместить эти $r - r_1$ частиц по $n - 1$ ячейкам будет $(n - 1)^{r - r_1}$. Из r частиц $r - r_1$ частиц, не попавших в ячейку k , можно выбирать $\binom{r}{r - r_1}$ способами. Поэтому искомая вероятность равна

$$\binom{r}{r - r_1} \frac{(n - 1)^{r - r_1}}{n^r} = \binom{r}{r - r_1} \left(\frac{1}{n}\right)^{r_1} \left(1 - \frac{1}{n}\right)^{r - r_1}$$

Эта вероятность совпадает с $P(r_1, r)$ в схеме Бернулли при $p = 1/n$.

2) Подсчитаем теперь вероятность того, что хотя бы одна ячейка окажется пустой. Обозначим это событие символом A . Пусть A_k означает, что k -я ячейка пуста, тогда

$$A = \bigcup_{k=1}^n A_k.$$

Чтобы найти вероятность события A , нам потребуется формула для вероятности суммы (объединения) событий. При этом мы не можем воспользоваться аддитивностью вероятности, поскольку в нашем случае события A_k пересекаются.

§ 4. Вероятность объединения событий.

Примеры

Вернемся к произвольному дискретному вероятностному пространству.

Теорема 2. Пусть A_1, A_2, \dots, A_n — события. Тогда

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i) - \sum_{i < j} P(A_i A_j) + \sum_{i < j < k} P(A_i A_j A_k) - \dots + (-1)^{n-1} P(A_1 \dots A_n).$$

Доказательство. Надо воспользоваться индукцией и свойством вероятности

$$P(A + B) = P(A) + P(B) - P(AB),$$

доказанным в § 1. При $n = 2$ утверждение теоремы верно. Предположим, что оно верно для произвольных $(n - 1)$ событий A_1, \dots, A_{n-1} . Тогда, обозначив $B = \bigcup_{i=1}^{n-1} A_i$, получим

$$P\left(\bigcup_{i=1}^n A_i\right) = P(B + A_n) = P(B) + P(A_n) - P(A_n B).$$

Подставляя сюда известные значения

$$P(B) = P\left(\bigcup_{i=1}^{n-1} A_i\right) \quad \text{и} \quad P(A_n B) = P\left(\bigcup_{i=1}^{n-1} (A_i A_n)\right),$$

получим утверждение теоремы. ■

Обратимся теперь к задаче о ячейках (см. § 3) и найдем вероятность события A , состоящего в том, что хотя бы одна ячейка окажется пустой. Мы представили A в виде $\bigcup_{k=1}^n A_k$, где A_k означает, что все r частиц не попадают в k -ю ячейку.

Имеем

$$P(A_k) = \frac{(n-1)^r}{n^r} = \left(1 - \frac{1}{n}\right)^r, \quad k \leq n.$$

Событие $A_k A_l$ означает, что все r частиц размещаются в $n-2$ ячейках с номерами, не равными k и l , поэтому

$$P(A_k A_l) = \frac{(n-2)^r}{n^r} = \left(1 - \frac{2}{n}\right)^r, \quad k, l \leq n.$$

Аналогично

$$P(A_k A_l A_m) = \frac{(n-3)^r}{n^r} = \left(1 - \frac{3}{n}\right)^r, \quad k, l, m \leq n,$$

и т. д. Вероятность события A по теореме 2 равна

$$P(A) = n \left(1 - \frac{1}{n}\right)^r - \binom{n}{2} \left(1 - \frac{2}{n}\right)^r + \dots = \sum_{j=1}^n (-1)^{j-1} \binom{n}{j} \left(1 - \frac{j}{n}\right)^r.$$

Обсуждение этой задачи будет продолжено в примере 4.5.

В качестве примера использования теоремы 2 рассмотрим еще одну задачу, имеющую много разного рода приложений. Это так называемая *задача о совпадениях*.

Имеется n элементов, расположенных в некотором порядке. Случайным образом они переставляются (все $n!$ перестановок равновероятны). Какова вероятность того, что хотя бы один элемент окажется на том же самом месте?

Всего перестановок $n!$. Пусть A_k означает, что k -й элемент окажется на своем месте. Это событие содержит $(n-1)!$ исходов, а его вероятность равна

$$P(A_k) = \frac{(n-1)!}{n!}.$$

Событие $A_k A_l$ означает, что k -й и l -й элементы попали на свои места; поэтому

$$P(A_k A_l) = \frac{(n-2)!}{n!}, \dots, \quad P(A_1 \dots A_n) = \frac{(n - (n-1))!}{n!} = \frac{1}{n!}.$$

Событие $\bigcup_{k=1}^n A_k$ и есть событие, состоящее в том, что хотя бы один элемент попал на свое место. Таким образом, мы можем воспользоваться только что доказанной формулой

$$\begin{aligned} P\left(\bigcup_{k=1}^n A_k\right) &= \binom{n}{1} \frac{(n-1)!}{n!} - \binom{n}{2} \frac{(n-2)!}{n!} + \binom{n}{3} \frac{(n-3)!}{n!} - \\ &\quad - \dots + (-1)^{n-1} \frac{1}{n!} = 1 - \frac{1}{2!} + \frac{1}{3!} - \dots + \frac{(-1)^{n-1}}{n!} = \\ &= 1 - \left(1 - 1 + \frac{1}{2!} - \frac{1}{3!} + \dots + \frac{(-1)^n}{n!}\right). \end{aligned}$$

Выражение в скобках есть первые $n+1$ членов разложения в ряд e^{-1} . Поэтому если $n \rightarrow \infty$, то

$$P\left(\bigcup_{k=1}^n A_k\right) \rightarrow 1 - e^{-1}.$$

Г л а в а 2

Произвольное пространство элементарных событий

§ 1. Аксиомы теории вероятностей. Вероятностное пространство

До сих пор мы рассматривали такие задачи, в которых множество исходов состояло не более чем из счетного числа элементов. В этом случае вероятность $P(A)$ определялась нами с помощью вероятностей $P(\omega)$ элементарных исходов ω . Ею оказалась функция, определенная на всех подмножествах A пространства элементарных событий Ω и обладающая свойствами:

1. $P(A) \geq 0$.
2. $P(\Omega) = 1$.
3. Для непересекающихся событий A_1, A_2, \dots

$$P(\cup A_j) = \sum P(A_j).$$

Однако, как уже отмечалось, легко можно представить задачу, где множество всех исходов несчетно. Например, случайное бросание точки в отрезок $[t_1, t_2]$ (скажем, эксперимент с измерением температуры) имеет континуум исходов, так как результатом может быть любая точка отрезка. При этом если в экспериментах, имеющих конечное или счетное множество исходов, любая совокупность исходов представляла событие, то в рассматриваемом примере дело обстоит иначе. Мы встретимся с большими трудностями, если будем считать событием *любое* подмножество этого отрезка. Здесь в качестве событий нужно выделить *специальный класс подмножеств*.

Пусть пространство элементарных событий Ω есть произвольное множество, а \mathcal{A} — некоторая система подмножеств множества Ω .

Определение 1. \mathcal{A} называется *алгеброй*, если:

A1. $\Omega \in \mathcal{A}$.

A2. Из того, что $A \in \mathcal{A}$ и $B \in \mathcal{A}$, следует

$$A \cup B \in \mathcal{A}, \quad A \cap B \in \mathcal{A}.$$

A3. Если $A \in \mathcal{A}$, то $\bar{A} \in \mathcal{A}$.

Нетрудно видеть, что в условии A2 достаточно требовать выполнения лишь одного из приведенных двух соотношений. Второе будет выполняться автоматически.

Алгебру \mathcal{A} иногда называют также *кольцом*, так как в \mathcal{A} определены две операции (сложения и умножения), не выводящие из \mathcal{A} . Алгебра \mathcal{A} является *кольцом с единицей*, так как $\Omega \in \mathcal{A}$ и $A\Omega = \Omega A = A$ для любого $A \in \mathcal{A}$.

Определение 2. Класс множеств \mathfrak{F} называется σ -алгеброй (или σ -кольцом, или борелевским полем событий), если свойство A2 выполняется для любых последовательностей множеств:

A2'. Если $\{A_n\}$ есть последовательность множеств из \mathfrak{F} , то

$$\bigcup_{n=1}^{\infty} A_n \in \mathfrak{F}, \quad \bigcap_{n=1}^{\infty} A_n \in \mathfrak{F}.$$

Здесь, как и в A2, достаточно, чтобы выполнялось лишь одно из этих двух соотношений. Второе будет следствием равенства $\bigcap A_n = \bigcup \overline{A_n}$.

Таким образом, алгебра есть класс множеств, замкнутый относительно *конечного* числа операций дополнения, объединения и пересечения; σ -алгебра есть класс множеств, замкнутый относительно *счетного* числа этих операций.

Если задано множество Ω и какая-нибудь алгебра или σ -алгебра \mathfrak{F} его подмножеств, то говорят, что задано *измеримое пространство* $\langle \Omega, \mathfrak{F} \rangle$.

На отрезке $[0, 1]$ все множества, составленные из конечного числа отрезков или интервалов, образуют алгебру, но не σ -алгебру.

Рассмотрим все σ -алгебры, содержащие интервалы (по крайней мере одна такая σ -алгебра имеется, поскольку совокупность всех подмножеств данного множества образует, очевидно, σ -алгебру). Легко видеть, что пересечение всех таких σ -алгебр (т.е. совокупность множеств, принадлежащих всем σ -алгебрам одновременно) снова образует σ -алгебру. Это есть *минимальная σ -алгебра, содержащая интервалы*; ее называют *борелевской*. Борелевскую σ -алгебру, грубо говоря, можно представлять себе как совокупность множеств, полученных из интервалов с помощью счетного числа операций объединения, пересечения и взятия дополнений. Это весьма богатый класс множеств, заведомо достаточный для практических целей. Элементы борелевской σ -алгебры называют *борелевскими множествами*.

Наряду с интервалами (a, b) борелевскими множествами являются одноточечные множества $\{a\}$ и множества вида $(a, b]$, $[a, b]$, $[a, b)$ (a и b могут принимать и бесконечные значения). Это утверждение следует, например, из представлений вида

$$\{a\} = \bigcap_{n=1}^{\infty} \left(a - \frac{1}{n}, a + \frac{1}{n} \right), \quad (a, b] = \bigcap_{n=1}^{\infty} \left(a, b + \frac{1}{n} \right).$$

Таким образом, все счетные множества и счетные объединения интервалов или отрезков также являются борелевскими.

Если дан какой-нибудь класс \mathcal{B} множеств из Ω , то мы можем рассмотреть снова пересечение всех σ -алгебр, содержащих \mathcal{B} , и получить таким способом *наименьшую σ -алгебру, содержащую \mathcal{B}* .

Определение 3. Наименьшая σ -алгебра, содержащая \mathcal{B} , называется σ -алгеброй, порожденной классом \mathcal{B} , и обозначается через $\sigma(\mathcal{B})$.

В этих терминах борелевская σ -алгебра в n -мерном евклидовом пространстве \mathbb{R}^n есть σ -алгебра, порожденная прямоугольниками или сферами. Если Ω счетно, то σ -алгебра, порожденная элементами $\omega \in \Omega$, очевидно, совпадает с σ -алгеброй всех подмножеств Ω .

В качестве упражнения предлагаем читателю описать алгебру и σ -алгебру множеств на $\Omega = [0, 1]$, порожденную:

- интервалами $(0, 1/3)$ и $(1/3, 1)$;
- полуинтервалами вида $(a, 1]$, $0 < a < 1$;
- отдельными точками.

Чтобы формализовать какую-либо вероятностную задачу, надо соответствующему эксперименту приписать измеримое пространство $\langle \Omega, \mathfrak{F} \rangle$. Символ Ω означает множество элементарных исходов эксперимента, а алгебра или σ -алгебра \mathfrak{F} выделяет класс событий. Все остальные подмножества Ω , не входящие в \mathfrak{F} , *событиями не являются*. При этом класс событий \mathfrak{F} часто будет удобно определять как σ -алгебру, порожденную той или иной алгеброй \mathcal{A} .

Выделение той или иной алгебры или σ -алгебры событий \mathfrak{F} обусловлено, с одной стороны, существом рассматриваемой задачи, с другой стороны — природой множества Ω . Как мы увидим, далеко не всегда можно определить вероятность так, чтобы она имела смысл для *любого* подмножества Ω .

В главе 1 уже отмечалось, что в теории вероятностей наряду с обычной теоретико-множественной терминологией используется несколько иная терминология, связанная с тем, что множества из Ω (принадлежащие \mathfrak{F}) интерпретируются как события. В связи с этим само множество Ω часто называют *достоверным событием*. По аксиомам A1 и A3 пустое множество \emptyset также принадлежит \mathfrak{F} ; его называют *невозможным событием*. \bar{A} называется *противоположным* или *дополнительным событием* к A . Если $A \cap B = \emptyset$, то события A и B называются *несовместными*.

Теперь осталось ввести понятие вероятности событий. Рассмотрим пространство Ω и какую-нибудь выделенную в нем систему множеств \mathcal{A} , образующую алгебру событий.

Определение 4. Вероятность на $\langle \Omega, \mathcal{A} \rangle$ есть числовая функция, определенная на множествах из \mathcal{A} и обладающая следующими свойствами:

P1. $P(A) \geq 0$ для любого $A \in \mathcal{A}$.

P2. $P(\Omega) = 1$.

P3. Если последовательность $\{A_n\}$ событий такова, что $A_i A_j = \emptyset$ при $i \neq j$,

$\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$, то

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n). \quad (1)$$

Эквивалентным аксиоме P3 будет требование аддитивности (1) для *конечного* набора событий A_j и следующая

Аксиома непрерывности.

P3'. Пусть последовательность $\{B_n\}$ событий такова, что $B_{n+1} \subset B_n$ и $\bigcap_{n=1}^{\infty} B_n = B \in \mathcal{A}$, тогда $P(B_n) \rightarrow P(B)$ при $n \rightarrow \infty$.

Доказательство эквивалентности. Пусть выполнено P3, и пусть $B_{n+1} \subset B_n$, $\bigcap_{n=1}^{\infty} B_n = B \in \mathcal{A}$. Тогда последовательность событий B , $C_k = B_k \bar{B}_{k+1}$, $k = 1, 2, \dots$, состоит из попарно несовместных событий и $B_n = B + \bigcup_{k=n}^{\infty} C_k$ для любого n .

Воспользовавшись теперь свойством P3, получим, что ряд $P(B_1) = P(B) + \sum_{k=1}^{\infty} P(C_k)$ сходится, а это значит, что при $n \rightarrow \infty$

$$P(B_n) = P(B) + \sum_{k=n}^{\infty} P(C_k) \rightarrow P(B).$$

Это и есть свойство P3'.

Наоборот, если $\{A_n\}$ — последовательность несовместных событий, то

$$P\left(\bigcup_{k=1}^{\infty} A_k\right) = P\left(\bigcup_{k=1}^n A_k\right) + P\left(\bigcup_{k=n+1}^{\infty} A_k\right)$$

и имеют место равенства

$$\begin{aligned} \sum_{k=1}^{\infty} P(A_k) &= \lim_{n \rightarrow \infty} \sum_{k=1}^n P(A_k) = \lim_{n \rightarrow \infty} P\left(\bigcup_{k=1}^n A_k\right) = \\ &= \lim_{n \rightarrow \infty} \left\{ P\left(\bigcup_{k=1}^{\infty} A_k\right) - P\left(\bigcup_{k=n+1}^{\infty} A_k\right) \right\} = P\left(\bigcup_{k=1}^{\infty} A_k\right). \end{aligned}$$

Последнее равенство следует из P3'.

Определение 5. Тройка $\langle \Omega, \mathcal{A}, P \rangle$ называется *вероятностным пространством в широком смысле*.

Если некоторая алгебра \mathcal{F} образует σ -алгебру ($\mathcal{F} = \sigma(\mathcal{F})$), то условие $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$ в аксиоме P3 (для вероятности на $\langle \Omega, \mathcal{F} \rangle$) будет выполняться автоматически.

Определение 6. Тройка $\langle \Omega, \mathcal{F}, P \rangle$, где \mathcal{F} — σ -алгебра, называется просто *вероятностным пространством*.

Вероятность P на $\langle \Omega, \mathcal{F} \rangle$ называют иногда также *распределением вероятностей* на Ω или просто *распределением* на Ω (на $\langle \Omega, \mathcal{F} \rangle$).

Таким образом, задание вероятностного пространства есть задание счетно-аддитивной неотрицательной меры на измеримом пространстве, такой, что мера Ω равна единице. В таком виде аксиоматика теории вероятностей была сформулирована А. Н. Колмогоровым. Введенная система аксиом является неполной и непротиворечивой.

Построение вероятностного пространства $\langle \Omega, \mathcal{F}, P \rangle$ является *основным этапом* в создании математической модели (формализации) того или иного эксперимента.

Дискуссии о том, что следует понимать под вероятностью, имеют длинную историю и связаны с желанием увязать определение вероятности с ее «физической» природой. Однако ввиду сложности последней такие попытки всегда наталкивались на существенные трудности не только математического, но и философского характера (см. введение). Наиболее важные этапы этой дискуссии связаны с именами Бореля, Мизеса, Бернштейна, Колмогорова. Появление аксиоматики Колмогорова как бы отделяет математическую сторону вопроса от всего остального. При таком подходе «физическая трактовка» понятия вероятности появляется уже в виде теоремы (усиленный закон больших чисел, см. гл. 5, 7), в силу которой частота появления некоторого события при неограниченном повторении независимых испытаний сближается (в строго определенном смысле) с вероятностью этого события.

Рассмотрим теперь примеры наиболее употребимых измеримых и вероятностных пространств.

1. *Дискретное измеримое пространство.* Так мы будем называть пространство $\langle \Omega, \mathcal{F} \rangle$, где Ω состоит из конечного или счетного числа элементов, а σ -алгебра \mathcal{F} обычно составлена из всех подмножеств множества Ω . *Дискретные вероятностные пространства*, построенные на дискретных измеримых пространствах, изучались в главе 1. Там же рассмотрено много конкретных примеров.

2. Измеримое пространство $\langle \mathbb{R}, \mathfrak{B} \rangle$, где \mathbb{R} — вещественная прямая (или ее часть), \mathfrak{B} — σ -алгебра борелевских множеств. Необходимость рассматривать такие пространства возникает в тех случаях, когда результаты интересующих нас наблюдений могут принимать любые значения из \mathbb{R} .

Пример 1. Рассмотрим эксперимент, состоящий в «случайном бросании» точки на отрезок $[0, 1]$. Под этим будем понимать следующее. Множеством Ω элементарных исходов здесь является отрезок $[0, 1]$. В качестве σ -алгебры \mathfrak{F} возьмем множество подмножеств B , для которых определено понятие длины (меры Лебега) $\mu(B)$, например, σ -алгебру \mathfrak{B} измеримых по Борелю множеств. Слова «произведено испытание» здесь означают, что выбрана точка $\omega \in \Omega = [0, 1]$ и вероятность события $\omega \in B$ равна $\mu(B)$. Ясно, что для вероятностного пространства $\langle [0, 1], \mathfrak{B}, \mu \rangle$ все аксиомы выполнены. Мы получили так называемое *равномерное распределение* на $[0, 1]$.

Почему в качестве \mathfrak{F} мы выбрали в этом примере σ -алгебру борелевских множеств \mathfrak{B} ? Если бы мы на $\Omega = [0, 1]$ рассмотрели σ -алгебру, порожденную «отдельными точками» отрезка, то получили бы множества, лебегова мера которых равна либо нулю, либо единице. Другими словами, мы получили бы множества либо очень «плотные», либо очень «худые» (счетные), так что интервалы (a, b) при $b - a < 1$ не принадлежат этой σ -алгебре.

С другой стороны, если рассмотреть на $\Omega = [0, 1]$ σ -алгебру всех подмножеств Ω , то задать вероятностную меру на ней так, чтобы $P([a, b]) = b - a$ (т.е. равномерное распределение), оказывается невозможным¹⁾.

Если обратиться снова к равномерному распределению P на $\Omega = [0, 1]$, то легко понять, что это распределение невозможно задать тем же способом, который мы использовали при определении вероятности на дискретном пространстве элементарных событий (т.е. путем задания вероятностей элементарных исходов ω). Так как в этом примере ω есть отдельные точки из $[0, 1]$, то, очевидно, $P(\omega) = 0$ для любого ω .

3. Измеримое пространство $\langle \mathbb{R}^n, \mathfrak{B}^n \rangle$ возникает в тех случаях, когда наблюдения представляют собой вектор. Здесь \mathbb{R}^n — n -мерное евклидово пространство ($\mathbb{R}^n = \mathbb{R}_1 \times \dots \times \mathbb{R}_n$, где $\mathbb{R}_1, \dots, \mathbb{R}_n$ суть n экземпляров вещественной прямой), \mathfrak{B}^n — σ -алгебра борелевских множеств в \mathbb{R}^n , т.е. алгебра, порожденная множествами $B = B_1 \times \dots \times B_n$, где $B_i \subset \mathbb{R}_i$ есть борелевские множества на прямой. Вместо \mathbb{R}^n можно рассматривать также какую-нибудь измеримую часть $\Omega \in \mathfrak{B}^n$ (например, куб, шар), а вместо \mathfrak{B}^n — сужение \mathfrak{B}^n на Ω . Например, можно аналогично предыдущему построить вероятностное пространство для случайного бросания точки в куб $\Omega = [0, 1]^n$. Здесь полагаем $P(\omega \in B) = \mu(B)$, где $\mu(B)$ — мера Лебега (объем) множества B . Вместо куба $[0, 1]^n$ мы могли бы рассматривать и любой другой куб, например, $[a, b]^n$, но только в этом случае надо полагать $P(\omega \in B) = \mu(B)/\mu(\Omega) = \mu(B)/(b - a)^n$. Это будет *равномерное распределение на кубе*.

В теории вероятностей возникают и более сложные вероятностные пространства. Как быть, если результатом эксперимента является бесконечная числовая последовательность? В этом случае наиболее подходящим часто оказывается пространство $\langle \mathbb{R}^\infty, \mathfrak{B}^\infty \rangle$.

4. Измеримое пространство $\langle \mathbb{R}^\infty, \mathfrak{B}^\infty \rangle$, где $\mathbb{R}^\infty = \prod_{j=1}^{\infty} \mathbb{R}_j$, есть пространство всех

¹⁾ См. [17, с. 80].

последовательностей (x_1, x_2, \dots) , а \mathfrak{B}^∞ есть σ -алгебра, порожденная множествами вида $\bigcap_{k=1}^N B_{j_k}$, $B_{j_k} \in \mathfrak{B}_{j_k}$ при любых N, j_1, \dots, j_N , где \mathfrak{B}_j — σ -алгебра борелевских множеств из \mathbb{R}_j .

5. Если результатом эксперимента является, скажем, непрерывная функция на отрезке $[a, b]$ (траектория движения частицы, электрокардиограмма больного и т.д.), то вероятностные пространства, рассмотренные выше, окажутся непригодными. В этом случае в качестве Ω следовало бы взять пространство $C(a, b)$ всех непрерывных функций на $[a, b]$ или пространство $\mathbb{R}^{[a, b]}$ всех функций на $[a, b]$. Вопрос о выборе подходящей σ -алгебры здесь несколько усложняется, и мы обсудим его позже в главе 17.

Вернемся теперь к определению вероятностного пространства.

Пусть тройка $\langle \Omega, \mathcal{A}, P \rangle$ образует вероятностное пространство в широком смысле (\mathcal{A} есть алгебра). Как мы уже видели, с каждой алгеброй \mathcal{A} связана σ -алгебра $\mathfrak{F} = \sigma(\mathcal{A})$, порожденная \mathcal{A} . Существенный интерес представляет следующий вопрос: определяли ли вероятностная мера P на \mathcal{A} меру на $\mathfrak{F} = \sigma(\mathcal{A})$ и однозначно ли это происходит? Другими словами, достаточно ли для построения вероятностного пространства $\langle \Omega, \mathfrak{F}, P \rangle$ задать вероятность P лишь на какой-нибудь алгебре \mathcal{A} , порождающей \mathfrak{F} (т.е. построить какое-нибудь вероятностное пространство в широком смысле $\langle \Omega, \mathcal{A}, P \rangle$, где $\sigma(\mathcal{A}) = \mathfrak{F}$). Ответ на этот важный вопрос содержится в следующей теореме Каратеодори.

Теорема о продолжении меры. Пусть $\langle \Omega, \mathcal{A}, P \rangle$ — вероятностное пространство в широком смысле. Тогда существует, и притом единственная, вероятностная мера Q , определенная на $\mathfrak{F} = \sigma(\mathcal{A})$ такая, что

$$Q(A) = P(A) \quad \text{для всех} \quad A \in \mathcal{A}.$$

Следствие 1. Каждое вероятностное пространство в широком смысле $\langle \Omega, \mathcal{A}, P \rangle$ автоматически определяет вероятностное пространство $\langle \Omega, \mathfrak{F}, P \rangle$ с $\mathfrak{F} = \sigma(\mathcal{A})$.

Этим фактом мы будем широко пользоваться в дальнейшем. Из него следует, в частности, что для определения вероятности на измеримом пространстве $\langle \mathbb{R}, \mathfrak{B} \rangle$ достаточно задать вероятность интервалов.

Доказательство теоремы Каратеодори отнесено в Приложение 1.

В заключение этого параграфа сделаем одно замечание общего характера. Математика качественно отличается от таких наук, как физика, химия и др. тем, что она в своих выводах не опирается постоянно на опытные данные, с помощью которых естествоиспытатель пытается получить ответы на свои вопросы, а развивается в рамках исходной конструкции или системы аксиом, которыми человек описал исследуемый объект. Таким образом, математика и, в частности, теория вероятностей, исследует природу окружающих нас явлений методологически иначе — изучаются не сами явления непосредственно, а модели этих явлений, созданные однажды на основе человеческого опыта. Ценность той или иной модели определяется соответствием выводов теории и наших наблюдений и поэтому зависит от выбора аксиом, характеризующих объект.

В этом смысле аксиомы P_1 , P_2 и аддитивность вероятности представляются бесспорными и естественными (ср. с замечаниями во введении о желательных свойствах вероятности). Счетная аддитивность вероятности и вместе с ней свойство σ -алгебры A_2' являются качествами более тонкими и менее контролируемыми

нашей интуицией (так же, впрочем, как и многое другое, связанное с понятием бесконечности). Введение последних двух свойств обусловлено, главным образом, возможностью построения содержательной математической теории. Многочисленные приложения теории вероятностей, развитой из сформулированной в этом параграфе системы аксиом, обнаруживают ее высокую эффективность и целесообразность.

§ 2. Свойства вероятности

1. $P(\emptyset) = 0$. Это следует из равенства $\emptyset + \Omega = \Omega$ и свойств вероятности P2, P3.
2. $P(\bar{A}) = 1 - P(A)$, так как $A + \bar{A} = \Omega$ и $A \cap \bar{A} = \emptyset$.
3. Если $A \subset B$, то $P(A) \leq P(B)$. Это следует из того, что $P(A) + P(\bar{A}B) = P(B)$.
4. $P(A) \leq 1$ (см. свойство 3 и P2).
5. $P(A \cup B) = P(A) + P(B) - P(AB)$, так как $A \cup B = A + (B - AB)$, $P(B - AB) = P(B) - P(AB)$.
6. $P(A \cup B) \leq P(A) + P(B)$, следует из предыдущего.
7.
$$P\left(\bigcup_{k=1}^n A_k\right) = \sum_{k=1}^n P(A_k) - \sum_{k < l} P(A_k A_l) +$$
$$+ \sum_{k < l < m} P(A_k A_l A_m) - \dots (-1)^{n-1} P(A_1 \dots A_n).$$

Эта формула была уже доказана и использована для дискретного Ω . Здесь читатель может доказать ее точно так же — используя индукцию и свойство 5.

8. $P\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} P(A_n)$. Так как

$$\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} B_n A_n, \quad \text{где} \quad B_n = \Omega - \bigcup_{k=1}^{n-1} A_k,$$

и события $A_n B_n$ не пересекаются, то

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(B_n A_n) \leq \sum_{n=1}^{\infty} P(A_n).$$

9. Если A_n есть монотонно возрастающая последовательность множеств ($A_n \subset A_{n+1}$) и $A = \bigcup_{n=1}^{\infty} A_n$, то

$$P(A) = \lim_{n \rightarrow \infty} P(A_n).$$

Это другая форма аксиомы непрерывности, эквивалентная P3'.

Действительно, введя в рассмотрение множества $B_n = A - A_n$, мы получим $B_{n+1} \subset B_n$, $\bigcap_{n=1}^{\infty} B_n = \emptyset$. Стало быть, по аксиоме непрерывности при $n \rightarrow \infty$

$$P(A - A_n) = P(A) - P(A_n) \rightarrow 0.$$

Утверждение в обратную сторону получается аналогичным образом.

§ 3. Условная вероятность.

Независимость событий и испытаний

Начнем с примеров. Пусть эксперимент состоит в троекратном подбрасывании симметричной монеты. Вероятность того, что герб выпадет ровно один раз, т.е. произойдет одно из элементарных событий (zpp) , (pzp) , (ppz) , в классической схеме равна $3/8$. Обозначим это событие буквой A . Предположим теперь, что об исходе эксперимента дополнительно известно, что произошло событие

$$B = \{\text{число выпавших гербов нечетно}\}.$$

Какова вероятность события A при этой дополнительной информации? Событие B состоит из четырех элементарных исходов. Событие же A составляется из трех исходов события B . В рамках классической схемы естественно принять новую вероятность события A равной $3/4$.

Рассмотрим еще один более общий пример. Пусть задана классическая схема с n исходами. Событие A состоит из r исходов, событие B — из m исходов, а событие AB содержит k исходов. Вероятность события A при условии, что произошло событие B , по аналогии с предыдущим примером естественно определить следующим образом:

$$P(A|B) = \frac{k}{m} = \frac{k/n}{m/n}.$$

Полученное отношение равно $P(AB)/P(B)$, так как

$$P(AB) = k/n, \quad P(B) = m/n.$$

Можно привести теперь общее

Определение. Пусть задано вероятностное пространство $\langle \Omega, \mathfrak{F}, P \rangle$, и пусть A и B — произвольные события. Если $P(B) > 0$, то *условная вероятность* события A при условии, что произошло событие B , по определению полагается равной

$$P(A|B) = \frac{P(AB)}{P(B)}.$$

Определение 7. События A и B называются *независимыми*, если

$$P(AB) = P(A)P(B).$$

Некоторые свойства независимых событий.

1. Если $P(B) > 0$, то независимость A и B эквивалентна равенству

$$P(A|B) = P(A).$$

Доказательство очевидно.

2. Если A и B независимы, то независимы \bar{A} и B .

Действительно,

$$P(\bar{A}B) = P(B - AB) = P(B) - P(AB) = P(B)(1 - P(A)) = P(\bar{A})P(B).$$

3. Пусть события A и B_1 независимы и независимы также события A и B_2 , при этом $B_1B_2 = \emptyset$. Тогда независимы события A и $B_1 + B_2$.

Следующие равенства доказывают это свойство:

$$\begin{aligned} P(A(B_1 + B_2)) &= P(AB_1 + AB_2) = P(AB_1) + P(AB_2) = \\ &= P(A)(P(B_1) + P(B_2)) = P(A)P(B_1 + B_2). \end{aligned}$$

Как мы увидим ниже, требование $B_1B_2 = \emptyset$ здесь существенно.

Пример 2. Пусть событие A означает выпадение герба в первом из двух бросаний симметричной монеты, а событие B — выпадение решетки во втором бросании. Вероятность каждого из этих событий равна $1/2$. Вероятность пересечения AB будет равна

$$P(AB) = \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} = P(A)P(B).$$

Таким образом, события A и B независимы.

Пример 3. Рассмотрим равномерное распределение на квадрате $[0, 1]^2$ (см. § 1). Пусть событие A состоит в том, что случайно брошенная точка попала в область, расположенную правее абсциссы a , событие B — в том, что точка попала в область, расположенную выше ординаты b .

На рис. 2 обе области заштрихованы. Событие AB на рисунке заштриховано в клеточку. Очевидно, $P(AB) = P(A)P(B)$, и, значит, события A и B независимы.

Легко проверить также, что если событие B означает попадание брошенной точки в треугольник FCD (см. рис. 2), то события A и B будут уже зависимыми.

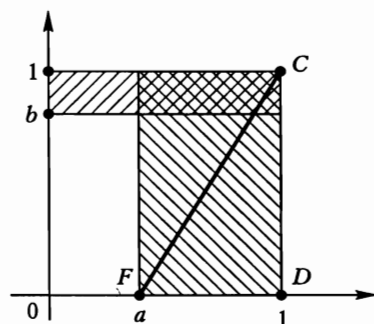


Рис. 2.

Определение 8. События B_1, B_2, \dots, B_n независимы в совокупности, если для любых $1 \leq i_1 < i_2 < \dots < i_r \leq n$, $r = 2, 3, \dots, n$,

$$P\left(\bigcap_{k=1}^r B_{i_k}\right) = \prod_{k=1}^r P(B_{i_k}).$$

Попарной независимости событий недостаточно для независимости n событий в совокупности. Это показывает следующий

Пример 4 (Пример Бернштейна). Рассмотрим такой эксперимент. На плоскость бросается тетраэдр, три грани которого покрашены соответственно в красный, синий и зеленый цвета, а на четвертую нанесены все три цвета. Событие K означает, что при бросании тетраэдра на плоскость выпала грань, содержащая красный цвет, событие C — грань, содержащая синий цвет, и событие $З$ — грань, содержащая зеленый цвет. Так как каждый из трех цветов содержится на двух гранях, то $P(K) = P(C) = P(З) = 1/2$. Вероятность пересечения любой пары введенных событий равна $1/4 = 1/2 \cdot 1/2$, так как любая пара цветов присутствует только на одной грани. Это означает попарную независимость всех трех событий. Но

$$P(KЗC) = \frac{1}{4} \neq P(K)P(З)P(C) = \frac{1}{8}.$$

Теперь легко построить пример, когда при $B_1 B_2 \neq \emptyset$ свойство 3 независимых событий не будет иметь места.

В качестве примера последовательности событий, независимых в совокупности, можно привести последовательность исходов в схеме Бернулли.

Если считать, что каждый исход получен в результате *отдельного испытания*, то мы обнаружим, что любое событие, относящееся к фиксированному испытанию,

будет независимым от любого события, относящегося к остальным испытаниям. В таких случаях говорят о последовательности *независимых испытаний*.

Чтобы дать общее определение, рассмотрим два произвольных эксперимента G_1 и G_2 и обозначим через $\langle \Omega_1, \mathfrak{F}_1, P_1 \rangle$ и $\langle \Omega_2, \mathfrak{F}_2, P_2 \rangle$ соответствующие им вероятностные пространства. Рассмотрим также «составной» эксперимент G с вероятностным пространством $\langle \Omega, \mathfrak{F}, P \rangle$, где $\Omega = \Omega_1 \times \Omega_2$ есть прямое произведение пространств Ω_1 и Ω_2 , а σ -алгебра \mathfrak{F} порождена прямым произведением $\mathfrak{F}_1 \times \mathfrak{F}_2$ (т.е. событиями $B = B_1 \times B_2$, $B_1 \in \mathfrak{F}_1$, $B_2 \in \mathfrak{F}_2$).

Определение 9. Мы будем говорить, что *испытания G_1 и G_2 независимы*, если для любых $B = B_1 \times B_2$, $B_1 \in \mathfrak{F}_1$, $B_2 \in \mathfrak{F}_2$ выполняется

$$P(B) = P_1(B_1)P_2(B_2) = P(B_1 \times \Omega_2)P(\Omega_1 \times B_2).$$

Независимость n испытаний G_1, \dots, G_n определяется аналогично с помощью равенства

$$P(B) = P_1(B_1) \dots P_n(B_n),$$

где $B = B_1 \times \dots \times B_n$, $B_k \in \mathfrak{F}_k$, $\langle \Omega_k, \mathfrak{F}_k, P_k \rangle$ — вероятностное пространство, соответствующее эксперименту G_k , $k = 1, \dots, n$.

В схеме Бернулли вероятность каждой последовательности исходов, состоящей из r нулей и единиц и содержащей k единиц, равна $p^k(1-p)^{r-k}$. Поэтому схему Бернулли можно рассматривать как результат независимых испытаний, в каждом из которых с вероятностью p появляется 1 (успех) и с вероятностью $1-p$ появляется 0 (неуспех). Таким образом, вероятность появления k успехов в r независимых испытаниях равна $\binom{r}{k} p^k (1-p)^{r-k}$.

Верно и в некотором смысле обратное утверждение: любая последовательность независимых одинаковых испытаний с двумя исходами образует схему Бернулли.

В главе 3 будет приведен ряд замечаний о связи введенных понятий независимости с общепринятым пониманием причинной зависимости.

§ 4. Формула полной вероятности и формула Байеса

Пусть A — некоторое событие, B_1, B_2, \dots, B_n — попарно несовместные события, имеющие положительные вероятности и такие, что $A \subset \bigcup_{j=1}^n B_j$. Тогда имеет место *формула полной вероятности*:

$$P(A) = \sum_{j=1}^n P(B_j)P(A | B_j).$$

Доказательство. Из приведенных условий следует, что $A = \bigcup_{j=1}^n B_j A$. Кроме того, события AB_1, AB_2, \dots, AB_n несовместны, поэтому

$$P(A) = \sum_{j=1}^n P(AB_j) = \sum_{j=1}^n P(B_j)P(A | B_j).$$

Пример 5. В экспериментах на встречных электрон-позитронных пучках вероятность того, что за единицу времени произойдет j столкновений, сопровождающихся рождением новых элементарных частиц, равна

$$p_j = \frac{e^{-\lambda} \lambda^j}{j!}, \quad j = 0, 1, \dots,$$

где λ — положительный параметр (это так называемое распределение Пуассона, которое подробно будет рассмотрено в главах 3, 5, 18). При каждом столкновении в результате взаимодействия могут возникнуть различные группы элементарных частиц, при этом вероятность появления каждой группы фиксирована и не зависит от результатов взаимодействия при остальных столкновениях. В качестве одной из таких групп рассмотрим пару μ -мезонов и обозначим через p вероятность их появления при одном столкновении. Чему равна тогда вероятность события A_k , при котором за единицу времени рождается k пар μ -мезонов?

Допустим, что наступило событие B_j , состоящее в том, что за единицу времени происходит j столкновений. При этом условии будем иметь последовательность j независимых испытаний, и вероятность появления k пар μ -мезонов будет равна $\binom{j}{k} p^k (1-p)^{j-k}$. Поэтому по формуле полной вероятности

$$\begin{aligned} P(A_k) &= \sum_{j=k}^{\infty} P(B_j) P(A_k | B_j) = \sum_{j=k}^{\infty} \frac{e^{-\lambda} \lambda^j}{j!} \frac{j!}{k!(j-k)!} p^k (1-p)^{j-k} = \\ &= \frac{e^{-\lambda} p^k \lambda^k}{k!} \sum_{j=0}^{\infty} \frac{(\lambda(1-p))^j}{j!} = \frac{e^{-\lambda p} (p\lambda)^k}{k!}. \end{aligned}$$

Таким образом, мы получили снова распределение Пуассона, но с параметром λp .

Приведенное решение задачи было не формализованным. Формальное решение требует прежде всего построения вероятностного пространства. Оно оказывается здесь достаточно сложным. Обозначим через Ω_j пространство элементарных событий в схеме Бернулли, соответствующей j испытаниям, и пусть ω_j — элемент Ω_j . Тогда в качестве Ω можно взять всевозможные пары $\{(j, \omega_j)\}_{j=0}^{\infty}$, где цифра указывает число столкновений, ω_j означает последовательность длины j «успехов» и «неуспехов» («успех» — появление μ -мезонов). Если ω_j содержит k «успехов», то следует положить

$$P((j, \omega_j)) = p_j p^k (1-p)^{j-k}.$$

Чтобы получить $P(A_k)$, остается просуммировать эти вероятности по всем ω_j , содержащим k успехов, и по всем $j \geq k$ (принцип формулы полной вероятности здесь используется неявно при расслоении A_k на события (j, ω_j)).

Тот факт, что число столкновений частиц описывается распределением Пуассона, нетрудно понять в силу следующих обстоятельств, вызванных существом физического процесса. Пусть событие $B_j(t, u)$ состоит в том, что за интервал времени $[t, t+u]$ произошло j столкновений. Тогда оказывается, что

а) пары событий $B_j(v, t)$ и $B_k(v+t, u)$, относящиеся к непересекающимся интервалам времени, независимы при всех v, t, u, j, k ;

б) при малых Δ вероятность столкновения за время Δ пропорциональна Δ :

$$P(B_1(t, \Delta)) = \lambda \Delta + o(\Delta),$$

и, кроме того, $P(B_k(t, \Delta)) = o(\Delta)$, $k \geq 2$.

Используя снова формулу полной вероятности при гипотезах $B_j(v, t)$, получим для вероятностей $p_k(t) = P(B_k(v, t))$ следующие соотношения:

$$\begin{aligned} p_k(t + \Delta) &= \sum_{j=0}^k p_j(t) P(B_k(v, t + \Delta) | B_j(v, t)) = \sum_{j=0}^k p_j(t) P(B_{k-j}(v + t, \Delta)) = \\ &= o(\Delta) + p_{k-1}(t)(\lambda\Delta + o(\Delta)) + p_k(t)(1 - \lambda\Delta - o(\Delta)), \quad k \geq 1; \\ p_0(t + \Delta) &= p_0(t)(1 - \lambda\Delta - o(\Delta)). \end{aligned}$$

Преобразуя последнее уравнение, находим

$$\frac{p_0(t + \Delta) - p_0(t)}{\Delta} = -\lambda p_0(t) + o(1).$$

Следовательно, существует производная

$$p'_0(t) = -\lambda p_0(t).$$

Аналогичным образом устанавливаем, что существуют

$$p'_k(t) = \lambda p_{k-1}(t) - \lambda p_k(t), \quad k \geq 1. \quad (1)$$

Заметим теперь, что так как функции $p_k(t)$ непрерывны, то следует положить $p_0(0) = 1$, $p_k(0) = 0$ при $k \geq 1$. Поэтому

$$p_0(t) = e^{-\lambda t}.$$

Пользуясь индукцией и подставляя в (1) $p_{k-1}(t) = \frac{(\lambda t)^{k-1} e^{-\lambda t}}{(k-1)!}$, устанавливаем (удобно произвести замену $p_k = e^{-\lambda t} u_k$, тогда уравнение (1) примет вид $u'_k = \frac{\lambda(\lambda t)^{k-1}}{(k-1)!}$), что

$$p_k(t) = \frac{(\lambda t)^k e^{-\lambda t}}{k!}, \quad k = 0, 1, \dots$$

Это есть распределение Пуассона с параметром λt .

Чтобы понять конструкцию вероятностного пространства в этой задаче, следует обратиться к множеству Ω всех ступенчатых неубывающих функций $x(t) \geq 0$, $t \geq 0$, принимающих значения $0, 1, 2, \dots$. Каждая такая функция может играть роль элементарного события: точки ее скачков будут указывать моменты столкновений, само значение $x(t)$ будет означать число столкновений за интервал времени $(0, t)$. Чтобы избежать здесь громоздких рассуждений, связанных с заданием σ -алгебры, можно для наших вычислений считать вероятность P заданной на алгебре \mathcal{A} (см. § 1), порожденной множествами $\{x(t) = k\}$, $t \geq 0$; $k = 0, 1, \dots$ (ведь все рассматриваемые в этой задаче события являются именно такими множествами). Приведенные выше рассуждения показывают, что следует положить

$$P(x(v + t) - x(v) = k) = \frac{e^{-\lambda t} (\lambda t)^k}{k!}$$

(см. также о пуассоновских процессах в главе 18).

Приведенными примерами мы хотели не только проиллюстрировать применение формулы полной вероятности, но и показать, что построение вероятностных пространств в реальных задачах далеко не всегда оказывается простым.

Разумеется, проведение таких построений в каждой задаче вовсе не обязательно, но рекомендуется все же делать это до появления достаточно твердых навыков.

Пусть события A и B_1, \dots, B_n удовлетворяют условиям, сформулированным в начале § 4. Если $P(A) > 0$, то при этих условиях имеет место *формула Байеса*:

$$P(B_j | A) = \frac{P(B_j)P(A | B_j)}{\sum_{k=1}^n P(B_k)P(A | B_k)}.$$

Эта формула есть просто иная форма записи равенства

$$P(B_j | A) = \frac{P(B_j A)}{P(A)},$$

где в числителе надо воспользоваться определением условной вероятности события A при условии B_j , а в знаменателе — формулой полной вероятности.

Пример 6. Некоторое изделие выпускается двумя заводами. При этом объем продукции второго завода в k раз превосходит объем продукции первого. Доля брака у первого завода P_1 , а у второго — P_2 . Предположим теперь, что изделия, выпущенные заводами за одинаковый промежуток времени, перемешали и в таком виде пустили в продажу. Какова вероятность того, что вы приобрели изделие со второго завода, если оно оказалось испорченным? Пусть событие B_1 означает, что выбранное вами изделие с первого завода, а B_2 — со второго. Легко видеть, что

$$P(B_1) = \frac{1}{1+k}, \quad P(B_2) = \frac{k}{1+k}.$$

Это так называемые *априорные* вероятности событий B_1, B_2 . Пусть событие A означает, что выбрано бракованное изделие. Нам даны условные вероятности $P(A | B_1) = P_1$ и $P(A | B_2) = P_2$. Теперь, воспользовавшись формулой Байеса, можно ответить на поставленный вопрос:

$$P(B_2 | A) = \frac{k/(1+k)P_2}{1/(1+k)P_1 + k/(1+k)P_2} = \frac{kP_2}{P_1 + kP_2}.$$

Аналогично, $P(B_1 | A) = P_1/(P_1 + kP_2)$.

Вероятности $P(B_1 | A)$ и $P(B_2 | A)$ называют иногда *апостериорными* вероятностями соответственно событий B_1, B_2 , после того как произошло событие A .

Пример 7. Студенту предложили решить числовую задачу. Ответ задачи известен и представляет собой одно из чисел $1, \dots, k$. Решая задачу, студент может либо найти правильный путь решения задачи, либо ошибиться. Подготовка студента такова, что он находит правильный путь решения задачи с вероятностью p . В этом случае ответ, полученный им, совпадает с данным. С дополнительной вероятностью $1 - p$ студент ошибается. В этом случае мы будем предполагать, что в качестве ответа может получиться любое из чисел $1, \dots, k$ с вероятностью $1/k$.

Известно, что студент получил правильный ответ. Какова при этом условии вероятность того, что он верно решил задачу?

Пусть B_1 (B_2) есть событие, означающее, что студент решил задачу верно (неверно). Тогда по условию априорные вероятности этих событий равны $P(B_1) = p$,

$P(B_2) = 1 - p$. Если событие A состоит в том, что студент получил правильный ответ, то

$$P(A|B_1) = 1, \quad P(A|B_2) = \frac{1}{k}.$$

По формуле Байеса искомая апостериорная вероятность $P(B_1|A)$ равна

$$P(B_1|A) = \frac{P(B_1)P(A|B_1)}{P(B_1)P(A|B_1) + P(B_2)P(A|B_2)} = \frac{p}{p + (1-p)/k} = \frac{1}{1 + (1-p)/kp}.$$

Очевидно, что $P(B_1|A) > P(B_1) = p$ и что $P(B_1|A)$ близка к 1 при больших k .

Глава 3

Случайные величины и функции распределения

§ 1. Определения и примеры

Пусть $\langle \Omega, \mathfrak{F}, P \rangle$ — произвольное вероятностное пространство.

Определение 1. Случайной величиной ξ называется измеримая функция $\xi = \xi(\omega)$, отображающая Ω в множество действительных чисел \mathbb{R} , т.е. функция, для которой прообраз $\xi^{-1}(B) = \{\omega: \xi(\omega) \in B\}$ любого борелевского множества $B \in \mathfrak{B}$ есть множество из σ -алгебры \mathfrak{F} .

В таких случаях говорят также, что ξ осуществляет измеримое отображение $\langle \Omega, \mathfrak{F} \rangle$ в $\langle \mathbb{R}, \mathfrak{B} \rangle$.

Например, при бросании монеты Ω состоит из двух точек: герба и решетки. Если гербу поставить в соответствие 1, а решетке 0, то, очевидно, получим случайную величину.

Случайной величиной будет также число очков на выпавшей грани кости.

Расстояние от начала координат до случайно брошенной точки в квадрат $[0 \leq x \leq 1, 0 \leq y \leq 1]$ тоже будет случайной величиной, так как множество $\{(x, y): x^2 + y^2 < t\}$ измеримо. Читатель, наверное, заметил, что в этих примерах придумать неизмеримую функцию от ω , которая отвечала бы какой-нибудь реальной задаче, очень трудно. Такое явление имеет место часто, но не всегда. В главе 17, посвященной случайным процессам, нас будут интересовать множества, которые, вообще говоря, событиями не являются и требуют специальных модификаций для того, чтобы их можно было считать событиями.

Как отмечалось выше, из определения случайной величины следует, что для любого множества B из σ -алгебры борелевских множеств \mathfrak{B} на прямой

$$\xi^{-1}(B) = \{\omega: \xi(\omega) \in B\} \in \mathfrak{F}.$$

Следовательно, на измеримом пространстве $\langle \mathbb{R}, \mathfrak{B} \rangle$ определена вероятность $P_\xi(B) = P(\xi \in B)$, которая порождает вероятностное пространство $\langle \mathbb{R}, \mathfrak{B}, P_\xi \rangle$.

Определение 2. Вероятность $P_\xi(B)$ называется *распределением случайной величины* ξ .

Если положить $B = (-\infty, x)$, то получим функцию $F_\xi(x) = P(\xi < x)$, определенную на всей вещественной оси, которая называется *функцией распределения случайной величины* ξ .

Ниже мы увидим, что функция распределения случайной величины полностью определяет ее распределение и часто используется для описания последнего.

Там, где это не вызовет недоразумений, вместо $F_\xi(x)$ будем писать просто $F(x)$.

Так как случайная величина ξ есть отображение Ω в \mathbb{R} , то $P(|\xi| < \infty) = 1$. Иногда наряду с такими случайными величинами бывает удобно рассматривать также случайные величины, которые могут принимать значения $\pm\infty$ (они будут осуществлять измеримое отображение Ω в $\mathbb{R} \cup \{\pm\infty\}$). Если $P(|\xi| = \infty) > 0$, то такие случайные величины $\xi(\omega)$ мы будем называть *несобственными*. Каждый случай, когда такие величины будут появляться, мы будем особо оговаривать.

Пример 1. Рассмотрим схему Бернулли с вероятностью успеха p и объемом выборки k (см. § 1.3). Как известно, множество элементарных исходов Ω в этом случае есть множество всевозможных последовательностей из нулей и единиц длины k . За σ -алгебру \mathfrak{F} возьмем систему всех подмножеств Ω . Определим на Ω случайную величину следующим образом: каждой последовательности из нулей и единиц поставим в соответствие число единиц в этой последовательности.

Вероятность r успехов, как мы уже знаем, есть

$$P(r, k) = \binom{k}{r} p^r (1-p)^{k-r}.$$

Поэтому функция распределения $F(x)$ введенной случайной величины будет определяться следующим образом:

$$F(x) = \sum_{r \leq x} P(r, k).$$

Суммирование здесь ведется по всем целым r , меньшим x . Если $x < 0$, то $F(x) = 0$, если же $x > k$, то $F(x) = 1$.

Пример 2. Пусть на отрезок $[a, b]$ вещественной прямой случайным образом бросается точка, т.е. вероятность попадания точки в некоторое множество из $[a, b]$ полагается пропорциональной лебеговой мере этого множества. Ω здесь есть отрезок $[a, b]$, σ -алгебра \mathfrak{F} есть класс борелевских подмножеств из $[a, b]$. Определим случайную величину ξ так:

$$\xi(\omega) = \omega, \quad \omega \in [a, b],$$

т.е. случайная величина равна тому числу из $[a, b]$, в которое попала бросаемая точка. Это измеримая функция. Если $x < a$, то $F(x) = P(\xi < x) = 0$. Пусть $x \in [a, b]$. Тогда $\{\xi < x\}$ означает, что точка попала в интервал $[a, x)$. Вероятность попадания в этот интервал пропорциональна длине этого интервала, значит,

$$F(x) = P(\xi < x) = \frac{x-a}{b-a}.$$

Если же $x > b$, то понятно, что $F(x) = 1$. Окончательно находим

$$F(x) = \begin{cases} 0, & x < a, \\ \frac{x-a}{b-a}, & a \leq x \leq b, \\ 1, & x > b. \end{cases} \quad (1)$$

Эта функция распределения определяет так называемое *равномерное распределение* на интервале $[a, b]$.

Если $\mu(B)$ — мера Лебега на $(\mathbb{R}, \mathfrak{B})$, то, как мы увидим в следующем параграфе, нетрудно установить также, что в этом случае $P_\xi(B) = 1/(b-a)\mu(B \cap [a, b])$.

§ 2. Свойства функций распределения и примеры

Пусть $F(x)$ есть функция распределения случайной величины ξ . Тогда имеют место следующие свойства $F(x)$:

F1. *Свойство монотонности*: если $x_1 \leq x_2$, то $F(x_1) \leq F(x_2)$.

F2. $\lim_{x \rightarrow -\infty} F(x) = 0$ и $\lim_{x \rightarrow \infty} F(x) = 1$.

F3. *Свойство непрерывности слева*: $\lim_{x \uparrow x_0} F(x) = F(x_0)$.

Доказательство. Так как $\{\xi < x_1\} \subseteq \{\xi < x_2\}$, где $x_1 \leq x_2$, то F1 сразу следует из свойства 3 вероятности (см. § 2.2).

Для доказательства F2 рассмотрим две числовые последовательности $\{x_n\}$ и $\{y_n\}$, при этом последовательность $\{x_n\}$ убывает так, что $x_n \rightarrow -\infty$, а последовательность y_n возрастает, $y_n \rightarrow \infty$. Введем обозначения $A_n = \{\xi < x_n\}$, $B_n = \{\xi < y_n\}$. Из того, что $\{x_n\}$ монотонно стремится к $-\infty$, следует, что последовательность множеств A_n убывает и $\bigcap A_n = \emptyset$. В силу аксиомы непрерывности P3' (см. § 1 главы 2) $P(A_n) \rightarrow 0$ при $n \rightarrow \infty$ или, что то же, $\lim_{n \rightarrow \infty} F(x_n) = 0$. Отсюда и из монотонности $F(x)$ следует, что

$$\lim_{x \rightarrow -\infty} F(x) = 0.$$

Так как последовательность $\{y_n\}$ монотонно сходится к ∞ , то последовательность множеств B_n возрастает и $\bigcup B_n = \Omega$, поэтому (см. свойство 9 из § 2.2) $P(B_n) \rightarrow 1$. Отсюда, как и прежде, следует, что

$$\lim_{n \rightarrow \infty} F(y_n) = 1, \quad \lim_{x \rightarrow \infty} F(x) = 1.$$

Свойство F3 доказывается аналогичным способом. Пусть $A = \{\xi < x_0\}$, $A_n = \{\xi < x_n\}$, где последовательность $\{x_n\}$ возрастает, $x_n \uparrow x_0$. Последовательность множеств A_n также возрастает и $\bigcup A_n = A$. Следовательно, $P(A_n) \rightarrow P(A)$. Это означает, что

$$\lim_{x \uparrow x_0} F(x) = F(x_0).$$

Нетрудно видеть, что функция F была бы *непрерывной справа*, если бы мы положили $F(x) = P(\xi \leq x)$.

При принятом определении функция F , вообще говоря, не является непрерывной справа, так как по аксиоме непрерывности

$$\begin{aligned} F(x+0) - F(x) &= \lim_{n \rightarrow \infty} \left(F\left(x + \frac{1}{n}\right) - F(x) \right) = \\ &= \lim_{n \rightarrow \infty} P\left(x \leq \xi < x + \frac{1}{n}\right) = P\left(\bigcap_{n=1}^{\infty} \left\{ \xi \in \left[x, x + \frac{1}{n}\right) \right\}\right) = P(\xi = x). \end{aligned}$$

Сказанное означает, что $F(x)$ непрерывна тогда и только тогда, когда $P(\xi = x) = 0$ при любом x . Примеры 1, 2 показывают, что весьма обычны как непрерывные, так и разрывные $F(x)$.

Из приведенных соотношений следует также, что

$$P(x \leq \xi \leq y) = P_{\xi}([x, y]) = F(y+0) - F(x).$$

Теорема 1. Если функция $F(x)$ обладает свойствами F1, F2 и F3, то существует вероятностное пространство $(\Omega, \mathfrak{F}, P)$ и случайная величина ξ на нем такая, что $F_{\xi}(x) = F(x)$.

Доказательство. Построим сначала вероятностное пространство $\langle \Omega, \mathfrak{F}, P \rangle$. В качестве Ω возьмем вещественную прямую \mathbb{R} , а в качестве \mathfrak{F} возьмем σ -алгебру \mathfrak{B} измеримых по Борелю множеств. Как мы уже знаем (см. § 2.1), для построения вероятностного пространства $\langle \mathbb{R}, \mathfrak{B}, P \rangle$ достаточно задать вероятность P на алгебре \mathcal{A} , порожденной, скажем, полуинтервалами вида $[\cdot, \cdot)$ (тогда $\sigma(\mathcal{A}) = \mathfrak{B}$). Произвольный элемент A алгебры \mathcal{A} имеет вид конечного объединения непересекающихся полуинтервалов:

$$A = \bigcup_{i=1}^n [a_i, b_i), \quad a_i < b_i$$

(числа a_i и b_i могут быть бесконечными). Полагаем по определению

$$P(A) = \sum_{i=1}^n (F(b_i) - F(a_i)).$$

Совершенно очевидно, что в силу F1, F2 аксиомы P1, P2 выполнены. Остается проверить счетную аддитивность или непрерывность P на алгебре \mathcal{A} . Пусть $B_n \in \mathcal{A}$, $B_{n+1} \subset B_n$, $\bigcap_{i=1}^{\infty} B_n = B \in \mathcal{A}$. Надо показать, что $P(B_n) \rightarrow P(B)$ или, что то же, $P(B_n \bar{B}) \rightarrow 0$ ($B_n \bar{B} \in \mathcal{A}$). Для этого достаточно убедиться, что $P(B_n \bar{B} C_N) \rightarrow 0$ при любом фиксированном N , где $C_N = [-N, N)$. Действительно, для заданного $\varepsilon > 0$ по свойству F2 выберем N так, что $P(\bar{C}_N) < \varepsilon$. Тогда $P(B_n \bar{B} C_N) \leq P(\bar{C}_N) \leq \varepsilon$,

$$\lim_{n \rightarrow \infty} P(B_n \bar{B}) \leq \lim_{n \rightarrow \infty} P(B_n \bar{B} C_N) + \varepsilon.$$

Так как ε произвольно, то сходимость $P(B_n \bar{B} C_N) \rightarrow 0$ при $n \rightarrow \infty$ означает требуемую сходимость $P(B_n \bar{B}) \rightarrow 0$. Из сказанного вытекает, что можно считать множества B_n ограниченными ($B_n \subset [-N, N)$ при каком-нибудь $N < \infty$). Кроме того, не ограничивая общности, можно считать, что B есть пустое множество. В силу сделанных выше замечаний B_n допускают представление

$$B_n = \bigcup_{i=1}^{k_n} [a_i^n, b_i^n), \quad k_n < \infty,$$

где a_i^n , b_i^n конечны. Заметим далее, что для заданного $\varepsilon > 0$ и для любого полуинтервала $[a, b)$ всегда можно найти вложенный в него полуинтервал $[a, b - \delta)$, $\delta > 0$, такой, что $P([a, b - \delta)) \geq P([a, b)) - \varepsilon$. Это вытекает непосредственно из свойства F3: $F(b - \delta) \rightarrow F(b)$ при $\delta \downarrow 0$. Следовательно, для заданного $\varepsilon > 0$ и множества B_n существуют $\delta_i^n > 0$, $i = 1, \dots, k_n$, такие, что

$$\tilde{B}_n = \bigcup_{i=1}^{k_n} [a_i^n, b_i^n - \delta_i^n) \subset B_n, \quad P(\tilde{B}_n) > P(B_n) - \varepsilon 2^{-n}.$$

Присоединим теперь к множеству \tilde{B}_n правые концы полуинтервалов и рассмотрим замкнутое ограниченное множество

$$K_n = \bigcup_{i=1}^{k_n} [a_i^n, b_i^n - \delta_i^n].$$

Очевидно, что

$$\widetilde{B}_n \subset K_n \subset B_n, \quad K = \bigcap_{n=1}^{\infty} K_n = \emptyset,$$

$$P(B_n - K_n) = P(B_n \overline{K}_n) \leq \varepsilon 2^{-n}.$$

Из соотношения $K = \emptyset$ следует, что $K_n = \emptyset$ при всех достаточно больших n . Действительно, все множества K_n принадлежат замыканию $[C_N] = [-N, N]$, которое является компактом. Множества $\{\Delta_n = [C_N] - K_n\}_{n=1}^{\infty}$ образуют *открытое покрытие* $[C_N]$, так как $\bigcup \Delta_n = [C_N](\bigcup \overline{K}_n) = [C_N](\overline{\bigcap K_n}) = [C_N]$. Таким образом, по лемме Гейне—Бореля существует *конечное подпокрытие* $\{\Delta_n\}_{n=1}^{n_0}$, $n_0 < \infty$, такое,

что $\bigcup_{n=1}^{n_0} \Delta_n = [C_N]$ или, что то же, $\bigcap_{n=1}^{n_0} K_n = \emptyset$. Поэтому

$$\begin{aligned} P(B_{n_0}) &= P\left(B_{n_0} \overline{\left(\bigcap_{n=1}^{n_0} K_n\right)}\right) = P\left(B_{n_0} \left(\bigcup_{n=1}^{n_0} \overline{K}_n\right)\right) = \\ &= P\left(\bigcup_{n=1}^{n_0} B_{n_0} \overline{K}_n\right) \leq P\left(\bigcup_{n=1}^{n_0} B_n \overline{K}_n\right) \leq \sum_{n=1}^{n_0} \varepsilon 2^{-n} < \varepsilon. \end{aligned}$$

Итак, для заданного $\varepsilon > 0$ нашлось n_0 (оно зависит от ε) такое, что $P(B_{n_0}) < \varepsilon$. Это значит, что $P(B_n) \rightarrow 0$ при $n \rightarrow \infty$. Выполнение аксиомы 3 доказано.

Итак, вероятностное пространство построено. Остается в качестве случайной величины ξ взять тождественное отображение действительных чисел на себя. Тогда

$$F_{\xi}(x) = P(\xi < x) = P(-\infty, x) = F(x).$$

При изучении функций распределений на основании доказанного утверждения часто используют удобную модель *выборочного* вероятностного пространства.

Определение 3. Вероятностное пространство $(\Omega, \mathfrak{F}, P)$ называется *выборочным* для случайной величины $\xi(\omega)$, если Ω есть подмножество вещественной прямой \mathbb{R} и $\xi(\omega) \equiv \omega$.

Вероятность P в соответствии с определением 1 из § 1 будет называться *распределением* ξ . Этот факт будем обозначать символом

$$\xi \in P. \quad (2)$$

Очевидно, что построение выборочного вероятностного пространства всегда возможно. Достаточно положить $\Omega = \mathbb{R}$, $\mathfrak{F} = \mathfrak{B}$, $P(B) = P(\xi \in B)$. Для целочисленных ξ пространство (Ω, \mathfrak{F}) можно выбрать и более «экономным» образом, положив $\Omega = \{\dots, -1, 0, 1, \dots\}$.

Так как функция распределения $F(x)$ случайной величины ξ по теореме 1 однозначно определяет распределение P этой случайной величины, то наряду с (2) будем писать также $\xi \in F$.

Приведем теперь примеры некоторых наиболее распространенных распределений.

1. *Вырожденное распределение* I_a . Распределение I_a определяется соотношением

$$I_a(B) = \begin{cases} 1, & \text{если } a \in B, \\ 0, & \text{если } a \notin B. \end{cases}$$

Это распределение, сосредоточенное в точке a : если $\xi \in I_a$, то $P(\xi = a) = 1$. Функция распределения для I_a имеет вид

$$F(x) = \begin{cases} 0 & \text{при } x \leq a, \\ 1 & \text{при } x > a. \end{cases}$$

Следующие два распределения были описаны в примерах 1, 2 из § 1.

2. *Биномиальное распределение* B_p^n . По определению $\xi \in B_p^n$ (n целое, $p \in (0, 1)$), если $P(\xi = k) = \binom{n}{k} p^k (1-p)^{n-k}$, $0 \leq k \leq n$. Распределение B_p^1 будем обозначать через B_p .

3. *Равномерное распределение* $U_{a,b}$. Если $\xi \in U_{a,b}$, то $P(\xi \in B) = \frac{\mu(B \cap [a,b])}{\mu([a,b])}$, где μ — мера Лебега. Мы видели, что это распределение имеет функцию распределения (1).

Следующее распределение играет особую роль в теории вероятностей, мы будем сталкиваться с ним многократно.

4. *Нормальное распределение* Φ_{α, σ^2} (нормальный закон или закон Гаусса). Будем писать $\xi \in \Phi_{\alpha, \sigma^2}$, если

$$P(\xi \in B) = \Phi_{\alpha, \sigma^2}(B) = \frac{1}{\sigma\sqrt{2\pi}} \int_B e^{-(u-\alpha)^2/(2\sigma^2)} du. \quad (3)$$

Распределение Φ_{α, σ^2} зависит от двух параметров; α и $\sigma > 0$. Если $\alpha = 0$, $\sigma = 1$, то нормальное распределение называют *стандартным*. Функция распределения $\Phi_{0,1}$ (будем обозначать ее через $\Phi(x)$), равна

$$\Phi(x) = \Phi_{0,1}((-\infty, x)) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{u^2}{2}} du.$$

Очевидно, что функция распределения для Φ_{α, σ^2} равна $\Phi\left(\frac{x-\alpha}{\sigma}\right)$, так что параметры α и σ имеют смысл параметров «сдвига» и «масштаба» распределения.

Тот факт, что формула (3) определяет распределение, следует хотя бы из теоремы 1 и того, что функция $\Phi(x)$ (или $\Phi\left(\frac{x-a}{\sigma}\right)$) удовлетворяет свойствам F1–F3, так как $\Phi(-\infty) = 0$, $\Phi(\infty) = 1$, $\Phi(x)$ непрерывна и монотонна. Можно использовать и непосредственно тот факт, что интеграл в (3) есть счетно-аддитивная функция множества (см. § 6 и Приложение 3).

5. *Показательное распределение* Γ_α . Соотношение $\xi \in \Gamma_\alpha$ означает, что ξ неотрицательна и

$$P(\xi \in B) = \Gamma_\alpha(B) = \alpha \int_{B \cap (0, \infty)} e^{-\alpha u} du.$$

Функция распределения $\xi \in \Gamma_\alpha$ имеет, очевидно, вид

$$P(\xi < x) = \begin{cases} 1 - e^{-\alpha x} & \text{при } x \geq 0, \\ 0 & \text{при } x < 0. \end{cases}$$

Показательное распределение есть частный случай Γ -распределения $\Gamma_{\alpha, \lambda}$, которое мы рассмотрим более подробно в § 7.7.

6. Дискретный аналог показательного распределения называют иногда *геометрическим*. Оно имеет вид

$$P(\xi = k) = (1 - p)p^k, \quad p \in (0, 1), \quad k = 0, 1, \dots$$

7. *Распределение Коши* $K_{\alpha, \sigma}$. Как и нормальное распределение, это есть распределение, зависящее от двух параметров α и $\sigma > 0$, которые также являются параметрами сдвига и масштаба. Если $\xi \in K_{\alpha, \sigma}$, то

$$P(\xi \in B) = \frac{1}{\pi\sigma} \int_B \frac{du}{1 + ((u - \alpha)/\sigma)^2}.$$

Функция распределения $K(x)$ распределения $K_{0,1}$ равна

$$K(x) = \frac{1}{\pi} \int_{-\infty}^x \frac{du}{1 + u^2}.$$

Функция распределения для $K_{\alpha, \sigma}$ равна $K\left(\frac{x - \alpha}{\sigma}\right)$. Здесь сохраняются все замечания, сделанные по поводу нормального распределения.

Пример 3. Предположим, что в некоторой точке (α, σ) , $\sigma > 0$, на плоскости находится источник излучения частиц. Следы излучения фиксируются детектором, который совпадает с осью x (осью абсцисс). Направление, в котором движутся излучаемые частицы, случайно и имеет равномерное распределение на окружности. Другими словами, угол η , составленный этим направлением с осью $(-y, 0)$, имеет равномерное распределение $U_{-\pi, \pi}$ на отрезке $[-\pi, \pi]$. Результатами наблюдений являются координаты ξ_1, ξ_2, \dots точек на оси абсцисс, в которых произошло взаимодействие частиц с детектором. Что из себя представляет распределение случайной величины $\xi = \xi_1$?

Чтобы найти это распределение, рассмотрим частицу, вылетевшую из точки (α, σ) при условии, что эта точка достигла детектора (т.е. при условии, что $\eta \in [-\pi/2, \pi/2]$). Ясно, что условное распределение η при этом условии ($P(\eta \in [-\pi/2, \pi/2]) = 1/2$) совпадает с $U_{-\pi/2, \pi/2}$. Так как $(\xi - \alpha)/\sigma = \operatorname{tg} \eta$, то

$$P(\xi < x) = P(\alpha + \sigma \operatorname{tg} \eta < x) = P\left(\frac{\eta}{\pi} < \frac{1}{\pi} \operatorname{arctg} \frac{x - \alpha}{\sigma}\right) = \frac{1}{2} + \frac{1}{\pi} \operatorname{arctg} \frac{x - \alpha}{\sigma}.$$

Поскольку $(\operatorname{arctg} u)' = 1/(1 + u^2)$, то

$$\begin{aligned} \operatorname{arctg} x &= \int_0^x \frac{du}{1 + u^2} = \int_{-\infty}^x \frac{du}{1 + u^2} - \frac{\pi}{2}, \\ P(\xi < x) &= \frac{1}{\pi} \int_{-\infty}^{\frac{x - \alpha}{\sigma}} \frac{du}{1 + u^2} = K\left(\frac{x - \alpha}{\sigma}\right). \end{aligned}$$

Таким образом, координаты наблюдаемых на оси абсцисс следов излучения из точки (α, σ) имеют распределение Коши $K_{\alpha, \sigma}$.

8. *Распределение Пуассона* Π_λ . Мы будем писать $\xi \in \Pi_\lambda$, если ξ принимает только целые неотрицательные значения с вероятностями

$$P(\xi = m) = \frac{\mu^m}{m!} e^{-\mu}, \quad \mu > 0, \quad m = 0, 1, 2, \dots$$

Функция распределения так же, как и в примере 1, будет иметь вид суммы:

$$F(x) = \begin{cases} \sum_{m \leq x} \frac{\mu^m}{m!} e^{-\mu} & \text{для } x > 0, \\ 0 & \text{для } x \leq 0. \end{cases}$$

Все распределения, рассмотренные в приведенных выше примерах, можно разбить на два типа.

Дискретные распределения.

Определение 4. Распределение случайной величины ξ называется *дискретным*, если ξ может принимать лишь конечное или счетное число значений x_1, x_2, \dots так, что

$$p_k = P(\xi = x_k) > 0, \quad \sum p_k = 1.$$

Очевидно, что дискретное распределение $\{p_k\}$ всегда можно задать на дискретном вероятностном пространстве. Характеризовать такое распределение бывает удобно с помощью таблиц:

Значения	x_1	x_2	\dots
Вероятности	p_1	p_2	\dots

К дискретным распределениям относятся распределения I_α , B_p^n , Π_λ , геометрическое распределение. Производная функций распределения $F(x)$ для этих распределений всюду равна нулю, за исключением точек x_1, x_2, \dots , где $F(x)$ терпит разрыв, причем

$$F(x_k + 0) - F(x_k) = p_k.$$

Абсолютно непрерывные распределения.

Определение 5. Распределение P случайной величины ξ называется *абсолютно непрерывным*¹⁾, если для любого борелевского множества B

$$P(B) = P(\xi \in B) = \int_B f(x) dx, \quad (4)$$

где $f(x) \geq 0$, $\int_{-\infty}^{\infty} f(x) dx = 1$.

¹⁾ Имеется в виду абсолютная непрерывность относительно меры Лебега. Если на $(\mathbb{R}, \mathfrak{B})$ задана мера μ (см. Приложение 3), то распределение P называется *абсолютно непрерывным относительно μ* , если для любого $B \in \mathfrak{B}$

$$P(B) = \int_B f(x) \mu(dx).$$

В этом смысле дискретные распределения также являются абсолютно непрерывными, но относительно считающей меры m . Действительно, если положить $f(x_k) = p_k$, $m(B) = \{\text{числу точек из множества } (x_1, x_2, \dots), \text{ попавших в } B\}$, то

$$P(B) = \sum_{x_k \in B} P_k = \sum_{x_k \in B} f(x_k) = \int_B f(x) m(dx)$$

(см. Приложение 3).

Функция $f(x)$ в равенстве (4) называется *плотностью* распределения.

Из доказательства теоремы 1 (точнее, из теоремы о единственности продолжения меры) нетрудно получить, что приведенное выше определение абсолютной непрерывности эквивалентно представлению

$$F_{\xi}(x) = \int_{-\infty}^x f(u) du$$

для всех $x \in \mathbb{R}$. Функции распределения, обладающие таким свойством, также называются абсолютно непрерывными.

Функция $f(x)$ определяется приведенными выше равенствами с точностью до значений на множестве меры нуль. Для нее почти всюду (по мере Лебега) имеет место равенство²⁾ $f(x) = dF(x)/dx$.

Абсолютно непрерывными являются распределения $U_{a,b}$, Φ_{α,σ^2} , $K_{\alpha,\sigma}$, Γ_{α} . Плотность нормального закона с параметрами α , σ равна

$$\varphi_{\alpha,\sigma^2}(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\alpha)^2/(2\sigma^2)}.$$

Из приведенных ранее равенств очевидным образом можно получить и плотности распределений $U_{a,b}$, $K_{\alpha,\sigma}$, Γ_{α} . Плотность для $K_{\alpha,\sigma}$ имеет вид, напоминающий вид нормальной плотности, но имеет более «толстые хвосты» (более медленное убывание при $|x| \rightarrow \infty$).

Итак, распределения могут быть дискретными и абсолютно непрерывными. Возможны, конечно, смеси двух типов распределений. Но существует еще третий тип, не описываемый предыдущими двумя. Это *сингулярный тип*. Характеризуется он тем, что функция распределения $F(x)$ непрерывна, но точки роста $F(x)$ образуют множество нулевой меры Лебега. (Точка x называется *точкой роста* $F(x)$, если для любого $\varepsilon > 0$ $F(x + \varepsilon) - F(x - \varepsilon) > 0$).

Таким образом, здесь $F(x)$ непрерывна, $dF(x)/dx = 0$ почти всюду, а $F(\infty) - F(-\infty) = 1$.

Примером такой функции распределения может служить известная кривая Кантора, все изменение которой сосредоточено на отрезке $[0, 1]$: $F(x) = 0$ при $x \leq 0$, $F(x) = 1$ при $x \geq 1$. Строится она следующим образом (процесс построения кривой Кантора показан на рис. 3). Отрезок $[0, 1]$ разбивается на три равные части $[0, 1/3]$, $[1/3, 2/3]$, $[2/3, 1]$. На внутреннем сегменте полагаем $F(x) = 1/2$. Оставшиеся два сегмента снова разбиваются на три равные части каждый, и на внутренних сегментах $F(x)$ полагается равной соответственно $1/4$ и $3/4$. Каждый из оставшихся сегментов снова делится на три части, и на внутренних $F(x)$ определяется как

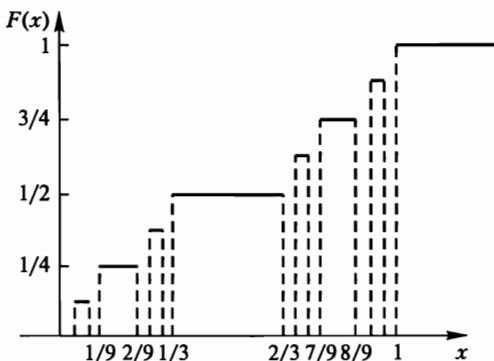


Рис. 3.

²⁾ Утверждение о единственности «почти всюду» функции f следует из теоремы Радона—Никодима (см. Приложение 3).

постоянная, равная среднему арифметическому между соседними, уже определенными значениями $F(x)$, и т.д. В точках, не принадлежащих таким внутренним сегментам, $F(x)$ определяется по непрерывности. Нетрудно видеть, что суммарная длина «внутренних» сегментов, на которых $F(x)$ постоянна, равна

$$\frac{1}{3} + \frac{2}{9} + \frac{4}{27} + \dots = \frac{1}{3} \sum_{k=0}^{\infty} \left(\frac{2}{3}\right)^k = \frac{1}{3} \frac{1}{1 - 2/3} = 1,$$

так что функция $F(x)$ растет на множестве меры нуль, но без скачков.

Известна теорема Лебега³⁾ о том, что любая функция распределения $F(x)$ может быть единственным образом представлена в виде суммы трех компонент: дискретной, абсолютно непрерывной и сингулярной. Отсюда следует, что произвольная функция распределения не может иметь более чем счетное число скачков (это видно и непосредственно: мы перечислим все скачки, если перенумеруем сначала все скачки, большие $1/2$, затем все скачки, большие $1/3$, затем — большие $1/4$, и т.д.). Это означает, в частности, что $F(x)$ непрерывна всюду, за исключением, быть может, счетного или конечного множества точек.

В заключение приведем ряд свойств функций распределения и плотностей, возникающих при переходе к новым случайным величинам.

Пусть $F_{\xi}(x) = P(\xi < x)$ и $g(x)$ есть борелевская функция ($g(x)$ называется борелевской, если прообраз $g^{-1}(B) = \{x: g(x) \in B\}$ борелевского множества B есть снова борелевское множество). Тогда функция распределения случайной величины $\eta = g(\xi)$ равна

$$F_{g(\xi)}(x) = P(g(\xi) < x) = P(\xi \in g^{-1}(-\infty, x)),$$

Если $g(x)$ — неубывающая функция, для которой определена обратная $g^{-1}(x)$, то

$$F_{g(\xi)}(x) = P(\xi < g^{-1}(x)) = F_{\xi}(g^{-1}(x)).$$

Полагая $g = F_{\xi}$, получим отсюда, в частности, что если F_{ξ} непрерывна, то случайная величина $\eta = F_{\xi}(\xi)$ распределена равномерно на $[0, 1]$. Наоборот, пусть η — равномерно распределенная случайная величина и F — заданная непрерывная функция распределения. Тогда случайная величина $\xi = F^{-1}(\eta)$ будет иметь функцию распределения $F(x)$. Мы получили, таким образом, важный способ построения случайных величин с наперед заданными распределениями с помощью случайных величин, распределенных равномерно.

В другом частном случае, когда $g(x) = a + bx$, $b > 0$, получаем $F_{g(\xi)}(x) = F_{\xi}\left(\frac{x-a}{b}\right)$. Этим соотношением мы уже в какой-то мере пользовались при рассмотрении распределений Φ_{α, σ^2} и $K_{\alpha, \sigma}$.

Если функция g строго возрастает и дифференцируема (в этом случае определена g^{-1}), а ξ имеет плотность $f(x)$, то существует плотность для $g(\xi)$, равная

$$f_{g(\xi)}(y) = f(g^{-1}(y))(g^{-1}(y))' = f(x) \frac{dx}{dy},$$

где $x = g^{-1}(y)$, $y = g(x)$. Аналогичное рассмотрение для убывающих g приводит к общей формуле

$$f_{g(\xi)}(y) = f(x) \left| \frac{dx}{dy} \right|.$$

³⁾ См. § 5 в Приложении 3.

При $g(x) = a + bx$, $b \neq 0$, получаем

$$f_{a+bx}(y) = \frac{1}{|b|} f\left(\frac{y-a}{b}\right).$$

§ 3. Многомерные случайные величины

Пусть $\xi_1, \xi_2, \dots, \xi_n$ — случайные величины, заданные на вероятностном пространстве $\langle \Omega, \mathfrak{F}, P \rangle$. Каждому ω эти случайные величины ставят в соответствие n -мерный вектор $\xi(\omega) = (\xi_1(\omega), \xi_2(\omega), \dots, \xi_n(\omega))$.

Определение 6. Отображение $\Omega \rightarrow \mathbb{R}^n$, задаваемое случайными величинами $\xi_1, \xi_2, \dots, \xi_n$, называется *случайным вектором* или *многомерной случайной величиной*.

Отображение $\Omega \rightarrow \mathbb{R}^n$ следует рассматривать как измеримое отображение $\langle \Omega, \mathfrak{F} \rangle$ на пространство $\langle \mathbb{R}^n, \mathfrak{B}^n \rangle$, где \mathfrak{B}^n есть σ -алгебра борелевских множеств в \mathbb{R}^n . Поэтому для любого борелевского множества B определена функция $P_\xi(B) = P(\xi \in B)$.

Определение 7. Функция $P_\xi(B)$ называется *распределением вектора* ξ .

Функция

$$F_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n) = P(\xi_1 < x_1, \dots, \xi_n < x_n)$$

называется *функцией распределения* случайного вектора (ξ_1, \dots, ξ_n) или *функцией совместного распределения* случайных величин ξ_1, \dots, ξ_n .

Приведем некоторые свойства функций распределения случайного вектора, которые называются свойствами *согласованности*.

$$\text{FF1. } \lim_{x_n \rightarrow \infty} F_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n) = F_{\xi_1, \dots, \xi_{n-1}}(x_1, \dots, x_{n-1}).$$

$$\text{FF2. } \lim_{x_n \rightarrow -\infty} F_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n) = 0.$$

То, что пределы рассматриваются по последнему аргументу, несущественно, так как всегда можно перенумеровать случайные величины.

Доказываются эти свойства так же, как свойство F2 функций распределения одномерной случайной величины. Их можно дополнить еще двумя свойствами, подобными F1 и F2, и тогда каждая функция $F(x_1, \dots, x_n)$ с набором таких свойств будет функцией распределения некоторой случайной величины.

Как и в одномерном случае, рассматривая случайные векторы $\xi = (\xi_1, \dots, \xi_n)$, мы можем пользоваться простейшей *выборочной* моделью вероятностного пространства $\langle \Omega, \mathfrak{F}, P \rangle$. Именно, пусть Ω совпадает с \mathbb{R}^n , а σ -алгебра $\mathfrak{F} = \mathfrak{B}^n$ составлена из борелевских множеств. Мы закончим построение требуемого вероятностного пространства, если для любого $B \in \mathfrak{B}^n$ положим $P(B) = P_\xi(B) = P(\xi \in B)$. Остается в качестве случайной величины принять само значение элементарного события, т. е. считать, что $\xi(\omega) = \omega$, где ω есть точка в \mathbb{R}^n .

Нетрудно видеть, что функция распределения F_{ξ_1, \dots, ξ_n} однозначно определяет распределение $P_\xi(B)$. Действительно, F_{ξ_1, \dots, ξ_n} определяет вероятность на алгебре \mathcal{A} , порожденной прямоугольниками $\{a_i \leq x_i < b_i, i = 1, \dots, n\}$. Например, в двумерном случае

$$\begin{aligned} P(a_1 \leq \xi_1 < b_1, a_2 \leq \xi_2 < b_2) &= \\ &= P(\xi_1 < b_1, a_2 \leq \xi_2 < b_2) - P(\xi_1 < a_1, a_2 \leq \xi_2 < b_2) = \\ &= [F_{\xi_1, \xi_2}(b_1, b_2) - F_{\xi_1, \xi_2}(b_1, a_2)] - [F_{\xi_1, \xi_2}(a_1, b_2) - F_{\xi_1, \xi_2}(a_1, a_2)]. \end{aligned}$$

Но $\mathfrak{B}^n = \sigma(\mathcal{A})$, и нам остается воспользоваться теоремой о продолжении меры (см. § 2.1).

Таким образом, по функции распределения $F_{\xi_1, \dots, \xi_n} = F$ мы всегда можем построить выборочное вероятностное пространство $(\mathbb{R}^n, \mathfrak{B}, P_\xi)$ и случайную величину $\xi(\omega) = \omega$ на нем, которая будет иметь заданное распределение F_ξ .

Так же, как и в одномерном случае, мы отнесем распределение случайного вектора к дискретному типу, если случайный вектор принимает не более чем счетное число значений. Распределение случайного вектора будет абсолютно непрерывного типа, если для любого борелевского множества $B \subset \mathbb{R}^n$

$$P_\xi(B) = P(\xi \in B) = \int_B f(x) dx,$$

где, очевидно, $f(x) \geq 0$, $\int_{\Omega} f(x) dx = 1$.

Это определение можно заменить на эквивалентное, потребовав, чтобы

$$F_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} f(t_1, \dots, t_n) dt_1 \dots dt_n. \quad (5)$$

Действительно, если имеет место (5), то мы определим счетно-аддитивную функцию множеств $Q(B) = \int_B f(x) dx$ (см. свойства интеграла в Приложении 3),

которая на прямоугольниках совпадает с $P_\xi(B)$. Следовательно, $P_\xi(B) = Q(B)$.

Функция $f(x)$ называется *плотностью распределения* ξ или *плотностью совместного распределения* ξ_1, \dots, ξ_n . Для нее почти всюду имеет место равенство

$$\frac{\partial^n F_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n)}{\partial x_1 \dots \partial x_n} = f(x_1, \dots, x_n).$$

Если вектор ξ имеет плотность $f(x_1, \dots, x_n)$, то, очевидно, любой «подвектор» $(\xi_{k_1}, \dots, \xi_{k_s})$, $k_i \leq n$, также имеет плотность, равную (пусть для простоты $k_i = i$, $i = 1, \dots, s$)

$$f(x_1, \dots, x_s) = \int f(x_1, \dots, x_n) dx_{s+1} \dots dx_n.$$

Пусть в некоторой области $A \subset \mathbb{R}^n$ заданы непрерывно дифференцируемые функции $y_i = g_i(x_1, \dots, x_n)$, которые однозначно разрешимы относительно x_1, \dots, x_n : $x_i = g_i^{-1}(y_1, \dots, y_n)$, при этом якобиан $J = |\partial x_i / \partial y_j| \neq 0$ в A . Обозначим через B образ A в пространстве значений (y_1, \dots, y_n) . Пусть, далее, случайный вектор $\xi = (\xi_1, \dots, \xi_n)$ имеет плотность $f_\xi(x)$. Тогда $\eta_i = g_i(\xi_1, \dots, \xi_n)$ будут случайными величинами с плотностью совместного распределения в точке $(y_1, \dots, y_n) \in B$, равной

$$f_\eta(y_1, \dots, y_n) = f_\xi(x_1, \dots, x_n) |J|; \quad (6)$$

кроме того,

$$\begin{aligned} P(\xi \in A) &= \int_A f_\xi(x_1, \dots, x_n) dx_1 \dots dx_n = \int_B f_\xi(x_1, \dots, x_n) |J| dy_1 \dots dy_n = \\ &= \int_B f_\eta(y_1, \dots, y_n) dy_1 \dots dy_n = P(\eta \in B). \end{aligned} \quad (7)$$

Это есть, очевидно, обобщение на многомерный случай свойства плотности, рассмотренного в конце § 2. Формулы (7) для интегралов хорошо известны в классическом анализе как формулы о замене переменной в интеграле и могут служить доказательством (6).

Рассмотрим два важных примера многомерных распределений (мы продолжаем список наиболее употребительных распределений, начатый в § 2).

9. *Полиномиальное распределение* \mathbf{B}_p^n . Мы используем здесь тот же символ \mathbf{B}_p^n , что и для биномиального распределения, с той лишь разницей, что под p будем понимать здесь вектор $p = (p_1, \dots, p_r)$, $p_j \geq 0$, $\sum_{j=1}^r p_j = 1$, который можно интерпретировать как набор вероятностей r несовместных событий A_j , $\bigcup A_j = \Omega$. Для целочисленного случайного вектора $\nu = (\nu_1, \dots, \nu_r)$ будем писать $\nu \in \mathbf{B}_p^n$, если для $k = (k_1, \dots, k_r)$, $k_j \geq 0$, $\sum_{j=1}^r k_j = n$, справедливо

$$P(\nu = k) = \frac{n!}{k_1! \dots k_r!} p_1^{k_1} \dots p_r^{k_r}. \quad (8)$$

Справа здесь стоит элемент разложения полинома $(p_1 + \dots + p_r)^n$ по степеням p_1, \dots, p_r . Это объясняет название распределения. Если p — число, то, очевидно, $\mathbf{B}_p^n = \mathbf{B}_{(p, 1-p)}^n$, так что биномиальное распределение есть полиномиальное при $r = 2$.

Числа ν_j можно интерпретировать как частоты появлений событий A_j после n независимых испытаний (вероятность появления A_j в каждом испытании равна p_j). Действительно, вероятность любой фиксированной последовательности исходов, содержащей соответственно k_1, \dots, k_r исходов A_1, \dots, A_r , равна $p_1^{k_1} \dots p_r^{k_r}$, а число различных таких последовательностей равно $n! / k_1! \dots k_r!$ (из $n!$ перестановок надо оставить те, которые отличаются не только перестановками элементов внутри групп k_1, \dots, k_r). Получим в результате вероятность (8).

Пример 4. Простейшей вероятностной моделью турнира между двумя шахматистами можно считать следующую. В каждой партии независимо от исходов предыдущих партий первый игрок выигрывает с вероятностью p , проигрывает с вероятностью q и делает ничью с вероятностью $1 - p - q$. В этом случае вероятность того, что в n партиях первый игрок одержит i побед и потерпит j поражений ($i + j \leq n$), будет равна

$$p(n; i, j) = \frac{n!}{i!j!(n-i-j)!} p^i q^j (1-p-q)^{n-i-j}.$$

Допустим, что турнир продолжается до тех пор, пока один из двух игроков не наберет N выигрышей (ему и присуждается победа). Если обозначить через η продолжительность турнира (число партий до его окончания), то

$$P(\eta = n) = \sum_{i=0}^{N-1} p(n-1; N-1, i) p + \sum_{i=0}^{N-1} p(n-1; i, N-1) q.$$

10. *Многомерное нормальное распределение* Φ_{α, σ^2} . Пусть α есть вектор $\alpha = (\alpha_1, \dots, \alpha_r)$ и $\sigma^2 = \|\sigma_{ij}\|$, $i, j = 1, \dots, r$, есть симметричная положительно определенная матрица, $A = \|\alpha_{ij}\|$ — матрица, обратная к $\sigma^2 = A^{-1}$. Мы будем говорить,

что вектор $\xi = (\xi_1, \dots, \xi_r)$ имеет нормальное распределение $\xi \in \Phi_{\alpha, \sigma^2}$, если это распределение имеет плотность

$$\varphi_{\alpha, \sigma^2}(x) = \frac{\sqrt{|A|}}{(2\pi)^{r/2}} \exp \left\{ -\frac{1}{2}(x - \alpha)A(x - \alpha)^T \right\}.$$

Здесь T означает операцию транспонирования:

$$xAx^T = \sum a_{ij}x_i x_j.$$

Нетрудно проверить, что

$$\int \varphi_{\alpha, \sigma^2}(x) dx_1 \dots dx_r = 1$$

(см. также § 6.6).

§ 4. Независимость случайных величин и классов событий

1. Независимость случайных величин.

Определение 8. Случайные величины ξ_1, \dots, ξ_n называются *независимыми*, если

$$P(\xi_1 \in B_1, \dots, \xi_n \in B_n) = P(\xi_1 \in B_1) \dots P(\xi_n \in B_n) \quad (9)$$

для любых борелевских множеств B_1, \dots, B_n на прямой.

Можно ввести понятие *последовательности* независимых случайных величин. Случайные величины последовательности $\{\xi_n\}_{n=1}^{\infty}$, заданные на вероятностном пространстве $(\Omega, \mathfrak{F}, P)$, *независимы*, если для любого конечного числа n выполняется (9), так что независимость последовательности случайных величин сводится к независимости любого конечного набора случайных величин из этой последовательности. Как увидим позже, для последовательности независимых случайных величин будут независимыми любые два события, относящиеся к непересекающимся группам случайных величин из этой последовательности.

Другое возможное определение независимости случайных величин вытекает из следующего утверждения.

Теорема 2. Случайные величины ξ_1, \dots, ξ_n независимы тогда и только тогда, когда

$$F_{\xi_1 \dots \xi_n}(x_1, \dots, x_n) = F_{\xi_1}(x_1) \dots F_{\xi_n}(x_n).$$

Доказательство этой теоремы отнесено нами в п. 3 этого параграфа.

Если распределение $\xi = (\xi_1, \dots, \xi_n)$ абсолютно непрерывно, то важный критерий независимости ξ_j содержится в теореме 3.

Теорема 3. Пусть случайные величины ξ_1, \dots, ξ_n имеют плотности $f_1(x), \dots, f_n(x)$ соответственно. Тогда для независимости ξ_1, \dots, ξ_n необходимо и достаточно, чтобы существовала плотность $f(x_1, \dots, x_n)$ вектора $\xi = (\xi_1, \dots, \xi_n)$, равная

$$f(x_1, \dots, x_n) = f_1(x_1) \dots f_n(x_n).$$

Таким образом, если окажется, что плотность ξ равна произведению плотностей ξ_j , то это будет означать независимость ξ_j .

Мы предоставляем читателю, пользуясь этой теоремой, убедиться, что компоненты нормального вектора (ξ_1, \dots, ξ_n) независимы тогда и только тогда, когда $a_{ij} = 0$, $\sigma_{ij} = 0$ при $i \neq j$.

Доказательство теоремы 3. Если функция распределения случайной величины ξ_i равна

$$F_{\xi_i}(x_i) = \int_{-\infty}^{x_i} f_i(t_i) dt_i$$

и ξ_i независимы, то совместная функция распределения будет определяться формулой

$$\begin{aligned} F_{\xi_1 \dots \xi_n}(x_1, \dots, x_n) &= F_{\xi_1}(x_1) \dots F_{\xi_n}(x_n) = \\ &= \int_{-\infty}^{x_1} f_1(t_1) dt_1 \dots \int_{-\infty}^{x_n} f_n(t_n) dt_n = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} f_1(t_1) \dots f_n(t_n) dt_1 \dots dt_n. \end{aligned}$$

Обратно, предположив, что

$$F_{\xi_1 \dots \xi_n}(x_1, \dots, x_n) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} f_1(t_1) \dots f_n(t_n) dt_1 \dots dt_n,$$

получим равенство

$$F_{\xi_1 \dots \xi_n}(x_1, \dots, x_n) = F_{\xi_1}(x_1) \dots F_{\xi_n}(x_n).$$

Рассмотрим теперь дискретный случай. Пусть для простоты компоненты ξ_j могут принимать лишь целочисленные значения. Тогда для независимости ξ_j необходимо и достаточно, чтобы при всех k_1, \dots, k_n

$$P(\xi_1 = k_1, \dots, \xi_n = k_n) = P(\xi_1 = k_1) \dots P(\xi_n = k_n).$$

Проверка этого утверждения не вызывает затруднений, и мы оставляем ее читателю.

Понятие независимости является весьма важным в теории вероятностей и будет использоваться на протяжении всего курса. Допустим, что мы формализуем какую-нибудь практическую задачу (строим подходящую вероятностную модель, в которой присутствуют различные случайные величины). Как установить, будут ли эти случайные величины (или события) независимыми? В такого рода ситуациях *оправдано правило считать независимыми события и случайные величины, не связанные причинно.*

В связи с этим выявление «вероятностной» независимости в математической модели случайного явления часто связано с глубоким проникновением в его физическую сущность.

Рассмотрим некоторые простые примеры. Известно, например, что вероятность новорожденного оказаться мальчиком имеет довольно устойчивое значение $P(A) = 22/43$. Если B означает условие, что рождение происходит в день соединения Юпитера с Марсом, то в предположении, что расположение планет не определяет индивидуальных судеб людей, условная вероятность $P(A|B)$ будет иметь то же самое значение: $P(A|B) = 22/43$, т. е. фактический подсчет частоты рождения мальчиков при таких специальных астрологических условиях привел бы именно к частоте $22/43$. Хотя такой подсчет в достаточно обширных размерах, возможно, никем не проводился, нет оснований сомневаться в его результатах.

Естественно, что связь между «математической» и причинной независимостью не следует абсолютизировать. Например, в силу закона всемирного тяготения несомненно, что полет одного артиллерийского снаряда оказывает некоторое

влияние на происходящий одновременно полет другого снаряда. Но очевидно, что на практике с этим влиянием можно не считаться. Сказанное поясняет также, что независимость событий и величин в том конкретном и относительном понимании этого термина, которое каждый раз имеет место, никак не противоречит принципу всеобщей связи всех явлений.

Любопытно отметить также, что формальное определение независимости событий или случайных величин значительно шире понятия реальной независимости в смысле принадлежности к причинно не связанным явлениям. Это следует из того, что «математическая» независимость может иметь место и в тех случаях, когда нет никаких оснований предполагать отсутствие причинной связи. Сказанное можно проиллюстрировать с помощью следующего примера. Пусть η есть равномерно распределенная на $[0, 1]$ случайная величина. Тогда в разложении η в двоичную дробь

$$\eta = \frac{\xi_1}{2} + \frac{\xi_2}{4} + \frac{\xi_3}{8} + \dots$$

случайные величины ξ_k будут независимыми (см. пример 10.1), хотя все они связаны по происхождению.

Понятно, что это обстоятельство лишь расширяет сферу применения всех утверждений, которые получены нами ниже при выполнении формального условия независимости⁴⁾.

С понятием независимости случайных величин тесно связано понятие *независимости σ -алгебр*.

2. Независимость классов событий. Пусть $\langle \Omega, \mathfrak{F}, P \rangle$ — вероятностное пространство, а \mathcal{A}_1 и \mathcal{A}_2 — классы событий, содержащиеся в σ -алгебре \mathfrak{F} .

Определение 9. Классы событий \mathcal{A}_1 и \mathcal{A}_2 называются *независимыми*, если для любых событий A_1 и A_2 таких, что $A_1 \in \mathcal{A}_1$ и $A_2 \in \mathcal{A}_2$, имеет место равенство

$$P(A_1 A_2) = P(A_1)P(A_2).$$

Следующее определение вводит понятие независимости последовательности событий.

Определение 10. Классы событий $\{\mathcal{A}_n\}_{n=1}^{\infty}$ *независимы*, если для каждого набора целых чисел n_1, \dots, n_k

$$P\left(\bigcap_{j=1}^k A_{n_j}\right) = \prod_{j=1}^k P(A_{n_j})$$

при любых $A_{n_j} \in \mathcal{A}_{n_j}$.

Например, в последовательности независимых испытаний σ -подалгебры событий, относящихся к различным испытаниям, будут независимы. Независимость *последовательности алгебр событий* также сводится к независимости любого конечного набора алгебр из этой последовательности. Ясно, что подалгебры событий независимых алгебр событий также независимы.

Теорема 4. σ -алгебры \mathcal{A}_1 и \mathcal{A}_2 , порожденные соответственно независимыми алгебрами событий \mathcal{A}_1 и \mathcal{A}_2 , независимы.

⁴⁾ Более подробное обсуждение связи причинной и вероятностной можно найти у А. Н. Колмогорова в книге [13], из которой заимствованы приведенные здесь примеры.

Прежде чем доказывать это утверждение, получим полезную для дальнейшего теорему об аппроксимации, в силу которой любое событие A из σ -алгебры \mathfrak{A} , порожденной некоторой алгеброй \mathcal{A} , можно приближать в известном смысле событиями из \mathcal{A} . Чтобы быть более точными, введем в рассмотрение «расстояние» между событиями

$$d(A, B) = P(\overline{AB} \cup \overline{AB}) = P(\overline{AB}) + P(\overline{AB}) = P(A - B) + P(B - A).$$

Оно обладает свойствами

$$\begin{aligned} d(\overline{A}, \overline{B}) &= d(A, B), \\ d(A, C) &\leq d(A, B) + d(B, C), \\ d(AB, CD) &\leq d(A, C) + d(B, D), \\ |P(A) - P(B)| &\leq d(A, B). \end{aligned} \tag{10}$$

Первое из этих соотношений очевидно. Неравенство треугольника вытекает из того, что

$$\begin{aligned} d(A, C) &= P(\overline{AC}) + P(\overline{AC}) = P(\overline{ACB}) + P(\overline{ACB}) + P(\overline{ACB}) + P(\overline{ACB}) \leq \\ &\leq P(\overline{CB}) + P(\overline{AB}) + P(\overline{AB}) + P(\overline{CB}) = d(A, B) + d(B, C). \end{aligned}$$

Третье соотношение в (10) получается аналогично путем укрупнения событий под знаком вероятности. Наконец, последнее неравенство в (10) есть следствие того, что

$$P(A) = P(AB) + P(\overline{AB}) = P(B) - P(B\overline{A}) + P(\overline{AB}).$$

Теорема 5 (теорема аппроксимации). Пусть $\langle \Omega, \mathfrak{F}, P \rangle$ — вероятностное пространство, и \mathfrak{A} есть σ -алгебра, порожденная некоторой алгеброй \mathcal{A} событий из \mathfrak{F} . Тогда для любого $A \in \mathfrak{A}$ существует последовательность $A_n \in \mathcal{A}$, обладающая свойством

$$\lim_{n \rightarrow \infty} d(A, A_n) = 0. \tag{11}$$

В силу последнего неравенства в (10) утверждение теоремы означает, что $P(A) = \lim_{n \rightarrow \infty} P(A_n)$, и каждое событие $A \in \mathfrak{A}$ может быть с точностью до множества нулевой вероятности представлено как предел последовательности событий из порождающей алгебры \mathcal{A} (см. также Приложение 1).

*Доказательство*⁵⁾. Событие $A \in \mathfrak{F}$ назовем *аппроксимируемым*, если для него существует последовательность $A_n \in \mathcal{A}$, обладающая свойством (11), т. е. $d(A_n, A) \rightarrow 0$.

Так как $d(A, A) = 0$, то совокупность аппроксимируемых событий \mathfrak{A}^* содержит \mathcal{A} . Поэтому для доказательства теоремы достаточно убедиться, что \mathfrak{A}^* есть σ -алгебра.

Тот факт, что \mathfrak{A}^* образует алгебру, очевиден, поскольку соотношения $A \in \mathfrak{A}^*$, $B \in \mathfrak{A}^*$ влекут за собой, что \overline{A} , $A \cup B$, $A \cap B \in \mathfrak{A}^*$. (Например, если $d(A, A_n) \rightarrow 0$, $d(B, B_n) \rightarrow 0$, то в силу третьего неравенства в (10) $d(AB, A_n B_n) \leq d(A, A_n) + d(B, B_n) \rightarrow 0$, $AB \in \mathfrak{A}^*$.)

Пусть теперь

$$C = \bigcup_{k=1}^{\infty} C_k, \quad C_k \in \mathfrak{A}^*.$$

⁵⁾ Эта теорема является также прямым следствием леммы из Приложения 1.

Так как \mathfrak{A}^* — алгебра, то $D_n = \bigcup_{k=1}^n C_k \in \mathfrak{A}^*$; при этом

$$d(D_n, C) = P(C - D_n) = P(C) - P(D_n) \rightarrow 0.$$

Поэтому можно выбрать $A_n \in \mathfrak{A}$ так, что $d(D_n, A_n) < 1/n$ и, следовательно, в силу (10)

$$d(C, A_n) \leq d(C, D_n) + d(D_n, A_n) \rightarrow 0.$$

Таким образом, $C \in \mathfrak{A}^*$ и \mathfrak{A}^* образует σ -алгебру. ■

Доказательство теоремы 4 получить теперь нетрудно. Если $A_1 \in \mathfrak{A}_1$ и $A_2 \in \mathfrak{A}_2$, то по теореме 5 существуют последовательности $A_{1n} \in \mathfrak{A}_1$ и $A_{2n} \in \mathfrak{A}_2$ такие, что $d(A_1, A_{1n}) \rightarrow 0$ при $n \rightarrow \infty$, $i = 1, 2$. Обозначив $B = A_1 A_2$, $B_n = A_{1n} A_{2n}$, получим

$$d(B, B_n) \leq d(A_i, A_{in}) + d(A_2, A_{2n}) \rightarrow 0$$

при $n \rightarrow \infty$,

$$P(A_1 A_2) = \lim_{n \rightarrow \infty} P(B_n) = \lim_{n \rightarrow \infty} P(A_{1n}) P(A_{2n}) = P(A_1) P(A_2).$$

3. Связь введенных понятий. Нам понадобится еще следующее определение. Пусть ξ — случайная величина (или вектор), определенная на вероятностном пространстве $\langle \Omega, \mathfrak{F}, P \rangle$.

Определение 11. Класс \mathfrak{F}_ξ множеств из \mathfrak{F} мы назовем σ -алгеброй, порожденной случайной величиной ξ , если он образован событиями вида $A = \xi^{-1}(B) = \{\omega : \xi(\omega) \in B\}$, где B — борелевские множества.

То, что \mathfrak{F}_ξ образует σ -алгебру, очевидно, поскольку каждой операции над множествами A соответствует такая же операция над множествами $B = \xi(A)$, образующими σ -алгебру.

σ -алгебру \mathfrak{F}_ξ , порожденную случайной величиной ξ , будем обозначать также символом $\sigma(\xi)$.

Рассмотрим, например, вероятностное пространство $\langle \Omega, \mathfrak{B}, P \rangle$, где $\Omega = \mathbb{R}$ есть вещественная прямая, а \mathfrak{B} — σ -алгебра борелевских множеств. Если

$$\xi = \xi(\omega) = \begin{cases} 0, & \omega < 0, \\ 1, & \omega \geq 0, \end{cases}$$

то \mathfrak{F}_ξ , очевидно, будет состоять из четырех множеств: \mathbb{R} , \emptyset , $\{\omega < 0\}$, $\{\omega \geq 0\}$. Более «тонких» событий так определенная случайная величина ξ «выделять» из \mathfrak{B} не может. В то же время очевидно, что ξ будет измерима ($\{\xi \in B\} \in \mathfrak{B}_1$) относительно любой другой более «богатой» σ -подалгебры \mathfrak{B}_1 такой, что $\sigma(\xi) \subset \mathfrak{B}_1 \subset \mathfrak{B}$.

Если $\xi = \xi(\omega) = [\omega]$ есть целая часть ω , то \mathfrak{F}_ξ будет σ -алгеброй множеств, составленных из событий $\{k \leq \omega < k+1\}$, $k = \dots, -1, 0, 1, \dots$.

Если, наконец, $\xi(\omega) = \varphi(\omega)$, где φ непрерывна и монотонна, $\varphi(\infty) = \infty$, $\varphi(-\infty) = -\infty$, то \mathfrak{F}_ξ будет совпадать с σ -алгеброй борелевских множеств \mathfrak{B} .

Лемма 1. Пусть на $\langle \Omega, \mathfrak{F}, P \rangle$ заданы две случайные величины ξ и η , величина ξ является измеримой относительно $\sigma(\eta)$. Тогда ξ и η функционально связаны, т. е. существует борелевская функция g такая, что $\xi = g(\eta)$.

Доказательство. По условию

$$A_{k,n} = \left\{ \xi \in \left[\frac{k}{2^n}, \frac{k+1}{2^n} \right) \right\} \in \sigma(\eta).$$

Обозначим через $B_{k,n} = \{\eta(\omega) : \omega \in A_{k,n}\}$ образы множеств $A_{k,n}$ на прямой \mathbb{R} при отображении $\eta(\omega)$ и положим $g_n(x) = k/2^n$ при $x \in B_{k,n}$. Тогда $g_n(\eta) = [2^n \xi]/2^n$, и поскольку $A_{k,n} \in \sigma(\eta)$, то $B_{k,n} \in \mathfrak{B}$ и g_n есть борелевская функция. Так как $g_n(x) \uparrow$ при каждом x , то существует $\lim_{n \rightarrow \infty} g_n(x) = g(x)$, который также является борелевской функцией. Остается заметить, что по самому построению $\xi = \lim_{n \rightarrow \infty} g_n(\eta) = g(\eta)$. ■

Теперь сформулируем очевидное предложение, связывающее независимость случайных величин и σ -алгебр.

Случайные величины ξ_1, \dots, ξ_n независимы тогда и только тогда, когда независимы σ -алгебры $\sigma(\xi_1), \dots, \sigma(\xi_n)$.

Это есть прямое следствие определения независимости случайных величин σ -алгебр.

Мы можем получить теперь доказательство теоремы 2. Заметим предварительно, что конечные объединения полуинтервалов вида $[\cdot, \cdot)$ (возможно, с бесконечно удаленными концами) образуют алгебру \mathcal{A} , которая порождает борелевскую σ -алгебру \mathfrak{B} на прямой: $\mathfrak{B} = \sigma(\mathcal{A})$.

Доказательство теоремы 2. Так как в одну сторону утверждение теоремы очевидно, то достаточно убедиться, что равенство для совместной функции распределения $F(x_1, \dots, x_n) = F_{\xi_1}(x_1) \dots F_{\xi_n}(x_n)$ влечет за собой независимость $\sigma(\xi_1), \dots, \sigma(\xi_n)$. Положим для простоты $n = 2$ и обозначим через Δ и Λ соответственно полуинтервалы $[x_1, x_2)$ и $[y_1, y_2)$. Тогда будут справедливы равенства

$$\begin{aligned} P(\xi_1 \in \Delta, \xi_2 \in \Lambda) &= P(\xi_1 \in [x_1, x_2), \xi_2 \in [y_1, y_2)) = \\ &= F(x_2, y_2) - F(x_1, y_2) - F(x_2, y_1) + F(x_1, y_1) = \\ &= (F_{\xi_1}(x_2) - F_{\xi_1}(x_1))(F_{\xi_2}(y_2) - F_{\xi_2}(y_1)) = P\{\xi_1 \in \Delta\}P\{\xi_2 \in \Lambda\}. \end{aligned}$$

Следовательно, если Δ_i , $i = 1, \dots, n$, и Λ_j , $j = 1, \dots, m$, — две системы непересекающихся интервалов, то

$$\begin{aligned} P\left(\xi_1 \in \bigcup_{i=1}^n \Delta_i, \xi_2 \in \bigcup_{j=1}^m \Lambda_j\right) &= \sum_{i,j} P(\xi_1 \in \Delta_i, \xi_2 \in \Lambda_j) = \\ &= \sum_{i,j} P(\xi_1 \in \Delta_i)P(\xi_2 \in \Lambda_j) = P\left(\xi_1 \in \bigcup_{i=1}^n \Delta_i\right)P\left(\xi_2 \in \bigcup_{j=1}^m \Lambda_j\right). \end{aligned} \quad (12)$$

Но класс событий $\{\omega : \xi(\omega) \in A\} = \xi^{-1}(A)$, где $A \in \mathcal{A}$, вместе с \mathcal{A} образует алгебру (обозначим ее $\alpha(\xi)$) и при этом $\sigma(\alpha(\xi)) = \sigma(\xi)$. Мы доказали в (12), что $\alpha(\xi_1)$ и $\alpha(\xi_2)$ независимы. Следовательно, по теореме 4 будут независимы $\sigma(\xi_1) = \sigma(\alpha(\xi_1))$ и $\sigma(\xi_2) = \sigma(\alpha(\xi_2))$. ■

Сформулируем теперь в виде теоремы следующий факт.

Теорема 6. Пусть φ_1, φ_2 — борелевские функции и ξ_1 и ξ_2 — независимые случайные величины, тогда $\eta_1 = \varphi_1(\xi_1)$ и $\eta_2 = \varphi_2(\xi_2)$ — также независимые случайные величины.

Доказательство. Нам надо убедиться, что

$$P(\varphi_1(\xi_1) \in B_1, \varphi_2(\xi_2) \in B_2) = P(\varphi_1(\xi_1) \in B_1)P(\varphi_2(\xi_2) \in B_2). \quad (13)$$

Но множества $\{x: \varphi_i(x) \in B_i\} = \varphi_i^{-1}(B_i) = B_i^*$ снова являются борелевскими. Поэтому

$$\{\omega: \varphi_i(\xi_i) \in B_i\} = \{\omega: \xi_i \in B_i^*\},$$

и требуемая мультипликативность вероятности (13) следует из независимости ξ_j . ■

Пусть $\{\xi_n\}_{n=1}^\infty$ — последовательность независимых случайных величин. Рассмотрим случайные величины $\xi_k, \xi_{k+1}, \dots, \xi_m$, где $k < m \leq \infty$. Пусть $\sigma(\xi_k, \dots, \xi_m)$ (при $m = \infty$ мы будем писать $\sigma(\xi_k, \xi_{k+1}, \dots)$) означает σ -алгебру событий, порожденную событиями $\bigcap_{j=k}^m A_j$, где $A_i \in \sigma(\xi_i)$.

Определение 12. σ -алгебра $\sigma(\xi_k, \dots, \xi_m)$ называется порожденной случайными величинами ξ_k, \dots, ξ_m .

Для дальнейшего нам понадобится следующее предложение.

Теорема 7. Для любого $k \geq 1$ σ -алгебра $\sigma(\xi_{n+k})$ не зависит от $\sigma(\xi_1, \dots, \xi_n)$.

Доказательство. Чтобы доказать это утверждение, воспользуемся теоремой 4. Для этого мы должны убедиться, что алгебра \mathcal{A} , порожденная множествами вида $B = \bigcup_{i=1}^n A_i$, где $A_i \in \sigma(\xi_i)$, не зависит от $\sigma(\xi_{n+k})$. Пусть $A \in \sigma(\xi_{n+k})$, тогда из независимости σ -алгебр $\sigma(\xi_1), \sigma(\xi_2), \dots, \sigma(\xi_n), \sigma(\xi_{n+k})$ следует

$$P(AB) = P(A)P(A_1) \dots P(A_n) = P(A) \cdot P(B).$$

Аналогично убеждаемся, что

$$P\left(\bigcup_{i=1}^n A_i A\right) = P\left(\bigcup_{i=1}^n A_i\right) P(A)$$

(надо лишь $\bigcup_{i=1}^n A_i$ представить в виде объединения непересекающихся множеств из \mathcal{A}). Таким образом, алгебра \mathcal{A} не зависит от $\sigma(\xi_{n+k})$. Следовательно, $\sigma(\xi_1, \dots, \xi_n)$ и $\sigma(\xi_{n+k})$ независимы. ■

Нетрудно понять, что заключения, аналогичные предыдущим, можно сделать и относительно векторных случайных величин ξ_1, ξ_2, \dots , определив их независимость с помощью соотношений

$$P(\xi_1 \in B_1, \dots, \xi_n \in B_n) = \prod P(\xi_j \in B_j),$$

где B_j — борелевские множества в пространствах соответствующей размерности.

В заключение этого параграфа отметим, что вероятностное пространство (Ω, \mathcal{F}, P) $((\mathbb{R}^n, \mathcal{B}^n, P_\xi))$, на котором заданы независимые случайные величины ξ_1, \dots, ξ_n с данными распределениями F_{ξ_j} компонент ξ_j , всегда можно построить, если эти распределения F_{ξ_j} известны. Это следует непосредственно из § 3, так как в нашем случае функция совместного распределения $F_\xi(x_1, \dots, x_n)$ вектора $\xi = (\xi_1, \dots, \xi_n)$ однозначно определяется функциями распределения $F_{\xi_j}(x)$ величины ξ_j

$$F_\xi(x_1, \dots, x_n) = \prod_{j=1}^n F_{\xi_j}(x_j).$$

§ 5*. О бесконечных последовательностях случайных величин

Бесконечные последовательности случайных величин уже упоминались. Такие последовательности в дальнейшем не раз будут предметом наших рассуждений. Однако возникает вопрос, всегда ли можно на вероятностном пространстве задать бесконечную последовательность так, чтобы ее координаты обладали наперед заданными свойствами (например, были одинаково распределены и независимы).

Как мы видели, *конечную* последовательность независимых случайных величин задать всегда можно, рассмотрев для «составной» случайной величины (ξ_1, \dots, ξ_n) выборочное вероятностное пространство $\mathbb{R}_1 \times \mathbb{R}_2 \times \dots \times \mathbb{R}_n = \mathbb{R}^n$ и σ -алгебру $\mathfrak{B}_1 \times \mathfrak{B}_2 \times \dots \times \mathfrak{B}_n = \mathfrak{B}$, порожденную множествами вида $B_1 \times B_2 \times \dots \times B_n \subset \mathbb{R}^n$, где B_i — борелевские множества. На алгебре этих множеств и достаточно задать вероятность. Однако в бесконечномерном случае дело обстоит сложнее. Здесь уже недостаточно теоремы 1 (см. § 2) и ее обобщения на многомерный случай.

Надо суметь задать вероятность на такой алгебре событий из $\mathbb{R}^\infty = \prod_{k=1}^{\infty} \mathbb{R}^k$, замыкание которой относительно счетного числа операций \cup и \cap образовывало бы σ -алгебру \mathfrak{B}^∞ , порожденную произведениями $\cap B_{j_k}$, $B_{j_k} \in \mathfrak{B}_{j_k}$.

Пусть N — некоторое подмножество целых чисел. Обозначим через $\mathbb{R}^N = \prod_{k \in N} \mathbb{R}_k$ прямое произведение пространств \mathbb{R}_k по $k \in N$, $\mathfrak{B}^N = \prod_{k \in N} \mathfrak{B}_k$. Мы будем говорить, что распределения $P_{N'}$ и $P_{N''}$ на $(\mathbb{R}^{N'}, \mathfrak{B}^{N'})$ и $(\mathbb{R}^{N''}, \mathfrak{B}^{N''})$ *согласованы*, если меры на пересечении $\mathbb{R}^N = \mathbb{R}^{N'} \cap \mathbb{R}^{N''}$ (для \mathbb{R}^N здесь $N = N' \cap N''$), которые индуцированы мерами $P_{N'}$ и $P_{N''}$, совпадают. Эти меры на \mathbb{R}^N называются *проекциями* соответственно $P_{N'}$ и $P_{N''}$ на \mathbb{R}^N . Ответ на поставленный выше вопрос о существовании бесконечной последовательности случайных величин содержится в следующей теореме Колмогорова (доказательство ее приведено в Приложении 2).

Теорема Колмогорова. *Задание на конечномерных пространствах \mathbb{R}^N согласованных распределений P_N определяет на $(\mathbb{R}^\infty, \mathfrak{B}^\infty)$ такую единственную вероятностную меру P_∞ , что каждая вероятность P_N есть проекция P_∞ на \mathbb{R}^N .*

Из этой теоремы следует, в частности, что на соответствующем пространстве всегда можно задать бесконечную последовательность произвольных независимых случайных величин. Ведь прямые произведения мер, заданных на $\mathbb{R}_1, \mathbb{R}_2, \dots$, для различных произведений $\mathbb{R}^{N'}$ и $\mathbb{R}^{N''}$, очевидно, всегда согласованы.

§ 6. Интегралы

1. Интеграл по мере. Как уже отмечалось, задание вероятностного пространства означает задание конечной счетно-аддитивной меры. Это делает возможным рассмотрение интегралов по введенной мере

$$\int g(\xi(\omega)) P(d\omega) \quad (14)$$

по множеству Ω для борелевской функции g и для любой случайной величины ξ на $(\Omega, \mathfrak{F}, P)$ (напомним, что $g(x)$ называется борелевской, если $\{x: g(x) < t\}$ при любом t есть борелевское множество на прямой).

Определение, конструкция и основные свойства интеграла по мере предполагаются известными читателю. Если читатель чувствует себя недостаточно подготовленным, то мы рекомендуем ему познакомиться с Приложением 3, где содержатся все необходимые сведения. Но читатель может и не делать этого, если он готов ограничить себя в дальнейшем только рассмотрением дискретных и абсолютно непрерывных распределений, для которых интегралы превращаются в суммы или в обычные римановы интегралы. Для дальнейшего полезно также знание интеграла Стильтьеса; комментарии по этому поводу см. в п. 2.

Мы уже знаем, что случайная величина $\xi(\omega)$ индуцирует меру P_ξ на прямой, задаваемую равенством

$$P_\xi([x, y)) = P(x \leq \xi < y) = F_\xi(y) - F_\xi(x).$$

С помощью этой меры интеграл (14) можно записать в виде

$$\int g(\xi(\omega))P(d\omega) = \int g(x)P_\xi(dx).$$

Это равенство есть результат подстановки $x = \xi(\omega)$. Доказать его можно, выписав определения обоих интегралов. Интеграл в правой части может быть записан также в форме

$$\int g(x) dF_\xi(x). \quad (15)$$

В таком виде он называется *интегралом Лебега—Стилтьеса* от функции $g(x)$ по распределению F_ξ .

2. Интеграл Стильтьеса. Интеграл (15) часто называется просто интегралом Стильтьеса, как, впрочем, и интеграл Римана—Стилтьеса, который определяется несколько иначе и для более узкого класса функций $g(x)$.

Если $g(x)$ — непрерывная функция, то интеграл Лебега—Стилтьеса совпадает с интегралом Римана—Стилтьеса, равным по определению

$$\int g(x) dF(x) = \lim_{\substack{b \rightarrow \infty \\ a \rightarrow -\infty}} \lim_{N \rightarrow \infty} \sum_{k=0}^N g(\bar{x}_k) [F(x_{k+1}) - F(x_k)], \quad (16)$$

где предел в правой части не зависит от разбиения x_0, x_1, \dots, x_N полуинтервала $[a, b)$ и выбора точек $\bar{x}_k \in \Delta_k = [x_k, x_{k+1})$. Разбиение x_0, x_1, \dots, x_N (свое для каждого N) обладает тем свойством, что $\max_k (x_{k+1} - x_k) \rightarrow 0$ при $N \rightarrow \infty$.

Действительно, как мы знаем (см. Приложение 3), интеграл Лебега—Стилтьеса есть

$$\int g(x) dF(x) = \lim_{\substack{b \rightarrow \infty \\ a \rightarrow -\infty}} \lim_{N \rightarrow \infty} \int_a^b g_N(x) P_\xi(dx), \quad (17)$$

где g_N — любая последовательность простых функций (принимających конечное число значений), монотонно сходящихся к $g(x)$. Из приведенных определений видно, что достаточно доказать совпадение интегралов \int_a^b в конечных пределах. Так как интеграл Лебега—Стилтьеса $\int_a^b g dF$ от непрерывной функции g всегда существует, то в его определении можно в качестве g_N взять две последовательности

простых функций g_N^* и g_N^{**} , постоянных на полуинтервалах Δ_k и равных на них соответственно

$$g^*(x_k) = \sup_{x \in \Delta_k} g(x) \quad \text{и} \quad g^{**}(x_k) = \inf_{x \in \Delta_k} g(x).$$

Обе последовательности в (17), построенные по g_N^* и g_N^{**} , будут, очевидно, монотонно сходиться с разных сторон к одному и тому же пределу, равному интегралу Лебега—Стилтьеса $\int_a^b g(x) dF(x)$. Но для любых $\tilde{x}_k \in \Delta_k$ будет выполняться

$$g^{**}(x_k) \leq g(\tilde{x}_k) \leq g^*(x_k),$$

и, следовательно, интегральная сумма в (1) будет заключена в границах

$$\int_a^b g_N^{**} dF(x) \leq \sum_{k=0}^N g(\tilde{x}_k) [F(x_{k+1}) - F(x_k)] \leq \int_a^b g_N^*(x) dF(x).$$

Эти неравенства и доказывают, очевидно, требуемое совпадение.

Нетрудно проверить, что такое же совпадение (17) и (16) будет и в том случае, если $F(x)$ непрерывна, а $g(x)$ есть функция ограниченной вариации, при этом

$$\int_a^b g(x) dF(x) = g(x)F(x) \Big|_a^b - \int_a^b F(x) dg(x).$$

Пользуясь здесь этим обстоятельством, мы можем определение интеграла Римана—Стилтьеса распространить и на случай, когда $g(x)$ есть функция ограниченной вариации, а $F(x)$ — произвольная функция распределения. Действительно, пусть $F(x) = F_H(x) + F_D(x)$ есть представление F в виде суммы непрерывной и дискретной компонент, и y_1, y_2, \dots есть точки скачков функции $F_D(x)$:

$$p_k = F_D(y_k + 0) - F_D(y_k) > 0.$$

Тогда следует по определению положить

$$\int g(x) dF(x) = \sum p_k g(y_k) + \int g(x) dF_H(x),$$

где интеграл Римана—Стилтьеса $\int g dF_H(x)$ можно понимать уже, как мы отмечали, в смысле определения (16).

Как общепринято, будем говорить, что $\int g dF$ существует, если $\int |g| dF$ конечен. Из определения интеграла Стилтьеса легко обнаружить, что для ступенчатых $F(x)$ (распределение дискретно) интеграл превращается в сумму

$$\int g(x) dF(x) = \sum_k g(x_k) (F(x_k + 0) - F(x_k)) = \sum_k g(x_k) P(\xi = x_k),$$

где x_1, x_2, \dots — точки скачков $F(x)$. Если же $F(x) = \int_{-\infty}^x p(x) dx$ абсолютно непрерывна, а $p(x)$ и $g(x)$ интегрируемы по Риману, то интеграл Стилтьеса

$$\int g(x) dF(x) = \int g(x) p(x) dx$$

превращается в обычный интеграл Римана.

Еще раз отметим, что читателю, плохо знакомому с техникой интеграла Стильеса и интегрированием по мере, можно продолжить чтение книги, имея в виду лишь эти две интерпретации интеграла. Этого вполне достаточно для понимания существа изложения. К тому же большинство практически важных распределений относится именно к этим типам — дискретному или абсолютно непрерывному.

Напомним некоторые другие свойства интеграла Стильеса (непосредственно вытекающие из определений (17) или (16)):

$$\begin{aligned}\int_a^b dF &= F(b) - F(a); \\ \int_a^b g dF &= \int_a^c g dF + \int_c^b g dF; \\ \int (g_1 + g_2) dF &= \int g_1 dF + \int g_2 dF; \\ \int cg dF &= c \int g dF \quad \text{при } c = \text{const}; \\ \int_a^b g dF &= gF|_a^b - \int_a^b F dg,\end{aligned}$$

если g — функция ограниченной вариации.

3. Интегралы от многомерных случайных величин. Распределение суммы независимых случайных величин. Интегралы по мере (14) сохраняют свой смысл и для многомерных $\xi(\omega) = (\xi_1(\omega), \dots, \xi_n(\omega))$ (этого нельзя сказать об интегралах Римана—Стильеса (16)). Имеется в виду интеграл

$$\int_{\Omega} g(\xi_1(\omega), \dots, \xi_n(\omega)) P(d\omega), \quad (18)$$

где g есть борелевская функция, отображающая \mathbb{R}^n в \mathbb{R} .

Если $(\mathbb{R}^n, \mathfrak{B}^n, P_\xi)$ — выборочное вероятностное пространство для ξ , то интеграл (18) можно записать в виде

$$\int_{\mathbb{R}^n} g(x) P_\xi(dx), \quad x = (x_1, \dots, x_n) \in \mathbb{R}^n.$$

Обратимся теперь к случаю, когда компоненты ξ_1, \dots, ξ_n вектора ξ независимы, и пусть сначала $n = 2$. Для множеств

$$B = B_1 \times B_2 = \{(x_1, x_2): x_1 \in B_1, x_2 \in B_2\} \subset \mathbb{R}^2,$$

где B_1 и B_2 принадлежат \mathbb{R} и измеримы, выполняется равенство

$$P(\xi \in B) = P(\xi_1 \in B_1, \xi_2 \in B_2) = P(\xi_1 \in B_1)P(\xi_2 \in B_2). \quad (19)$$

В этом случае говорят, что мера $P_{\xi_1, \xi_2}(dx_1, dx_2) = P(\xi_1 \in dx_1, \xi_2 \in dx_2)$ в \mathbb{R}^2 , соответствующая (ξ_1, ξ_2) , является *прямым произведением мер*

$$P_{\xi_1}(dx_1) = P(\xi_1 \in dx_1) \quad \text{и} \quad P_{\xi_2}(dx_2) = P(\xi_2 \in dx_2).$$

Как мы уже знаем, равенство (19) однозначно определяет меру в $\langle \mathbb{R}^2, \mathfrak{B}^2 \rangle$ по заданным распределениям ξ_1 и ξ_2 в $\langle \mathbb{R}, \mathfrak{B} \rangle$. При этом оказывается, что интеграл

$$\int g(x_1, x_2) P_{\xi_1 \xi_2}(dx_1, dx_2) \quad (20)$$

по мере $P_{\xi_1 \xi_2}$ можно выразить через интегралы по мерам P_{ξ_1} и P_{ξ_2} . Именно, справедлива следующая теорема Фубини (доказательство ее см. в Приложении 3 или см. свойство 5А в § 4.8).

Теорема 8 (теорема о повторном интегрировании). Для борелевской функции $g(x, y) \geq 0$ и для независимых ξ_1 и ξ_2

$$\int g(x_1, x_2) P_{\xi_1 \xi_2}(dx_1, dx_2) = \int \left[\int g(x_1, x_2) P_{\xi_2}(dx_2) \right] P_{\xi_1}(dx_1). \quad (21)$$

Если $g(x, y)$ разноточна, то для выполнения равенства (21) требуется существование интеграла в левой части (21). Индексы 1 и 2 в правой части можно менять местами.

В Приложении 3 показано, что измеримость $g(x, y)$ влечет измеримость подынтегральных функций в правой части (21).

Следствие 1. Пусть $g(x_1, x_2) = g_1(x_1)g_2(x_2)$. Тогда, если выполнено хотя бы одно из трех условий:

- 1) $g_1 \geq 0, g_2 \geq 0$,
- 2) $\int g_1(x_1)g_2(x_2)P_{\xi_1 \xi_2}(dx_1, dx_2)$ существует,
- 3) $g_j(x_j)P_{\xi_j}(dx_j), j = 1, 2$, существуют,

то

$$\int g_1(x_1)g_2(x_2)P_{\xi_1 \xi_2}(dx_1, dx_2) = \int g_1(x_1)P_{\xi_1}(dx_1) \int g_2(x_2)P_{\xi_2}(dx_2). \quad (22)$$

Чтобы избежать тривиальных затруднений, предполагаем, что $P(g_j(\xi_j) = 0) \neq 1, j = 1, 2$.

Доказательство. При выполнении первых двух условий утверждение следствия немедленно вытекает из теоремы Фубини. Для произвольных g_1, g_2 положим $g_j = g_j^+ - g_j^-$, $g_j^\pm \geq 0, j = 1, 2$. Если $\int g_j^\pm dP_{\xi_j} < \infty$ (будем использовать здесь эту сокращенную запись интегралов), то

$$\begin{aligned} \int g_1 g_2 dP_{\xi_1} dP_{\xi_2} &= \int g_1^+ g_2^+ dP_{\xi_1} dP_{\xi_2} - \\ &- \int g_1^+ g_2^- dP_{\xi_1} dP_{\xi_2} - \int g_1^- g_2^+ dP_{\xi_1} dP_{\xi_2} + \int g_1^- g_2^- dP_{\xi_1} dP_{\xi_2} = \\ &= \int g_1^+ dP_{\xi_1} \int g_2^+ dP_{\xi_2} - \int g_1^+ dP_{\xi_1} \int g_2^- dP_{\xi_2} - \\ &- \int g_1^- dP_{\xi_1} \int g_2^+ dP_{\xi_2} + \int g_1^- dP_{\xi_1} \int g_2^- dP_{\xi_2} = \int g_1 dP_{\xi_1} \int g_2 dP_{\xi_2}. \end{aligned}$$

Следствие 2. В частном случае, когда $g(x_1, x_2) = I_B(x_1, x_2)$ есть индикатор множества $B \in \mathfrak{B}^2$, получаем формулу о последовательном вычислении меры множества B :

$$P((\xi_1, \xi_2) \in B) = \int P((x_1, \xi_2) \in B) P_{\xi_1}(dx_1).$$

Вероятность события $\{(x_1, \xi_2) \in B\}$ можно записать также в виде $P(\xi_2 \in B_{x_1}) = P_{\xi_2}(B_{x_1})$, где $B_{x_1} = \{x_2: (x_1, x_2) \in B\}$ есть «сечение» множества B в точке x_1 .

Если $B = \{(x_1, x_2): x_1 + x_2 < x\}$, мы получим

$$\begin{aligned} P((\xi_1, \xi_2) \in B) &= P(\xi_1 + \xi_2 < x) \equiv F_{\xi_1 + \xi_2}(x) = \\ &= \int P(x_1 + \xi_2 < x) P_{\xi_1}(dx_1) = \int F_{\xi_2}(x - x_1) dF_{\xi_1}(x_1). \end{aligned} \quad (23)$$

Мы получили формулу для функции распределения суммы независимых случайных величин, выражающую $F_{\xi_1 + \xi_2}$ через F_{ξ_1} и F_{ξ_2} . Интеграл в правой части (23) называется *сверткой* функций распределения $F_{\xi_1}(x)$ и $F_{\xi_2}(x)$ и обозначается $F_{\xi_1} * F_{\xi_2}(x)$. Таким же образом можно получить равенство

$$P(\xi_1 + \xi_2 < x) = \int_{-\infty}^{\infty} F_{\xi_1}(x - t) dF_{\xi_2}(t).$$

Правую часть здесь можно рассматривать также как результат интегрирования $\int dF_{\xi_1}(t) F_{\xi_2}(x - t)$ по частям.

Если хотя бы одна функция распределения имеет плотность, то и свертка имеет плотность. Это непосредственно следует из формул для свертки. Пусть, например,

$F_{\xi_2}(x) = \int_{-\infty}^x f_{\xi_2}(u) du$. Тогда

$$F_{\xi_1 + \xi_2}(x) = \int_{-\infty}^x P_{\xi_1}(dt) \int_{-\infty}^x f_{\xi_2}(u - t) du = \int_{-\infty}^x \left(\int_{-\infty}^{\infty} P_{\xi_1}(dt) f_{\xi_2}(x - t) \right) du,$$

так что плотность $f(x)$ суммы $\xi_1 + \xi_2$ равна

$$f(x) = \int_{-\infty}^{\infty} P_{\xi_1}(dt) f_{\xi_2}(x - t) = \int_{-\infty}^{\infty} f_{\xi_2}(x - t) dF_{\xi_1}(t).$$

Пример 5. Пусть ξ_1, ξ_2, \dots независимы и равномерно распределены на $[0, 1]$, т. е. ξ_1, ξ_2, \dots имеют одну и ту же функцию распределения с плотностью

$$f(x) = \begin{cases} 1, & x \in [0, 1], \\ 0, & x \notin [0, 1]. \end{cases} \quad (24)$$

Тогда плотность суммы $\xi_1 + \xi_2$ равна

$$f_{\xi_1 + \xi_2}(x) = \int_0^1 f(x - t) dt = \begin{cases} 0, & x \notin [0, 2], \\ x, & x \in [0, 1], \\ 2 - x, & x \in [1, 2]. \end{cases} \quad (25)$$

Присутствующий здесь интеграл равен, очевидно, длине пересечения отрезков $[0, 1]$ и $[x - 1, x]$. Плотность суммы $\xi_1 + \xi_2 + \xi_3$ будет уже склеена из трех кусков парабол:

$$f_{\xi_1+\xi_2+\xi_3}(x) = \int_0^1 f_{\xi_1+\xi_2}(x-t) dt = \begin{cases} 0, & x \notin [0, 3], \\ \frac{x^2}{2}, & x \in [0, 1], \\ 1 - \frac{(2-x)^2}{2} - \frac{(x-1)^2}{2}, & x \in [1, 2], \\ \frac{(3-x)^2}{2}, & x \in [2, 3]. \end{cases}$$

Вычисление этого интервала становится наглядным, если воспользоваться рис. 4, а, где заштрихованные площади соответствуют значениям $f_{\xi_1+\xi_2+\xi_3}(x)$ при разных x . Вид плотностей для ξ_1 , $\xi_1 + \xi_2$, $\xi_1 + \xi_2 + \xi_3$ представлен на рис. 4, б. Плотность суммы $\xi_1 + \xi_2 + \xi_3 + \xi_4$ будет склеена уже из четырех кусков кривых третьего порядка и т.д. Если начало координат перенести в точку $n/2$, то с точностью до преобразования сжатия внешний вид плотности суммы $\xi_1 + \dots + \xi_n$ с ростом n все больше будет напоминать график функции e^{-x^2} . Мы увидим дальше, что это обстоятельство не является случайным.

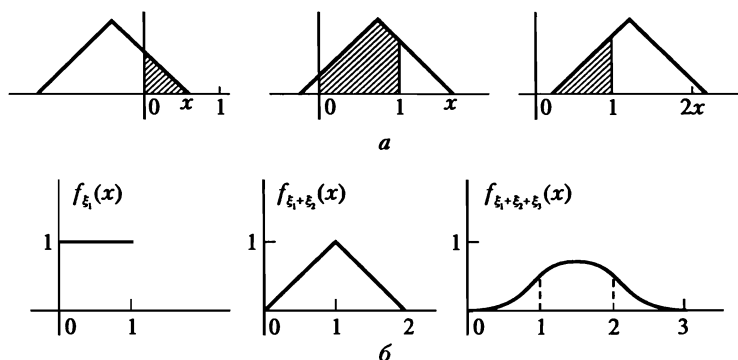


Рис. 4.

В связи с рассмотренным примером можно заметить, что если есть две независимые случайные величины ξ и η , где ξ имеет функцию распределения $F(x)$, а η распределена равномерно на $[0, 1]$, то плотность суммы $\xi + \eta$ в точке x равна

$$f_{\xi+\eta}(x) = \int dF(t) f_{\eta}(x-t) = \int_{x-1}^x dF(t) = F(x) - F(x-1).$$

Г л а в а 4

Числовые характеристики случайных величин

§ 1. Математическое ожидание

Определение 1. Математическим ожиданием или средним значением случайной величины ξ , заданной на вероятностном пространстве $\langle \Omega, \mathfrak{F}, P \rangle$, называется число

$$E\xi = \int_{\Omega} \xi(\omega) P(d\omega),$$

если интеграл, стоящий в правой части, существует.

В силу сделанных выше замечаний (см. § 3.6) мы можем $E\xi$ определить также как

$$E\xi = \int x P_{\xi}(dx) = \int x dF(x), \quad (1)$$

где $F(x)$ — функция распределения ξ . Из определения следует, что $E\xi$ существует, если $E|\xi| < \infty$. Как нетрудно видеть, $E\xi$ не существует, если, например, $1 - F(x) > 1/x$ при всех достаточно больших x .

Мы уже знаем, что если $F(x)$ — ступенчатая функция, то интеграл Стильеса в (1) превратится в сумму

$$E\xi = \sum_k x_k P(\xi = x_k).$$

Если $F(x)$ имеет плотность $f(x)$, то

$$E\xi = \int x f(x) dx,$$

так что $E\xi$ есть координата «центра тяжести» распределения F единичной массы на прямой и соответствует естественному пониманию среднего значения этого распределения.

Если $g(x)$ — борелевская функция, то $\eta = g(\xi)$ есть снова случайная величина и

$$Eg(\xi) = \int g(\xi(\omega)) P(d\omega) = \int g(x) dF(x) = \int x dF_{g(\xi)}(x).$$

Последнее равенство следует из определения (1).

Основные свойства математических ожиданий совпадают со следующими свойствами интеграла.

Е1. Если a и b постоянные, то $E(a + b\xi) = a + bE\xi$.

Е2. $E(\xi_1 + \xi_2) = E\xi_1 + E\xi_2$, если существуют какие-нибудь два участвующих в равенстве математических ожидания.

Е3. Если $a \leq \xi \leq b$, то $a \leq E\xi \leq b$. Всегда $E\xi \leq E|\xi|$.

Е4. Если $\xi \geq 0$ и $E\xi = 0$, то $\xi = 0$ с вероятностью 1.

Е5. Вероятность события A может быть записана в терминах математических ожиданий с помощью равенства

$$P(A) = EI(A),$$

где $I(A)$ есть случайная величина, равная индикатору события A : $I(A) = 1$, если $\omega \in A$, и $I(A) = 0$ в противном случае.

Дальнейшие свойства математических ожиданий см. в Приложении 3.

Рассмотрим несколько примеров.

Пример 1. Математические ожидания, связанные со схемой Бернулли. Пусть $\xi \in \mathbf{B}_p$, т.е. ξ принимает два значения: 0 с вероятностью q и 1 с вероятностью p , где $p + q = 1$. Тогда

$$E\xi = 0 \cdot P(\xi = 0) + 1 \cdot P(\xi = 1) = p.$$

Рассмотрим теперь последовательность исходов с схеме Бернулли до момента η появления первого «успеха». Другими словами, рассмотрим последовательность независимых величин ξ_1, ξ_2, \dots , распределенных как ξ , до момента времени

$$\eta = \min\{k \geq 1: \xi_k = 1\}.$$

Очевидно, что η является случайной величиной,

$$P(\eta = k) = q^{k-1}p, \quad k \geq 1,$$

так что $\eta - 1$ имеет геометрическое распределение. Следовательно,

$$E\eta = \sum_{k=1}^{\infty} k q^{k-1} p = \frac{p}{(1-q)^2} = \frac{1}{p}.$$

Если обозначить $S_n = \sum_{k=1}^n \xi_k$, то, очевидно, $ES_n = np$. Определим теперь при некотором целом $N \geq 1$ случайную величину $\eta = \min\{k \geq 1: S_k = N\}$ как «время» достижения последовательностью S_n уровня N . Имеем

$$P(\eta = k) = P(S_{k-1} = N - 1)p,$$

$$E\eta = p \sum_{k=N}^{\infty} k \binom{k-1}{N-1} p^{N-1} q^{k-N} = \frac{p^N}{(N-1)!} \sum_{k=N}^{\infty} k(k-1) \dots (k-N+1) q^{k-N}.$$

Сумма здесь равна производной N -го порядка в точке $z = q$ от функции $\psi(z) = \sum_{k=0}^{\infty} z^k = \frac{1}{1-z}$, т.е. равна $\frac{N!}{p^{N+1}}$. Итак, $E\eta = \frac{N}{p}$. Как мы увидим ниже, это равенство может быть получено как очевидное следствие результатов § 4.

Пример 2. Если $\xi \in \Phi_{\alpha, \sigma^2}$, то

$$\begin{aligned} E\xi &= \int t \varphi_{\alpha, \sigma^2}(t) dt = \int t \frac{1}{\sigma\sqrt{2\pi}} e^{-(t-\alpha)^2/2\sigma^2} dt = \\ &= \frac{1}{\sigma\sqrt{2\pi}} \int (t - \alpha) e^{-(t-\alpha)^2/2\sigma^2} dt + \frac{\alpha}{\sigma\sqrt{2\pi}} \int e^{-(t-\alpha)^2/2\sigma^2} dt = \\ &= \frac{1}{\sigma\sqrt{2\pi}} \int z e^{-z^2/2\sigma^2} dz + \alpha = \alpha. \end{aligned}$$

Таким образом, параметр a нормального закона равен его математическому ожиданию.

Пример 3. Если $\xi \in \Pi_\lambda$, то $E\xi = \lambda$. Действительно,

$$E\xi = \sum_{k=0}^{\infty} k \frac{\mu^k}{k!} e^{-\mu} = \mu e^{-\mu} \sum_{k=1}^{\infty} \frac{\mu^{k-1}}{(k-1)!} = \mu.$$

Пример 4. Если $\xi \in U_{0,1}$, то

$$E\xi = \int_0^1 x dx = \frac{1}{2}.$$

Из свойства E1 следует, что для $\xi \in U_{a,b}$

$$E\xi = a + \frac{b-a}{2} = \frac{a+b}{2}.$$

Если $\xi \in K_{0,1}$, то $E\xi$ не существует. Это следует из того, что интеграл $\int \frac{x dx}{1+x^2}$ расходится.

Рассмотрим теперь пример, близкий по типу к примеру 1 о подсчете $E\eta$, но более сложный, связанный с вычислением среднего значения продолжительности шахматного турнира в реальной ситуации. В примере 3.4 была описана простейшая вероятностная модель шахматного турнира. Первый игрок в каждой партии независимо от исходов предыдущих партий выигрывает с вероятностью p , проигрывает с вероятностью q , $p+q < 1$, и делает ничью с вероятностью $1-p-q$. Конечно, это весьма грубое первое приближение, поскольку независимость на самом деле, по-видимому, отсутствует. С другой стороны, трудно предположить, что для уравновешенных игроков высокого класса названные вероятности сильно менялись бы от партии к партии или в зависимости от предыдущих результатов. Более сложная модель, учитывающая зависимость p и q от исходов предыдущих партий, будет рассмотрена в примере 12.6.

Допустим, что турнир продолжается до тех пор, пока один из двух участников не наберет N выигранных (ему и присуждается победа). Именно по такой схеме при $N=6$ был организован матч на первенство мира по шахматам в 1984 году между А. Карповым и Г. Каспаровым. Что можно сказать в этом случае о математическом ожидании $E\eta$ времени η продолжительности матча?

В примере 3.4 было показано, что

$$P(\eta = n) = p \sum_{i=0}^{N-1} p(n-1; N-1, i) + q \sum_{i=0}^{N-1} p(n-1; i, N-1),$$

где $p(n; i, j) = \frac{n!}{i!j!(n-i-j)!} p^i q^j (1-p-q)^{n-i-j}$. Поэтому при очевидных соглашениях относительно индексов суммирования

$$E\eta = \frac{1}{(N-1)!} \sum_{i=0}^{N-1} \frac{p^N q^i + p^i q^N}{i!} \sum_{n=0}^{\infty} n(n-1) \dots (n-i-N+1) (1-p-q)^{n-N}$$

Сумма по n подсчитана в примере 1 и равна $\frac{(N+i)!}{(p+q)^{N+i+1}}$. Следовательно,

$$\begin{aligned} E\eta &= \frac{N}{p+q} \sum_{i=0}^{N-1} \frac{(p^N q^i + p^i q^N)(N+i)!}{i! N! (p+q)^{i+N}} = \\ &= \frac{N}{p+q} \sum_{i=0}^{N-1} \binom{N+i}{i} [r^N (1-r)^i + r^i (1-r)^N], \end{aligned}$$

где $r = p/(p+q)$.

В интервью газете «Известия» от 3 марта 1985 г. Карпов сообщил, что в среднем в предматчевых турнирах он проигрывал одну партию из двадцати. Такими же показателями, по его словам, характеризуется и Каспаров. Если в этих условиях применительно к нашей простейшей модели положить $p = q = 1/20$ (строго говоря, такое заключение из приведенных данных сделать нельзя; соотношение $p = q = 1/20$ скорее можно рассматривать как одну из возможных гипотез), то при $N = 6$ ($r = 1/2$) путем непосредственного подсчета получим

$$E\eta = \frac{15}{8} \left[1 + 21 \left(1 + \frac{5}{8} + \frac{11}{16} \right) \right] \approx 93.$$

Таким образом, если верить в адекватность нашей простейшей модели, то ожидаемая продолжительность турнира оказывается очень большой. Некоторым подтверждением правильности сделанных предположений может служить тот факт, что матч между Карповым и Каспаровым был прерван по решению президента шахматной федерации на 48-й партии ввиду того, что этот матч чрезмерно затянулся.

Если учитывать результаты проведенного турнира и последующих партий между Карповым и Каспаровым, то оценка (приближенное значение) чисел p и q может отличаться от $1/20$.

Для рассмотренной модели справедливо также следующее простое неравенство: $N/(p+q) < E\eta < (2N-1)/(p+q)$.

Оно вытекает из соотношения $\eta_N \leq \eta \leq \eta_{2N-1}$, где η_N — число партий до момента, когда суммарное число очков обоих игроков достигнет N . В силу примера 1 $E\eta_N = N/(p+q)$.

Пример 5. В задаче о ячейках в §§ 3, 4 в главе 1 рассматривалась вероятность события A , состоящего в том, что хотя бы одна из n ячеек, куда наудачу бросают r частиц, остается пустой. Найдем математическое ожидание числа $S_{n,r}$ пустых ячеек после r бросаний. Если A_k означает событие, состоящее в том, что k -я ячейка пуста, а $I(A_k)$ означает индикатор этого события, то

$$S_{n,r} = \sum_{k=1}^n I(A_k), \quad ES_{n,r} = \sum_{k=1}^n P(A_k) = n \left(1 - \frac{1}{n} \right)^r.$$

Заметим теперь, что $ES_{n,r}$ близко к 0, если $(1 - 1/n)^r$ мало по сравнению с $1/n$, т. е. когда $-r \ln(1 - 1/n) - \ln n$ велико. При больших n

$$\ln \left(1 - \frac{1}{n} \right) = -\frac{1}{n} + O \left(\frac{1}{n^2} \right)$$

и требуемое соотношение будет иметь место, если $(r - n \ln n)/n$ велико. Ясно, что в нашем случае (ср. со свойством E4) малость $ES_{n,r}$ будет влечь за собой малость $P(A) = P(S_{n,r} > 0)$.

§ 2. Условные функции распределения и условные математические ожидания

Пусть $\langle \Omega, \mathfrak{F}, P \rangle$ — вероятностное пространство, и событие $B \in \mathfrak{F}$ обладает свойством $P(B) > 0$. Образует новое вероятностное пространство $\langle \Omega, \mathfrak{F}, P_B \rangle$, где P_B для любого $A \in \mathfrak{F}$ зададим с помощью равенства

$$P_B(A) = P(A|B).$$

Легко проверить свойства вероятности P_1, P_2, P_3 для меры P_B . Пусть ξ — случайная величина на $\langle \Omega, \mathfrak{F}, P \rangle$. Это есть, очевидно, случайная величина и в пространстве $\langle \Omega, \mathfrak{F}, P_B \rangle$.

Определение 2. Математическое ожидание ξ в пространстве $\langle \Omega, \mathfrak{F}, P_B \rangle$ называется *условным относительно B математическим ожиданием* и обозначается через $E(\xi|B)$:

$$E(\xi|B) = \int_{\Omega} \xi(\omega) P_B(d\omega).$$

По определению меры P_B

$$E(\xi|B) = \int_{\Omega} \xi(\omega) P(d\omega|B) = \frac{1}{P(B)} \int_{\Omega} \xi(\omega) P(d\omega \cap B) = \frac{1}{P(B)} \int_B \xi(\omega) P(d\omega).$$

Последний интеграл отличается от $E\xi$ тем, что интегрирование в нем ведется лишь по множеству B . Этот интеграл будем обозначать

$$E(\xi; B) = \int_B \xi(\omega) P(d\omega),$$

так что

$$E(\xi|B) = \frac{1}{P(B)} E(\xi; B).$$

Нетрудно заметить, что функция

$$F(x|B) = P_B(\xi < x) = P(\xi < x|B)$$

является функцией распределения случайной величины ξ , рассматриваемой на $\langle \Omega, \mathfrak{F}, P_B \rangle$.

Определение 3. Функция $F(x|B)$ называется *условной функцией распределения ξ* (в условном пространстве $\langle \Omega, \mathfrak{F}, P \rangle$) *при условии B* .

Значение $E(\xi|B)$ можно записать, очевидно, также в виде $\int x dF(x|B)$.

Если σ -алгебра $\sigma(\xi)$, порожденная случайной величиной ξ , не зависит от события B , то $P_B(A) = P(A)$ для любого $A \in \sigma(\xi)$. Следовательно,

$$F(x|B) = F(x), \quad E(\xi|B) = E\xi, \quad E(\xi; B) = P(B)E(\xi). \quad (2)$$

Пусть $\{B_n\}$ — последовательность, быть может, конечная, непересекающихся событий такая, что $\bigcup B_n = \Omega$ и $P(B_n) > 0$ для любого n . Тогда

$$E\xi = \int_{\Omega} \xi(\omega) P(d\omega) = \sum_n \int_{B_n} \xi(\omega) P(d\omega) = \sum_n E(\xi; B_n) = \sum_n P(B_n) E(\xi|B_n). \quad (3)$$

Мы получили формулу полной вероятности для математических ожиданий. Эта формула бывает весьма полезной.

Пример 6. Пусть время службы какого-нибудь механизма есть случайная величина ξ с функцией распределения $F(x)$. Известно, что механизм проработал время a . Каково распределение оставшегося срока службы? Чему равно его математическое ожидание?

Очевидно, в этой задаче нам надо найти $P(\xi - a \geq x | \xi \geq a)$ и $E(\xi - a | \xi \geq a)$. Предполагается, конечно, что

$$P(a) = P(\xi \geq a) > 0.$$

В силу выписанных выше формул

$$P(\xi - a \geq x | \xi \geq a) = \frac{P(x + a)}{P(a)},$$

$$E(\xi - a | \xi \geq a) = \frac{1}{P(a)} \int_0^{\infty} x dF(x + a).$$

Интересно отметить следующее. Во многих прикладных задачах, особенно когда речь идет о работе сложных механизмов, состоящих из большого числа надежных деталей, распределение ξ можно считать *показательным*:

$$P(x) = P(\xi \geq x) = e^{-\mu x}, \quad \mu > 0.$$

(Причина этого явления станет ясной позже, при рассмотрении теоремы Пуассона и пуассоновских процессов. Примерами таких механизмов могут служить электронно-вычислительные устройства.) Но для показательного распределения оказывается, что распределение оставшегося срока службы механизма

$$P(\xi - a \geq x | \xi \geq a) = \frac{P(x + a)}{P(a)} = e^{-\mu x} = P(x). \quad (4)$$

совпадает с распределением срока службы нового прибора. Другими словами, новый прибор и прибор, проработавший непрерывно некоторое время a , с точки зрения длительности дальнейшей безотказной работы эквивалентны.

Нетрудно понять, что показательное распределение (вместе со своим дискретным аналогом $P(\xi = k) = q^k(1 - q)$, $k = 0, 1, \dots$) является единственным, которое обладает отмеченным выше замечательным свойством. Это видно из равенства (4), в силу которого с необходимостью

$$P(x + a) = P(x)P(a).$$

Пример 7. n агрегатов расположены так, что расстояние между i -м и j -м агрегатом равно a_{ij} , $1 \leq i, j \leq n$. Каждый агрегат время от времени нуждается в обслуживании (наладке, ремонте и т. п.). Предположим, что обслуживание производится одним рабочим и что каждый очередной вызов на обслуживание с вероятностью p_j ($\sum_{j=1}^n p_j = 1$) возникает у j -го агрегата. Если, например, в данный момент рабочий закончил обслуживание i -го агрегата, то с вероятностью p_j (от i не зависящей) он следующим будет обслуживать агрегат с номером j ; при этом он пройдет к нему путь длиной a_{ij} . Чему равна средняя длина пройденного пути за один переход?

Пусть событие B_i означает, что номер агрегата, обслуживание которого предшествовало данному переходу, равен i . Тогда $P(B_i) = p_i$ и вероятность того,

что рабочий от i -го агрегата направится к j -му, $j = 1, \dots, n$, равна p_j . Длина перехода при этом составит a_{ij} . Поэтому

$$E(\xi | B_i) = \sum_{j=1}^n p_j a_{ij},$$

и по формуле полной вероятности

$$E\xi = \sum_{i=1}^n P(B_i)E(\xi | B_i) = \sum_{i,j=1}^n p_i p_j a_{ij}.$$

Полученное выражение позволяет сравнивать различные варианты расположения агрегатов с точки зрения минимизации величины $E\xi$ при заданных ограничениях на a_{ij} . Например, если $a_{ij} \geq 1$ и все агрегаты однотипны ($p_j = 1/n$), то при их расположении вдоль одной прямой линии получим $a_{ij} = |j - i|$,

$$E\xi = \frac{1}{n^2} \sum_{i,j=1}^n |j - i| = \frac{2}{n^2} \sum_{k=1}^{n-1} k(n - k) = \frac{n-1}{3} \left(1 + \frac{1}{n}\right)^{1)},$$

так что $E\xi$ при больших n равно приблизительно $n/3$. Таким образом, если в день происходит s вызовов, то средняя длина суммарного перехода составит примерно $sn/3$. Нетрудно показать, что расположение агрегатов по окружности будет более выгодным, однако и оно не будет оптимальным.

Пример 8. Как уже отмечалось, не для всех случайных величин (распределений) существует математическое ожидание. И соответствующие примеры вовсе не являются патологическими — этим свойством обладает, например, распределение Коши $K_{\alpha, \sigma}$. Мы рассмотрим теперь одну задачу о случайных блужданиях, в которой также появляются случайные величины, не имеющие математических ожиданий. Это задача о так называемой *безобидной* игре. В игре участвуют два игрока. Начальный капитал первого игрока равен z . Каждую партию этот игрок выигрывает или проигрывает с вероятностью $1/2$ независимо от исходов предыдущих партий, при этом его капитал соответственно увеличивается или уменьшается на единицу. Пусть $z + S_k$ — капитал первого игрока после k -й партии, а $\eta(z)$ — число шагов до его разорения при игре с бесконечно богатым противником, когда z есть целое число, т. е.

$$\eta(z) = \min\{k: z + S_k = 0\}, \quad \eta(0) = 0.$$

Если $\inf_k S_k > -z$ (т. е. первый игрок никогда не разорится), мы положим $\eta(z) = \infty$.

Покажем прежде всего, что $\eta(z)$ — *собственная* случайная величина, т. е. случайная величина, с вероятностью 1 принимающая конечные значения. Это обстоятельство для первого игрока будет означать, что он с вероятностью 1 рано или поздно разорится, с каким бы капиталом z он ни начинал играть. В качестве Ω

¹⁾ Для подсчета этой суммы достаточно заметить, что

$$\sum_{k=1}^{n-1} k(k-1) = \frac{1}{3}(n-2)(n-1)n$$

(сравните начальные значения и приращения этих функций).

здесь можно взять «выборочное» пространство, состоящее из всевозможных последовательностей из 1 и -1 . Каждая такая последовательность ω будет описывать «траекторию» игры. (Например, -1 на k -м месте означает, что первый игрок проиграл k -ю партию.) Мы предоставляем читателю в виде упражнения построить здесь вероятностное пространство $\langle \Omega, \mathfrak{F}, \mathbf{P} \rangle$. Очевидно, что надо сделать так, чтобы вероятность первых n фиксированных исходов игры (первые n координат ω фиксированы) равнялась 2^{-n} .

Положим

$$u(z) = \mathbf{P}(\eta(z) < \infty), \quad u(0) = 1$$

и обозначим через B_1 событие, состоящее в том, что первая координата ω равна 1 (игрок выиграл), и B_2 — что эта координата равна -1 (проиграл). Замечая, что $\mathbf{P}(\eta(z) < \infty | B_1) = u(z+1)$ (если первая партия выиграна, то капитал стал равным $z+1$), мы по формуле полной вероятности при $z \geq 1$ получим

$$u(z) = \mathbf{P}(B_1)\mathbf{P}(\eta(z) < \infty | B_1) + \mathbf{P}(B_2)\mathbf{P}(\eta(z) < \infty | B_2) = \frac{1}{2}u(z+1) + \frac{1}{2}u(z-1).$$

Полагая $\delta(z) = u(z+1) - u(z)$, $z \geq 0$, мы находим отсюда $\delta(z) - \delta(z-1) = 0$, $\delta(z) = \delta = \text{const}$. Так как $u(z+1) = u(0) + \sum_{k=1}^z \delta(k) = u(0) + z\delta$, то ясно, что δ не может быть ничем иным, кроме 0, $u(z) \equiv 1$ при всех z .

Итак, в игре с бесконечно богатым противником игрок разоряется с вероятностью 1. Это объясняет в какой-то мере тот факт, что все азартные игроки (не останавливающиеся «вóвремя»; выбор этого «времени» — отдельная задача) рано или поздно кончают разорением. Даже если игра является безобидной.

Покажем теперь, что хотя $\eta(z)$ — конечная случайная величина, $\mathbf{E}\eta(z) = \infty$. Допустим противное, что

$$v(z) = \mathbf{E}\eta(z) < \infty.$$

Аналогично предыдущему замечаем, что $\mathbf{E}(\eta(z) | B_1) = 1 + v(z+1)$ (капитал стал равным $z+1$, и один ход уже сделан). Поэтому по формуле полной вероятности для математических ожиданий при $z \geq 1$ находим

$$v(z) = \frac{1}{2}(1 + v(z+1)) + \frac{1}{2}(1 + v(z-1)), \quad v(0) = 0.$$

Из этой формулы видно, что если $v(z) < \infty$, то $v(k) < \infty$ при всех k . Обозначим $\Delta(z) = v(z+1) - v(z)$. Тогда предыдущее равенство при $z \geq 1$ можно записать в виде

$$-1 = \frac{1}{2}\Delta(z) - \frac{1}{2}\Delta(z-1),$$

или

$$\Delta(z) = \Delta(z-1) - 2.$$

Из этого равенства находим $\Delta(z) = \Delta(0) - 2z$. Поэтому

$$v(z) = \sum_{k=0}^{z-1} \Delta(k) = z\Delta(0) - z(z-1) = zv(1) - z(z-1).$$

Отсюда следует, что $\mathbf{E}\eta(z) < 0$ для достаточно больших z . Но $\eta(z)$ — положительная случайная величина, и поэтому $\mathbf{E}\eta(z) \geq 0$. Полученное противоречие показывает, что предположение о конечности математического ожидания $\eta(z)$ неверно.

§ 3. Математические ожидания функций независимых случайных величин

Теорема 1. 1. Пусть ξ и η — независимые случайные величины, $g(x, y)$ — борелевская функция. Тогда, если $g \geq 0$ или если $Eg(\xi, \eta)$ конечно, то

$$Eg(\xi, \eta) = E[Eg(x, \eta)|_{x=\xi}]. \quad (5)$$

2. Пусть $g(x, y) = g_1(x)g_2(y)$. Если $g_1(\xi) \geq 0$, $g_2(\eta) \geq 0$ или если $Eg_1(\xi)$, $Eg_2(\eta)$ существуют, то

$$Eg(\xi, \eta) = Eg_1(\xi)Eg_2(\eta). \quad (6)$$

Для существования $Eg(\xi, \eta)$ необходимо и достаточно существования $Eg_1(\xi)$ и $Eg_2(\eta)$. (Случаи $P(g_1(\xi) = 0) = 1$ или $P(g_2(\eta) = 0) = 1$ во избежание тривиальных затруднений здесь исключаются.)

Доказательство. Первое утверждение теоремы есть переформулировка теоремы Фубини в терминах математических ожиданий. Первая часть второго утверждения следует непосредственно из следствия 3.1 теоремы Фубини. Так как $|g_1(\xi)| \geq 0$, $|g_2(\eta)| \geq 0$ и эти случайные величины независимы, то

$$E|g_1(\xi)g_2(\eta)| = E|g_1(\xi)|E|g_2(\eta)|.$$

Поскольку $E|g_1(\xi)| \neq 0$, $E|g_2(\eta)| \neq 0$, то отсюда следует последнее утверждение теоремы. ■

Замечание 1. Формулу (5) можно рассматривать как формулу полной вероятности для вычисления $Eg(\xi, \eta)$. Утверждение (6), не ограничивая общности, можно записать в виде

$$E\xi\eta = E\xi E\eta. \quad (7)$$

Чтобы получить отсюда (6), надо в качестве ξ взять $g_1(\xi)$, а в качестве η взять $g_2(\eta)$ — это снова будут независимые случайные величины.

Примеры использования теоремы 1 были приведены в § 3.6 и будут встречаться в дальнейшем.

Обратное утверждение к (6) или (7) неверно. Существуют зависимые случайные величины ξ и η такие, что

$$E\xi\eta = E\xi E\eta.$$

Пусть, например, ζ и ξ независимы, $E\xi = E\zeta = 0$. Положим $\eta = \xi\zeta$. Тогда ξ и η , очевидно, зависимы (за исключением некоторых тривиальных случаев, таких, например, когда $\xi = \text{const}$), но

$$E\xi\eta = E\xi^2\zeta = E\xi^2 E\zeta = 0 = E\xi E\eta.$$

§ 4. Математическое ожидание случайных величин, не зависящих от будущего

Пусть на $(\Omega, \mathfrak{F}, P)$ заданы последовательность $\{\xi_n\}_{n=1}^\infty$ независимых случайных величин и целочисленная случайная величина $\nu \geq 0$. Пусть $\mathfrak{F}_{k,n} = \sigma(\xi_k, \dots, \xi_n)$ означает σ -алгебру, порожденную $n - k + 1$ случайными величинами ξ_k, \dots, ξ_n .

Определение 4. Случайная величина ν называется не зависящей от будущего, если событие $\{\nu \leq n\}$ не зависит от $\mathfrak{F}_{n+1, \infty}$.

Случайная величина ν называется *марковской* или *моментом остановки*, если $\{\nu \leq n\} \in \mathfrak{F}_{1,n}$. Другими словами, в этом случае знание значений $\xi_1; \dots, \xi_n$ позволяет сказать, произошло событие $\{\nu \leq n\}$ или нет.

На основании предыдущего (см. § 5 в главе 3) ясно, что в нашем случае для марковской случайной величины событие $\{\nu \leq n\}$ (или $\{\nu > n\}$) и любое событие A из $\mathfrak{F}_{n+1,\infty}$ независимы. Это значит, что марковская величина для последовательности независимых ξ_k не зависит от будущего.

Пример 9. Пусть ν — номер первой случайной величины в последовательности $\{\xi_n\}_{n=1}^{\infty}$, которая оказалась больше или равной N , т. е.

$$\nu = \inf \{k: \xi_k \geq N\}.$$

Если ξ_n независимы, то, очевидно, ν не зависит от будущего, так как событие $\{\nu \leq n\} = \bigcup_{k=1}^n \{\xi_k \geq N\} \in \mathfrak{F}_{1,n}$.

Любая случайная величина ν , не зависящая от последовательности ξ_1, ξ_2, \dots (не зависят $\sigma(\nu)$ и $\mathfrak{F}_{1,\infty}$), является, очевидно, не зависящей от будущего случайной величиной.

Термин «марковская» станет несколько понятнее после введения понятия марковости в главе 12. Термин «момент остановки» связан с характером большого количества прикладных задач, в которых такие случайные величины появляются. В качестве довольно типичного примера мы можем рассмотреть следующую процедуру. Пусть ξ_k — количество брака в k -й партии продукции предприятия. Статистический приемочный контроль качества продукции состоит в следующем. Вся продукция бракуется, если при последовательной проверке партий обнаружится, что при некотором n значение суммы $S_n = \sum_{k=1}^n \xi_k$ превзошло заданный допустимый уровень $a + bn$. Номер партии ν , при которой это событие произошло,

$$\nu = \min \{n: S_n \geq a + bn\}$$

есть момент остановки всей процедуры проверки. Чтобы избежать длительной проверки, моментом остановки (в буквальном смысле) делают также и случайную величину

$$\nu^* = \min \{n: S_n \leq -A + Bn\}, \quad B \leq b,$$

где $A > 0$ подбирается настолько большим, чтобы гарантировать с высокой вероятностью достаточный уровень качества всей продукции (предположив, например, что ξ_k одинаково распределены). Ясно, что ν и ν^* удовлетворяют определению марковской величины или момента остановки.

Мы предлагаем читателю привести пример случайной величины ν , зависящей от будущего.

Пусть $S_\nu = \xi_1 + \dots + \xi_\nu$. Это есть сумма случайного числа случайных величин.

Теорема 2 (Колмогорова—Прохорова). Пусть целочисленная неотрицательная случайная величина ν не зависит от будущего. Тогда если

$$\sum_{k=1}^{\infty} P(\nu \geq k) E|\xi_k| < \infty, \quad (8)$$

то

$$ES_\nu = \sum_{k=1}^{\infty} P(\nu \geq k) E\xi_k. \quad (9)$$

Если $\xi_k \geq 0$, то условие (8) излишне.

Доказательство. Имеем

$$\begin{aligned} ES_\nu &= \sum_{n=1}^{\infty} E(S_\nu; \nu = n) = \sum_{n=1}^{\infty} E(S_n; \nu = n) = \\ &= \sum_{n=1}^{\infty} \sum_{k=1}^n E(\xi_k; \nu = n) = \sum_{k=1}^{\infty} E(\xi_k; \nu \geq k). \end{aligned} \quad (10)$$

Законность перемены порядка суммирования здесь мы поясним позже. Так как событие $\{\nu \geq k\} = \{\nu > k-1\}$ не зависит от σ -алгебры $\mathfrak{F}_{k,\infty}$, то оно не зависит и от $\sigma(\xi_k)$. Следовательно, $E(\xi_k; \nu \geq k) = P(\nu \geq k)E\xi_k$, и формула (9) доказана. Перемена порядка суммирования в (10) возможна благодаря абсолютной суммируемости соответствующих рядов, так как

$$\sum_{k=1}^{\infty} \sum_{n=k}^{\infty} E(|\xi_k|; \nu = n) = \sum_{k=1}^{\infty} E(|\xi_k|; \nu \geq k) = \sum_{k=1}^{\infty} P(\nu \geq k)E|\xi_k| < \infty$$

по условию теоремы. Если $\xi_k \geq 0$, то все слагаемые будут неотрицательны и перемена порядка суммирования станет законной в любом случае. ■

Следствие 1 (тождество Вальда). Если случайные величины ξ_1, ξ_2, \dots независимы и одинаково распределены, $E|\xi_k| < \infty$, а случайная величина ν не зависит от будущего, $E\nu < \infty$, то $ES_\nu = E\xi_1 E\nu$.

Доказательство сразу получается, если заметить, что

$$\sum_{k=1}^{\infty} P(\nu \geq k) = \sum_{k=1}^{\infty} \sum_{i=k}^{\infty} P(\nu = i) = \sum_{i=1}^{\infty} iP(\nu = i) = E\nu. \quad \blacksquare$$

Пример 10. Схема цепной реакции. Пусть мы имеем в качестве исходной одну частицу, которая либо исчезает с вероятностью q , либо превращается в m таких же частиц с вероятностью $p = 1 - q$. Каждая частица нового поколения ведет себя таким же образом независимо от судьбы остальных частиц. Чему равно математическое ожидание числа ζ_n частиц в n -м поколении?

Рассмотрим «двойную последовательность» $\{\xi_k^{(n)}\}_{k=1, n=1}^{\infty, \infty}$ независимых одинаково распределенных случайных величин, принимающих значения m и 0 с вероятностями p и q . Очевидно, что последовательности $\{\xi_k^{(1)}\}_{k=1}^{\infty}, \{\xi_k^{(2)}\}_{k=2}^{\infty}, \dots$ будут независимыми между собой. С помощью этих последовательностей случайные величины ζ_n ($\zeta_0 = 1$) можно представить в виде

$$\begin{aligned} \zeta_1 &= \xi_{\zeta_0}^{(1)} = \xi_1^{(1)}, \\ \zeta_2 &= \xi_1^{(2)} + \dots + \xi_{\zeta_1}^{(2)}, \\ &\dots \dots \dots \\ \zeta_n &= \xi_1^{(n)} + \dots + \xi_{\zeta_{n-1}}^{(n)}, \end{aligned}$$

где в равенстве для ζ_n число слагаемых ζ_{n-1} есть число «частиц-родителей». Так как последовательность $\{\xi_k^{(n)}\}$ не зависит от ζ_{n-1} и $E\xi_k^{(n)} = pm$, то в силу тождества Вальда

$$E\zeta_n = E\xi_1^{(n)} E\zeta_{n-1} = pm E\zeta_{n-1} = (pm)^n.$$

Пример 11. Вернемся к безобидной игре двух игроков, описанной в примере 8, но уже в случае, когда оба капитала $z_1 > 0$ и $z_2 > 0$ первого и второго игроков конечны. Введем случайные величины ξ_k , изображающие выигрыш первого игрока в k -й партии. Очевидно, ξ_k независимы,

$$\xi_k = \begin{cases} 1 & \text{с вероятностью } 1/2, \\ -1 & \text{с вероятностью } 1/2. \end{cases}$$

Число $z_1 + S_k = z_1 + \sum_{j=1}^k \xi_j$ будет означать капитал первого игрока, а $z_2 - S_k$ — капитал второго игрока после k партий. Величина

$$\eta = \min\{k: z_1 + S_k = 0 \text{ или } z_2 - S_k = 0\}$$

есть время до окончания игры, т. е. до разорения одного из игроков. Спрашивается, чему равна вероятность P_i того, что выиграет i -й игрок ($i = 1, 2$)?

Ясно, что η есть марковский момент, при этом $S_\eta = -z_1$ с вероятностью P_2 и $S_\eta = z_2$ с вероятностью $P_1 = 1 - P_2$. Поэтому

$$ES_\eta = P_1 z_2 - P_2 z_1.$$

Если $E\eta < \infty$, то в силу тождества Вальда мы имели бы

$$P_1 z_2 - P_2 z_1 = E\eta E\xi_1 = 0.$$

Отсюда находим, что $P_i = z_i / (z_1 + z_2)$.

Остается проверить конечность $E\eta$. Пусть для простоты $z_1 + z_2 = 2z$ четно. С вероятностью $2^{-\min(z_1, z_2)} \geq 2^{-z}$ игра может закончиться за $\min(z_1, z_2) \leq z$ партий. Так как в течение игры суммарный капитал обоих игроков не меняется, то

$$P(\eta > z) \leq 1 - 2^{-z}, \dots, P(\eta > zN) \leq (1 - 2^{-z})^N.$$

Это, очевидно, влечет за собой конечность

$$E\eta = \sum_{k=0}^{\infty} P(\eta > k).$$

Мы уже отмечали, что не зависящая от $\{\xi_j\}$ случайная величина ν не зависит от будущего. В то же время ν не будет моментом остановки в смысле определения 4. Приведем теперь менее простой пример случайной величины ν , не зависящей от будущего, но не являющейся моментом остановки.

Пример 9А. Рассмотрим две независимые между собой последовательности независимых случайных величин ζ_1, ζ_2, \dots и η_1, η_2, \dots ; $\zeta_j \in F$, $\eta_j \in G$. Рассмотрим далее некую систему, состоящую из двух одинаковых приборов. При пуске системы сначала работает первый прибор случайное время ζ_1 , после чего он выходит из строя. Его заменяет второй прибор, который работает время ζ_2 (на интервале времени $(\zeta_1, \zeta_1 + \zeta_2)$). Одновременно в момент выхода из строя первого прибора он поступает в ремонт (восстанавливается), время ремонта длится η_2 . Если $\eta_2 > \zeta_2$, то в момент $\zeta_1 + \zeta_2$ отказа второго прибора оба прибора окажутся неисправными и система дает отказ. Если же $\eta_2 \leq \zeta_2$, то в момент $\zeta_1 + \zeta_2$ в строй снова вступает первый прибор и работает время ζ_3 , а второй прибор поступает в ремонт, который длится время η_3 . Если $\eta_3 > \zeta_3$, то система отказывает. Если $\eta_3 \leq \zeta_3$, то второй прибор продолжит работу, и т. д. Чему равно математическое ожидание времени τ безотказной работы системы?

Пусть $\nu = \min\{k \geq 2: \eta_k > \zeta_k\}$. Тогда, очевидно, $\tau = \zeta_1 + \dots + \zeta_\nu$, где ζ_j независимы и одинаково распределены, $\{\nu \leq n\} \in \sigma(\zeta_1, \dots, \zeta_n; \eta_2, \dots, \eta_n)$. Это означает, что ν не зависит от будущего. В то же время, если $\eta_j \neq \text{const}$, то $\{\nu \leq n\} \notin \sigma(\zeta_1, \dots, \zeta_n)$ и ν не является марковским моментом относительно последовательности ζ_1, ζ_2, \dots . В силу тождества Вальда $E\tau = E\nu E\zeta_1$. Так как

$$\{\nu = k\} = \bigcap_{j=2}^{k-1} \{\eta_j \leq \zeta_j\} \cap \{\eta_k > \zeta_k\}, \quad k \geq 2,$$

то $P(\nu = k) = q^{k-2}(1-q)$, $k \geq 2$, где

$$q = P(\eta_k \leq \zeta_k) = \int dF(t)G(t+0).$$

Следовательно,

$$E\nu = \sum_{k=2}^{\infty} kq^{k-2}(1-q) = 1 + \sum_{k=1}^{\infty} kq^{k-1}(1-q) = 1 + \frac{1}{1-q},$$

$$E\tau = E\zeta_1 \frac{2-q}{1-q}.$$

Тождество Вальда допускает целый ряд обобщений (подробнее об этом см. § 14.2).

§ 5. Дисперсия

Введем еще одну числовую характеристику случайных величин.

Определение 5. Дисперсией $D\xi$ случайной величины ξ называется число

$$D\xi = E(\xi - E\xi)^2.$$

Это есть мера «разброса» или «рассеивания» распределения. Дисперсия равна моменту инерции распределения единичной массы на прямой. Находим

$$D\xi = E\xi^2 - 2E\xi E\xi + (E\xi)^2 = E\xi^2 - (E\xi)^2. \quad (11)$$

Дисперсию можно было бы определить также как $\min_a E(\xi - a)^2$. В самом деле, в силу этого определения

$$D\xi = E\xi^2 + \min_a (a^2 - 2aE\xi) = E\xi^2 - (E\xi)^2,$$

так как минимум $a^2 - 2aE\xi$ достигается при $a = E\xi$. Это замечание показывает, что число $a = E\xi$ является наилучшей оценкой (приближением) в среднеквадратичном случайной величины ξ .

Величина $\sqrt{D\xi}$ называется *стандартным отклонением*.

Пример 12. $\xi \in \Phi_{\alpha, \sigma^2}$. В примере 2 было выяснено, что $\alpha = E\xi$. Следовательно,

$$D\xi = \int (x - \alpha)^2 \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\alpha)^2/2\sigma^2} dx = \frac{\sigma^2}{\sqrt{2\pi}} \int t^2 e^{-t^2/2} dt.$$

Последнее равенство получено заменой $(x - \alpha)/\sigma = t$. Интегрируя по частям, получим

$$D\xi = -\frac{\sigma^2}{\sqrt{2\pi}} t e^{-t^2/2} \Big|_{-\infty}^{\infty} + \frac{\sigma^2}{\sqrt{2\pi}} \int e^{-t^2/2} dt = \sigma^2.$$

Пример 13. $\xi \in \Pi_\mu$. В примере 3 было вычислено $E\xi = \mu$. Поэтому

$$\begin{aligned} D\xi &= E\xi^2 - (E\xi)^2 = \sum_{k=0}^{\infty} k^2 \frac{\mu^k e^{-\mu}}{k!} - \mu^2 = \\ &= \sum_{k=2}^{\infty} \frac{k(k-1)\mu^k}{k!} e^{-\mu} + \sum_{k=0}^{\infty} \frac{k\mu^k}{k!} e^{-\mu} - \mu^2 = \mu^2 + \mu - \mu^2 = \mu. \end{aligned}$$

Пример 14. Для $\xi \in U_{0,1}$ имеем

$$E\xi^2 = \int_0^1 x^2 dx = \frac{1}{3}, \quad E\xi = \frac{1}{2}.$$

В силу (11) $D\xi = 1/12$.

Пример 15. Для $\xi \in B_p$ в силу соотношений $\xi^2 = \xi$, $E\xi^2 = E\xi = p$ получаем $D\xi = p - p^2 = p(1 - p)$.

Рассмотрим теперь некоторые свойства дисперсии.

D1. $D\xi \geq 0$, $D\xi = 0$ тогда и только тогда, когда $P(\xi = c) = 1$, где c — постоянная (не зависит от ω).

Первое очевидно, так как $D\xi = E(\xi - E\xi)^2 \geq 0$. Пусть $P(\xi = c) = 1$, тогда $(E\xi)^2 = E\xi^2 = c^2$ и, значит,

$$D\xi = c^2 - c^2 = 0.$$

Если $D\xi = E(\xi - E\xi)^2 = 0$, то (так как $(\xi - E\xi)^2 \geq 0$) $P(\xi - E\xi = 0) = 1$ или $P(\xi = E\xi) = 1$ (см. свойство E4).

D2. Если a и b постоянные, то

$$D(a + b\xi) = b^2 D\xi.$$

Это свойство следует непосредственно из определения $D\xi$.

D3. Если случайные величины ξ и η независимы, то

$$E(\xi + \eta) = D\xi + D\eta.$$

Действительно,

$$\begin{aligned} D(\xi + \eta) &= E(\xi + \eta)^2 - (E\xi + E\eta)^2 = \\ &= E\xi^2 + 2E\xi E\eta + E\eta^2 - (E\xi)^2 - (E\eta)^2 - 2E\xi E\eta = \\ &= E\xi^2 - (E\xi)^2 + E\eta^2 - (E\eta)^2 = D\xi + D\eta. \end{aligned}$$

Из этих вычислений видно, что аддитивность дисперсии будет иметь место не только для независимых ξ и η , но и во всех случаях, когда

$$E\xi\eta = E\xi E\eta.$$

Пример 16. Пусть $\nu \geq 0$ — целочисленная величина, не зависящая от последовательности $\{\xi_j\}$ независимых одинаково распределенных величин, $E\nu < \infty$, $E\xi_j = a$. Тогда, как мы знаем, $ES_\nu = aE\nu$. Чему равна DS_ν ?

По формуле полной вероятности

$$\begin{aligned} D S_{\nu} &= E(S_{\nu} - E S_{\nu})^2 = \sum P(\nu = k) E(S_k - E S_{\nu})^2 = \\ &= \sum P(\nu = k) [E(S_k - a k)^2 + (a k - a E \nu)^2] = \\ &= \sum P(\nu = k) k D \xi_1 + a^2 E(\nu - E \nu)^2 = D \xi_1 E \nu + a^2 D \nu. \end{aligned}$$

Это равенство эквивалентно соотношению

$$E(S_{\nu} - \nu a)^2 = E \nu \cdot D \xi_1.$$

В такой форме полученное соотношение остается справедливым для любого момента остановки ν (см. гл. 14). Пользуясь им, можно в примере 11 найти математическое ожидание времени η до окончания безобидной игры, когда капиталы игроков z_1 и z_2 конечны. Действительно, в этом случае $a = 0$,

$$E S_{\eta}^2 = D \xi_1 E \eta = z_1^2 P_2 + z_2^2 P_1.$$

Отсюда находим

$$E \eta = z_1 z_2.$$

§ 6. Коэффициент корреляции и другие числовые характеристики

Две случайные величины ξ и η могут быть функционально (детерминированно) зависимы: $\xi = g(\eta)$; могут быть зависимы, но не функциональным образом; наконец, могут быть независимыми. Коэффициент корреляции двух величин есть число, с помощью которого пытаются количественно характеризовать степень зависимости двух случайных величин друг от друга.

У всех случайных величин, которые встретятся в этом разделе, предполагается существование дисперсии, не равной нулю.

Случайная величина ξ называется *нормированной*, если $E \xi = 0$ и $D \xi = 1$. Любую случайную величину ξ можно линейным преобразованием привести к нормированной, положив $\xi_1 = (\xi - E \xi) / \sqrt{D \xi}$. Пусть ξ и η — некоторые случайные величины, а ξ_1 и η_1 — соответствующие им нормированные случайные величины.

Определение 6. Коэффициентом корреляции между случайными величинами ξ и η называется число $\rho(\xi, \eta) = E \xi_1 \eta_1$.

Свойства коэффициента корреляции следующие.

1. $|\rho(\xi, \eta)| \leq 1$.

Действительно,

$$0 \leq D(\xi_1 \pm \eta_1) = E(\xi_1 \pm \eta_1)^2 = 2 \pm 2\rho(\xi, \eta).$$

Отсюда следует, что $|\rho(\xi, \eta)| \leq 1$. ■

2. Если ξ и η независимы, то $\rho(\xi, \eta) = 0$.

Доказательство следует из того, что ξ_1 и η_1 при этом тоже независимы.

Обратное утверждение, конечно, не верно. В § 3 был приведен пример зависимых случайных величин ξ и η таких, что $E \xi = 0$, $E \eta = 0$, $E \xi \eta = 0$. Для них коэффициент корреляции равен нулю, хотя они и зависимы. Однако, как

мы увидим в главе 7, для нормально распределенных (ξ, η) равенство $\rho(\xi, \eta) = 0$ является необходимым и достаточным для независимости этих величин.

Другой пример, когда коррелированность случайных величин влечет их независимость, дает схема Бернулли. Пусть $P(\xi = 1) = p$, $P(\xi = 0) = 1 - p$, $P(\eta = 1) = q$, $P(\eta = 0) = 1 - q$. Тогда $E\xi = p$, $E\eta = q$, $D\xi = p(1 - p)$, $D\eta = q(1 - q)$, $\rho(\xi, \eta) = \frac{E(\xi - p)(\eta - q)}{\sqrt{pq(1-p)(1-q)}}$. Равенство $\rho(\xi, \eta) = 0$ означает, что $E\xi\eta = pq$ или, что то же, $P(\xi = 1, \eta = 1) = P(\xi = 1)P(\eta = 1)$, $P(\xi = 1, \eta = 0) = P(\xi = 1) - P(\xi = 1, \eta = 1) = p - pq = P(\xi = 1)P(\eta = 0)$, и т. д.

Отсюда легко получить, что в общем случае ξ и η независимы, если $\rho(f(\xi), g(\eta)) = 0$ для любых ограниченных измеримых функций f и g . Достаточно взять $f = I_{(-\infty, x]}$, $g = I_{(-\infty, y]}$, получить отсюда, что $P(\xi < x, \eta < y) = P(\xi < x) \times P(\eta < y)$ и воспользоваться результатами предыдущей главы.

3. $|\rho(\xi, \eta)| = 1$ тогда и только тогда, когда существуют такие числа a и $b \neq 0$, что $P(\eta = a + b\xi) = 1$.

Доказательство. Пусть $P(\eta = a + b\xi) = 1$. Обозначим $E\xi = \alpha$ и $\sqrt{D\xi} = \beta$; тогда

$$\rho(\xi, \eta) = E \frac{\xi - \alpha}{\beta} \cdot \frac{a + b\xi - a - b\alpha}{|b|\beta} = \text{sign } b.$$

Предположим теперь, что $|\rho(\xi, \eta)| = 1$. Пусть, например, $\rho(\xi, \eta) = 1$. Тогда

$$D(\xi_1 - \eta_1) = 2(1 - \rho(\xi, \eta)) = 0.$$

По свойству дисперсии D1 это может быть только тогда, когда

$$P(\xi_1 - \eta_1 = c) = 1.$$

Если $\rho(\xi, \eta) = -1$, то, рассмотрев $D(\xi_1 + \eta_1)$, получим

$$P(\xi_1 + \eta_1 = c) = 1.$$

Если $\rho > 0$, то случайные величины ξ и η называются *положительно коррелированными*; если $\rho < 0$, то ξ и η называются *отрицательно коррелированными*.

Пример 17. Рассмотрим передающее устройство. Случайная величина ξ означает величину передаваемого сигнала. В результате помех приемник принимает величину $\eta = \alpha\xi + \Delta$ (α — коэффициент усиления, Δ — помехи). Предположим, что случайные величины Δ и ξ независимы. Пусть $E\xi = a$, $D\xi = 1$, а $E\Delta = 0$, $D\Delta = \sigma^2$. Подсчитаем коэффициент корреляции ξ и η :

$$\rho(\xi, \eta) = E \left((\xi - a) \frac{\alpha\xi + \Delta - a\alpha}{\sqrt{\alpha^2 + \sigma^2}} \right) = \frac{\alpha}{\sqrt{\alpha^2 + \sigma^2}}.$$

Если σ — большое число по сравнению с усилением α , то ρ близко к нулю и η практически не зависит от ξ . Если же σ мало по сравнению с α , тогда ρ близко к единице и по η бывает легко восстанавливать ξ !

Рассмотрим некоторые другие числовые характеристики случайных величин. В качестве таковых часто используются моменты более высоких порядков.

Определение 7. Моментом порядка k случайной величины ξ называется число $E\xi^k$. Значение $E(\xi - E\xi)^k$ называется *центральный моментом порядка k* , так что дисперсия есть центральный второй момент ξ .

Если дан случайный вектор (ξ_1, \dots, ξ_n) , то величина $E\xi_1^{k_1} \dots \xi_n^{k_n}$ называется *смешанным моментом порядка $k_1 + \dots + k_n$* . Аналогично $E(\xi_1 - E\xi_1)^{k_1} \dots (\xi_n - E\xi_n)^{k_n}$ называется *центральный смешанным моментом* того же порядка.

Для независимых величин смешанные моменты равны, очевидно, произведению соответствующих обычных моментов.

§ 7. Неравенства

Теорема 3 (неравенства Коши—Буняковского). Если ξ_1 и ξ_2 — две произвольные случайные величины, то

$$E|\xi_1 \xi_2| \leq [E\xi_1^2 E\xi_2^2]^{1/2}.$$

Это неравенство иногда называют также *неравенством Шварца*.

Доказательство. Требуемое соотношение получается из неравенства $2|ab| \leq a^2 + b^2$, если положить в нем $a^2 = \xi_1^2/E\xi_1^2$, $b^2 = \xi_2^2/E\xi_2^2$ и взять математическое ожидание обеих его частей. ■

Неравенство Коши—Буняковского является частным случаем более общих неравенств.

Теорема 4. При $r > 1$, $1/r + 1/s = 1$ справедливы неравенство Гёльдера

$$E|\xi_1 \xi_2| \leq [E|\xi_1|^r]^{1/r} [E|\xi_2|^s]^{1/s}$$

и неравенство Минковского

$$[E|\xi_1 + \xi_2|^r]^{1/r} \leq [E|\xi_1|^r]^{1/r} + [E|\xi_2|^r]^{1/r}.$$

Доказательство. Так как x^r при $r > 1$ есть в области $x > 0$ выпуклая вниз функция, равная единице в точке $x = 1$ и имеющая там производную r , то $r(x - 1) \leq x^r - 1$ при всех $x > 0$. Полагая $x = (a/b)^{1/r}$ ($a > 0$, $b > 0$), получим

$$a^{1/r} b^{1-1/r} - b \leq \frac{a}{r} - \frac{b}{r},$$

или, что то же, $a^{1/r} b^{1/s} \leq a/r + b/s$. Если положить здесь $a = |\xi_1|^r/E|\xi_1|^r$, $b = |\xi_2|^s/E|\xi_2|^s$ и взять математическое ожидание, то мы получим неравенство Гёльдера.

Чтобы доказать неравенство Минковского, заметим, что в силу неравенства $|\xi_1 + \xi_2| \leq |\xi_1| + |\xi_2|$ выполняется

$$E|\xi_1 + \xi_2|^r \leq E|\xi_1| |\xi_1 + \xi_2|^{r-1} + E|\xi_2| |\xi_1 + \xi_2|^{r-1}.$$

Применяя здесь к слагаемым правой части неравенство Гёльдера, получим

$$E|\xi_1 + \xi_2|^r \leq \left\{ [E|\xi_1|^r]^{1/r} + [E|\xi_2|^r]^{1/r} \right\} [E|\xi_1 + \xi_2|^{(r-1)s}]^{1/s}.$$

Так как $(r-1)s = r$, $1 - 1/s = 1/r$, то отсюда следует неравенство Минковского. ■

Очевидно, что при $r = s = 2$ неравенство Гёльдера превращается в неравенство Коши—Буняковского.

Теорема 5 (неравенство Иенсена). Если $E\xi$ существует и $g(x)$ — выпуклая вниз функция, то $g(E\xi) \leq Eg(\xi)$.

Доказательство. Если $g(x)$ выпуклая вниз, то для каждого y найдется число $g^1(y)$ такое, что

$$g(x) \geq g(y) + (x - y)g^1(y).$$

Положив здесь $x = \xi$, $y = E\xi$ и взяв математическое ожидание обеих частей этого неравенства, получим

$$Eg(\xi) \geq g(E\xi).$$

Часто бывает полезным также

Следствие 2. При любых $0 < v < u$

$$[E|\xi|^v]^{1/v} \leq [E|\xi|^u]^{1/u}. \quad (12)$$

Это неравенство показывает, в частности, что если существует момент порядка u , то существует и момент любого порядка $v < u$.

Неравенство (12) получается из неравенства Гёльдера, если положить в нем $\xi_1 = |\xi|^v$, $\xi_2 = 1$, $r = u/v$, или из неравенства Иенсена при $g(x) = |x|^{u/v}$ и замене ξ на ξ^v .

Теорема 6. Пусть $\xi \geq 0$ с вероятностью 1. Тогда для любого $\varepsilon > 0$

$$P(\xi \geq \varepsilon) \leq \frac{E\xi}{\varepsilon}. \quad (13)$$

Доказательство. Неравенство доказывают следующие соотношения:

$$E\xi \geq E(\xi; \xi \geq \varepsilon) \geq \varepsilon E(1; \xi \geq \varepsilon) = \varepsilon P(\xi \geq \varepsilon).$$

Если функция $g(x) \geq 0$ монотонно возрастает, то, очевидно, $\{\xi: g(\xi) \geq g(\varepsilon)\} = \{\xi: \xi \geq \varepsilon\}$ и, применяя теорему 6 к случайной величине $\eta = g(\xi)$, получим

Следствие 3. Если $g(x) \uparrow$, $g(x) \geq 0$, то

$$P(\xi \geq \varepsilon) \leq \frac{Eg(\xi)}{g(\varepsilon)}.$$

В частности, при $g(x) = e^{\lambda x}$

$$P(\xi \geq \varepsilon) \leq e^{-\lambda \varepsilon} E e^{\lambda \xi}, \quad \lambda > 0.$$

Следствие 4 (неравенство Чебышева). Для произвольной случайной величины ξ , имеющей математическое ожидание, справедливо неравенство

$$P(|\xi - E\xi| \geq \varepsilon) \leq \frac{D\xi}{\varepsilon^2}. \quad (14)$$

Для доказательства (14) достаточно применить теорему 5 к случайной величине $\eta = (\xi - E\xi)^2 \geq 0$.

Утверждение теоремы 6 также часто называют *неравенством Чебышева* (или *неравенством типа Чебышева*), поскольку оно является по технике доказательства несущественным обобщением (14).

С помощью неравенства Чебышева мы имеем возможность оценивать вероятности различных отклонений ξ , зная лишь $E\xi$ и $D\xi$. В качестве одного из первых применений этого неравенства можно получить так называемый *закон больших чисел в форме Чебышева* (в более общей форме закон больших чисел будет получен в главе 8).

Теорема 7. Пусть ξ_1, ξ_2, \dots — независимые одинаково распределенные случайные величины с математическим ожиданием $E\xi_j = a$ и конечной дисперсией σ^2 , $S_n = \sum_{j=1}^n \xi_j$.

Тогда для любого $\varepsilon > 0$

$$P\left(\left|\frac{S_n}{n} - a\right| > \varepsilon\right) \leq \frac{\sigma^2}{n\varepsilon^2} \rightarrow 0$$

при $n \rightarrow \infty$.

Обсуждение этого утверждения будет проведено в главах 5, 6, 8.

Доказательство теоремы вытекает из неравенства Чебышева, так как

$$E\frac{S_n}{n} = a, \quad D\frac{S_n}{n} = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}.$$

Приведем пример использования неравенства Чебышева в вычислениях.

Пример 18. Предположим, что мы решили измерить диаметр лунного диска, используя фотографии луны, полученные с помощью телескопа. Вследствие атмосферных помех замеры фотографий, сделанных в разное время, дадут разные результаты. Пусть $\xi - a$ означает отклонение замера от истинного значения, равного в некотором масштабе a . И пусть $E\xi = a$, $\sigma = \sqrt{D\xi} = 1$. Проведем подряд n независимых друг от друга замеров и обозначим $\zeta_n = \frac{1}{n}(\xi_1 + \dots + \xi_n)$. Тогда, как мы видели, $E\zeta_n = a$, $D\zeta_n = \sigma^2/n$. Так как дисперсия среднего арифметического всех наблюдений уменьшается с ростом числа наблюдений, то величину a естественно оценивать с помощью ζ_n .

Сколько надо сделать наблюдений, чтобы $|\zeta_n - a| \leq 0,1$ с вероятностью, большей 0,99? То есть чтобы выполнялось $P(|\zeta_n - a| \leq 0,1) > 0,99$ или $P(|\zeta_n - a| > 0,1) \leq 0,01$. По неравенству Чебышева $P(|\zeta_n - a| > 0,1) \leq \sigma^2/(n \cdot 0,01)$. Значит, если n выбрать таким, чтобы $\sigma^2/(n \cdot 0,01) \leq 0,01$, то требуемое неравенство будет выполнено. Отсюда получаем $n \geq 10^4$.

Приведенный пример иллюстрирует возможность использования неравенства Чебышева для оценок вероятностей тех или иных отклонений случайных величин. Но этот пример, пожалуй, еще больше иллюстрирует, насколько грубым с точки зрения практических подсчетов является неравенство Чебышева. Если читатель вернется к примеру 18 после знакомства в главе 8 с центральной предельной теоремой, то он может легко подсчитать, что на самом деле для получения заданной погрешности достаточно провести не 10^4 , а около 670 наблюдений.

§ 8. Обобщение понятия условного математического ожидания

В заключение этой главы введем понятие, которое, наряду с уже введенными, является полезным и важным инструментом в теории вероятностей. Чтобы дать возможность читателю при первом чтении книги этот раздел опустить, мы в дальнейшем будем избегать прямого использования проводимых здесь рассуждений вплоть до глав 12, 14.

1. Определение условного математического ожидания (у.м.о.). В § 2 введено понятие у.м.о. относительно произвольного события B , $P(B) > 0$, которое определяется равенством

$$E(\xi | B) = \frac{E(\xi; B)}{P(B)}, \quad (15)$$

где $E(\xi; B) = \int_B \xi dP = E\xi I_B$, $I_B = I_B(\omega)$ есть случайная величина, равная индикатору множества B . Мы уже видели и будем убеждаться не раз в дальнейшем, что это очень полезное понятие. Однако определение (15), с помощью которого это понятие было введено, обладает тем недостатком, что оно теряет смысл в случае $P(B) = 0$. Как преодолеть этот недостаток?

Тот факт, что условие $P(B) > 0$ не должно играть существенной роли, можно проиллюстрировать следующими соображениями. Допустим, что ξ и η независимы, $B = \{\eta = x\}$ и $P(B) > 0$. Тогда для любой измеримой функции $\varphi(x, y)$, согласно (15),

$$E[\varphi(\xi, \eta) | \eta = x] = \frac{E\varphi(\xi, \eta) I_{\{\eta=x\}}}{P(\eta = x)} = \frac{E\varphi(\xi, x) I_{\{\eta=x\}}}{P(\eta = x)} = E\varphi(\xi, x). \quad (16)$$

Последнее равенство справедливо, так как случайные величины $\varphi(\xi, x)$ и $I_{\{\eta=x\}}$ как функции соответственно от ξ и η независимы и, следовательно,

$$E\varphi(\xi, x) I_{\{\eta=x\}} = E\varphi(\xi, x) P(\eta = x).$$

Соотношения (16) показывают, что понятие у.м.о. может сохранять свой смысл и в случае, когда вероятность условия равна 0 — ведь само по себе равенство

$$E[\varphi(\xi, \eta) | \eta = x] = E\varphi(\xi, x)$$

для независимых ξ и η представляется естественным и с предположением $P(\eta = x) > 0$ никак не связано.

Пусть \mathfrak{A} есть σ -подалгебра \mathfrak{F} . Определим теперь понятие у.м.о. случайной величины ξ относительно \mathfrak{A} , которое будем обозначать через $E(\xi | \mathfrak{A})$. Определение дадим сначала для «дискретного» случая, но так, чтобы оно легко обобщалось.

Напомним, что дискретным называется случай, когда σ -алгебра \mathfrak{A} образована (порождена) не более чем счетной последовательностью непересекающихся событий A_1, A_2, \dots ; $\bigcup A_j = \Omega$, $P(A_j) > 0$. Этот факт мы записываем в виде $\mathfrak{A} = \sigma(A_1, A_2, \dots)$, он означает, что элементами \mathfrak{A} являются всевозможные объединения множеств A_1, A_2, \dots .

Пусть L_2 есть совокупность всех случайных величин (всех функций $\xi(\omega)$, определенных на $(\Omega, \mathfrak{F}, P)$), для которых $E\xi^2 < \infty$. В линейном пространстве L_2 можно ввести скалярное произведение $(\xi, \eta) = E\xi\eta$ (это делает пространство L_2 гильбертовым).

Рассмотрим теперь линейное подпространство $H_{\mathfrak{A}}$ всех функций вида $\xi(\omega) = \sum_k c_k I_{A_k}(\omega)$, где $I_{A_k}(\omega)$ — индикаторы множеств A_k . Ясно, что $H_{\mathfrak{A}}$ есть пространство всех функций, измеримых относительно \mathfrak{A} , и его можно представлять себе как пространство, натянутое на ортогональный базис $\{I_{A_k}(\omega)\}$ в L_2 .

Обратимся к определению у.м.о. Мы знаем, что обычное математическое ожидание $a = E\xi$ для $\xi \in L_2$ можно определить как единственное значение a , при котором достигается минимум функции $\varphi(a) = E(\xi - a)^2$. Рассмотрим теперь задачу

о минимизации функционала $\varphi(a) = E(\xi - a(\omega))^2$, $\xi \in L_2$, по всем \mathfrak{A} -измеримым функциям $a(\omega)$ из $H_{\mathfrak{A}}$.

Определение 8. Пусть $\xi \in L_2$. \mathfrak{A} -измеримая случайная величина $a(\omega)$, для которой достигается $\min_{a \in H_{\mathfrak{A}}} \varphi(a)$, называется у.м.о. относительно \mathfrak{A} и обозначается через $E(\xi | \mathfrak{A})$.

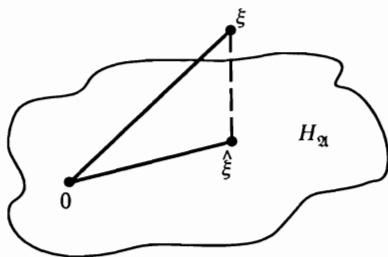


Рис. 5.

Таким образом, в отличие от обычных математических ожиданий, у.м.о. $E(\xi | \mathfrak{A})$ есть случайная величина. Рассмотрим ее подробнее. Ясно, что минимум $\varphi(a)$ достигается, если в качестве $a(\omega)$ взять в L_2 проекцию $\hat{\xi}$ элемента ξ на $H_{\mathfrak{A}}$, т.е. элемент $\hat{\xi} \in H_{\mathfrak{A}}$, для которого $\xi - \hat{\xi} \perp H_{\mathfrak{A}}$ (см. рис. 5). В этом случае для любого $a \in H_{\mathfrak{A}}$

$$\hat{\xi} - a \in H_{\mathfrak{A}}, \quad \xi - \hat{\xi} \perp \xi - a,$$

$$\varphi(a) = E(\xi - \hat{\xi} + \hat{\xi} - a)^2 = E(\xi - \hat{\xi})^2 + E(\hat{\xi} - a)^2,$$

$$\varphi(a) \geq \varphi(\hat{\xi}),$$

$$\varphi(a) = \varphi(\hat{\xi}), \text{ если только } a = \hat{\xi} \text{ п. в.}$$

Итак, в L_2 операция у.м.о. есть обычный ортопроектор на $H_{\mathfrak{A}}$ ($\hat{\xi} = E(\xi | \mathfrak{A})$ есть проекция ξ на $H_{\mathfrak{A}}$).

Так как для дискретной σ -алгебры \mathfrak{A} $\hat{\xi}$ как элемент $H_{\mathfrak{A}}$ имеет вид $\hat{\xi} = \sum c_k I_{A_k}$, то условия ортогональности $\xi - \hat{\xi} \perp H_{\mathfrak{A}}$ (или, что то же, $E(\xi - \hat{\xi}) I_{A_k} = 0$) однозначно определяют коэффициенты c_k :

$$E(\xi I_{A_k}) = c_k P(A_k), \quad c_k = \frac{E(\xi; A_k)}{P(A_k)} = E(\xi | A_k),$$

так что

$$E(\xi | \mathfrak{A}) = \hat{\xi} = \sum_k E(\xi | A_k) I_{A_k}.$$

Таким образом, случайная величина $E(\xi | \mathfrak{A})$ постоянна на A_k и равна на этих множествах усреднению ξ по A_k .

Если ξ и \mathfrak{A} независимы (т.е. $P(\xi \in B; A_k) = P(\xi \in B)P(A_k)$), то, очевидно, $E(\xi; A_k) = E\xi P(A_k)$ и $\hat{\xi} = E\xi$. Если же $\mathfrak{A} = \mathfrak{F}$, то \mathfrak{F} тоже «дискретна», ξ постоянна на множествах A_k и, значит, $\hat{\xi} = \xi$.

Отметим теперь следующие основные свойства у.м.о., которые позволяют нам освободиться от двух частных предположений ($\xi \in L_2$ и дискретность \mathfrak{A}), которые мы ввели вначале, чтобы лучше понять смысл у.м.о.:

1) $\hat{\xi}$ измерима относительно \mathfrak{A} ;

2) для любого события $A \in \mathfrak{A}$

$$E(\hat{\xi}; A) = E(\xi; A).$$

Первое свойство очевидно. Второе следует из того, что любое событие $A \in \mathfrak{A}$ представимо в виде $A = \bigcup_k A_{j_k}$ и, стало быть,

$$E(\hat{\xi}; A) = \sum_k E(\hat{\xi}; A_{j_k}) = \sum_k c_{j_k} P(A_{j_k}) = \sum_k E(\xi; A_{j_k}) = E(\xi; A).$$

Это свойство довольно ясно: усреднение по множеству A величины ξ дает тот же результат, что усреднение уже усредненной по A_{j_k} величины $\widehat{\xi}$.

Лемма 1. Свойства 1), 2) однозначно определяют у. м. о. и эквивалентны определению 8.

Доказательство. В одну сторону утверждение леммы уже доказано. Пусть теперь выполнены условия 1), 2). Измеримость ξ относительно \mathfrak{A} означает, что $\widehat{\xi}$ постоянна на множествах A_k . Обозначим значение $\widehat{\xi}$ на A_k через c_k . Так как $A_k \in \mathfrak{A}$, то из свойства 2 следует, что

$$E(\widehat{\xi}; A_k) = c_k P(A_k) = E(\xi; A_k)$$

и, следовательно, при $\omega \in A_k$

$$\widehat{\xi} = c_k = \frac{E(\xi; A_k)}{P(A_k)}.$$

Теперь можно дать общее определение у. м. о.

Определение 9. Пусть ξ есть случайная величина на вероятностном пространстве $\langle \Omega, \mathfrak{F}, P \rangle$ и $\mathfrak{A} \subset \mathfrak{F}$ есть σ -подалгебра \mathfrak{F} . Условным математическим ожиданием ξ относительно \mathfrak{A} называется случайная величина $\widehat{\xi}$, обозначаемая через $E(\xi | \mathfrak{A})$, которая обладает следующими свойствами:

- 1) $\widehat{\xi}$ измерима относительно \mathfrak{A} ;
- 2) для всякого $A \in \mathfrak{A}$ справедливо $E(\widehat{\xi}; A) = E(\xi; A)$.

В этом определении случайная величина ξ может быть как скалярной, так и векторной.

Сразу возникают вопросы: существует ли такая величина $\widehat{\xi}$ и единственна ли она? В «дискретном» случае мы видели, что ответ на эти вопросы положителен. В общем случае справедлива

Теорема 8. Если $E|\xi|$ конечно, то функция $\widehat{\xi} = E(\xi | \mathfrak{A})$ в определении 9 всегда существует и единственна с точностью до значений на множестве вероятности 0.

Доказательство. Предположим сначала, что ξ скалярна, $\xi \geq 0$. Тогда функция множества

$$Q(A) = \int_A \xi dP = E(\xi; A), \quad A \in \mathfrak{A},$$

будет мерой на (Ω, \mathfrak{A}) , которая абсолютно непрерывна относительно P , поскольку $P(A) = 0$ влечет за собой $Q(A) = 0$. Следовательно, по теореме Радона—Никодима (см. Приложение 3) существует \mathfrak{A} -измеримая функция $\widehat{\xi} = E(\xi | \mathfrak{A})$, единственная с точностью до значений на множестве меры нуль, такая, что

$$Q(A) = \int_A \widehat{\xi} dP = E(\widehat{\xi}; A).$$

В общем случае положим $\xi = \xi^+ - \xi^-$, $\xi^+ = \max(0, \xi) \geq 0$, $\xi^- = \max(0, -\xi) \geq 0$, $\widehat{\xi} = \widehat{\xi}^+ - \widehat{\xi}^-$, где $\widehat{\xi}^\pm$ — у. м. о. для ξ^\pm . Этим доказано существование у. м. о., так как $\widehat{\xi}$ будет удовлетворять условиям 1), 2) определения 9. Отсюда же следует и единственность, так как предположение о неединственности $\widehat{\xi}$ будет означать

неединственность $\widehat{\xi}^+$ или $\widehat{\xi}^-$. Доказательство для векторных $\widehat{\xi}$ сводится к одномерному случаю, так как свойствами 1), 2) будут обладать координаты $\widehat{\xi}$, существование и единственность которых нами доказана. ■

Существо проведенного доказательства достаточно прозрачно: ведь по условию 2) для всякого $A \in \mathfrak{A}$ задано $E(\widehat{\xi}; A) = \int_A \widehat{\xi} dP$, т.е. заданы значения интегралов от $\widehat{\xi}$ по всем множествам $A \in \mathfrak{A}$. Ясно, что это должно определять \mathfrak{A} -измеримую функцию $\widehat{\xi}$ однозначно с точностью до значений на множестве меры нуль.

Смысл $E(\xi|\mathfrak{A})$ остается прежним — грубо говоря, это есть усреднение ξ по «неделимым» элементам \mathfrak{A} .

Если $\mathfrak{A} = \mathfrak{F}$, то, очевидно, $\widehat{\xi} = \xi$ удовлетворяют свойствам 1), 2) и, стало быть, $E(\xi|\mathfrak{F}) = \xi$.

Определение 10. Пусть ξ и η — случайные величины на $(\Omega, \mathfrak{F}, P)$, $\mathfrak{A} = \sigma(\eta)$ есть σ -алгебра, порожденная случайной величиной η . Тогда $E(\xi|\mathfrak{A})$ называется также *условным математическим ожиданием величины ξ относительно η* .

Иногда для упрощения записи мы будем вместо $E(\xi|\sigma(\eta))$ писать $E(\xi|\eta)$. Это не приводит к недоразумениям.

Так как $E(\xi|\eta)$ по определению есть $\sigma(\eta)$ -измеримая случайная величина, то это означает (см. § 5 главы 3), что существует измеримая функция $g(x)$, для которой $E(\xi|\eta) = g(\eta)$. По аналогии с дискретным случаем можно истолковывать здесь величину $g(x)$ как результат усреднения ξ по множеству $\{\eta = x\}$. (Напомним, что в дискретном случае $g(x) = E(\xi|\eta = x)$.)

Определение 11. Если $\xi = I_C$ есть индикатор множества $C \in \mathfrak{F}$, то $E(I_C|\mathfrak{A})$ будет называться *условной вероятностью $P(C|\mathfrak{A})$ события C относительно \mathfrak{A}* . Если $\mathfrak{A} = \sigma(\eta)$, то мы будем говорить об *условной вероятности $P(C|\eta)$ события C относительно η* .

2. Свойства у.м.о. 1. У.м.о. обладает свойствами обычных математических ожиданий с той лишь разницей, что они выполняются почти наверное (с вероятностью 1):

- а) $E(a + b\xi|\mathfrak{A}) = a + bE(\xi|\mathfrak{A})$;
- б) $E(\xi_1 + \xi_2|\mathfrak{A}) = E(\xi_1|\mathfrak{A}) + E(\xi_2|\mathfrak{A})$;
- в) Если $\xi_1 \leq \xi_2$ п. н., то $E(\xi_1|\mathfrak{A}) \leq E(\xi_2|\mathfrak{A})$ п. н.

Чтобы доказать, например, свойство а), надо убедиться согласно определению 9, что:

- 1) $a + bE(\xi|\mathfrak{A})$ есть \mathfrak{A} -измеримая функция;
- 2) $E(a + b\xi; A) = E(a + bE(\xi|\mathfrak{A}); A)$ для любого $A \in \mathfrak{A}$.

Выполнение 1) очевидно; выполнение 2) вытекает из свойств линейности обычного математического ожидания (или обычного интеграла).

Свойство б) доказывается точно так же.

Чтобы доказать свойство в), положим для краткости $\widehat{\xi}_i = E(\xi_i|\mathfrak{A})$. Тогда для любого $A \in \mathfrak{A}$

$$\int_A \widehat{\xi}_1 dP = E(\widehat{\xi}_1; A) = E(\xi_1; A) \leq E(\xi_2; A) = \int_A \widehat{\xi}_2 dP,$$

$$\int_A (\widehat{\xi}_2 - \widehat{\xi}_1) dP \geq 0.$$

Отсюда вытекает, что $\widehat{\xi}_2 - \xi_1 \geq 0$ п. н.

2. *Неравенство Чебышева.* Если $\xi \geq 0$, $x > 0$, то $P(\xi \geq x | \mathfrak{A}) \leq E(\xi | \mathfrak{A})/x$.

Это свойство вытекает из в), так как $P(\xi \geq x | \mathfrak{A}) = E(I_{\{\xi \geq x\}} | \mathfrak{A})$, где I_A — индикатор события A , и справедливо неравенство $I_{\{\xi \geq x\}} \leq \xi/x$.

3. Если \mathfrak{A} и $\sigma(\xi)$ независимы, то $E(\xi | \mathfrak{A}) = E\xi$. Так как $\widehat{\xi} = E\xi$ является \mathfrak{A} -измеримой функцией, то остается проверить лишь второе условие определения 9: для любого $A \in \mathfrak{A}$

$$E(\widehat{\xi}; A) = E(\xi; A).$$

Справедливость этого равенства вытекает из независимости случайных величин I_A и ξ и соотношений $E(\xi; A) = E(\xi I_A) = E\xi E I_A = E(\widehat{\xi}; A)$.

Отсюда следует, в частности, что если ξ и η независимы, то $E(\xi | \eta) = E\xi$. Если σ -алгебра \mathfrak{A} тривиальна, то, очевидно, также $E(\xi | \mathfrak{A}) = E\xi$.

4. Для у. м. о. верны теоремы сходимости, справедливые для обычных математических ожиданий, например

Теорема о монотонной сходимости (см. Приложение 3). Если $0 \leq \xi_n \uparrow \xi$ п. н., то $E(\xi_n | \mathfrak{A}) \uparrow E(\xi | \mathfrak{A})$ п. н.

Действительно, из $\xi_{n+1} \geq \xi_n$ п. н. следует $\widehat{\xi}_{n+1} \geq \widehat{\xi}_n$ п. н., где $\widehat{\xi}_n = E(\xi_n | \mathfrak{A})$. Поэтому существует \mathfrak{A} -измеримая ξ , такая, что $\widehat{\xi}_n \uparrow \widehat{\xi}$ п. н. В силу обычной теоремы о монотонной сходимости для любого $A \in \mathfrak{A}$

$$\int_A \widehat{\xi}_n dP \rightarrow \int_A \xi dP, \quad \int_A \xi_n dP \rightarrow \int_A \xi dP.$$

Так как левые части этих соотношений совпадают, то совпадают и правые. Это и означает, что $\widehat{\xi} = E(\xi | \mathfrak{A})$.

5. Если η скалярна и измерима относительно \mathfrak{A} , $E|\xi| < \infty$, $E|\xi\eta| < \infty$, то

$$E(\eta\xi | \mathfrak{A}) = \eta E(\xi | \mathfrak{A}). \quad (17)$$

Другими словами, \mathfrak{A} -измеримые случайные величины ведут себя относительно операции у. м. о. как постоянные (ср. со свойством 1, а)).

Чтобы доказать (17), заметим, что если $\eta = I_B$ (индикатор множества $B \in \mathfrak{A}$), то утверждение справедливо, так как при любом $A \in \mathfrak{A}$

$$\int_A E(I_B \xi | \mathfrak{A}) dP = \int_A I_B \xi dP = \int_{AB} \xi dP = \int_{AB} E(\xi | \mathfrak{A}) dP = \int_A I_B E(\xi | \mathfrak{A}) dP.$$

Отсюда и из линейности у. м. о. следует, что утверждение справедливо и для любых простых функций η .

Если $\xi \geq 0$, $\eta \geq 0$, то, взяв последовательность простых функций $0 \leq \eta_n \uparrow \eta$ и пользуясь теоремой о монотонной сходимости в равенстве

$$E(\eta_n \xi | \mathfrak{A}) = \eta_n E(\xi | \mathfrak{A}),$$

мы получим (17). Переход к случаю произвольных ξ и η осуществляется обычным образом — путем рассмотрения положительных и отрицательных частей случайных величин ξ и η . При этом для того, чтобы полученные разности и суммы имели смысл, нужно требовать существования $E|\xi|$, $E|\xi\eta|$.

6. Для у.м.о. остаются справедливыми все основные неравенства для обычных математических ожиданий, в частности, неравенство Коши—Буняковского

$$E(|\xi_1 \xi_2| | \mathfrak{A}) \leq [E(\xi_1^2 | \mathfrak{A}) E(\xi_2^2 | \mathfrak{A})]^{1/2}$$

и неравенство Иенсена: если $E|\xi| < \infty$, то для любой выпуклой вниз функции

$$g(E(\xi | \mathfrak{A})) \leq E(g(\xi) | \mathfrak{A}). \quad (18)$$

Неравенство Коши—Буняковского доказывается точно так же, как для обычных математических ожиданий, поскольку доказательство не использует других свойств математических ожиданий, кроме линейности.

Неравенство Иенсена вытекает из следующего соотношения. В силу выпуклости $g(x)$ для каждого y найдется число $g_1(y)$ такое, что $g(x) \geq g(y) + (x - y)g_1(y)$. Положим здесь $x = \xi$, $y = \hat{\xi} = E(\xi | \mathfrak{A})$ и возьмем у.м.о. обеих частей этого неравенства. Так как в силу свойства 5

$$E[(\xi - \hat{\xi})g_1(\hat{\xi}) | \mathfrak{A}] = g_1(\hat{\xi})E(\xi - \hat{\xi} | \mathfrak{A}) = 0,$$

то получим (18).

7. Формула полной вероятности

$$E\xi = EE(\xi | \mathfrak{A})$$

вытекает непосредственно из свойства 2) определения 9 при $A = \Omega$.

8. Последовательное усреднение (обобщение свойств 7): если $\mathfrak{A} \subset \mathfrak{A}_1 \subset \mathfrak{F}$, то

$$E(\xi | \mathfrak{A}) = E[E(\xi | \mathfrak{A}_1) | \mathfrak{A}].$$

Действительно, для любого $A \in \mathfrak{A}$ в силу того, что $A \subset \mathfrak{A}_1$, имеем

$$\int_A E[E(\xi | \mathfrak{A}_1) | \mathfrak{A}] dP = \int_A E(\xi | \mathfrak{A}_1) dP = \int_A \xi dP = \int_A E(\xi | \mathfrak{A}) dP.$$

Очевидно, что свойства 1, 3–5, 7, 8 справедливы как для скалярных, так и для векторных случайных величин ξ . Следующее свойство у.м.о. мы выделим особо.

9. Для $\xi \in L_2$ при $a(\omega) = E(\xi | \mathfrak{A})$ достигается минимум $E(\xi - a(\omega))^2$ среди всех \mathfrak{A} -измеримых функций $a(\omega)$.

Действительно, $E(\xi - a(\omega))^2 = EE((\xi - a(\omega))^2 | \mathfrak{A})$, но $a(\omega)$ ведет себя как постоянная относительно операции $E(\cdot | \mathfrak{A})$ (см. свойство 5), так что

$$E((\xi - a(\omega))^2 | \mathfrak{A}) = E((\xi - E(\xi | \mathfrak{A}))^2 | \mathfrak{A}) + (E(\xi | \mathfrak{A}) - a(\omega))^2,$$

и минимум этого выражения достигается при $a(\omega) = E(\xi | \mathfrak{A})$.

Это свойство доказывает эквивалентность определений 8 и 9 в случае $\xi \in L_2$ (в обоих определениях у.м.о. определяется однозначно с точностью до значений на множестве меры нуль). В связи с этим свойством отметим еще раз, что в L_2 операция у.м.о. есть операция проектирования на $H_{\mathfrak{A}}$ (см. пояснения к определению 8).

Свойство 9 можно перенести на многомерный случай в следующей форме: при $a(\omega) = E(\xi | \mathfrak{A})$ достигается $\min(\xi - a(\omega))^T V (\xi - a(\omega))$ для любой неотрицательно определенной матрицы V , где \min берется по всем \mathfrak{A} -измеримым функциям $a(\omega)$.

В заключение отметим, что свойство 5 при широких предположениях допускает следующее обобщение.

5А. Если η измерима относительно \mathfrak{A} , $g(\omega, \eta)$ есть измеримая функция переменных $\omega \in \Omega$ и $\eta \in \mathbb{R}^k$, то

$$E(g(\omega, \eta) | \mathfrak{A}) = E(g(\omega, y) | \mathfrak{A})|_{y=\eta}. \quad (19)$$

Отсюда следует формула (последовательного усреднения или полной вероятности)

$$Eg(\omega, \eta) = E[E(g(\omega, y) | \mathfrak{A})|_{y=\eta}],$$

которую можно рассматривать как обобщение теоремы Фубини (см. § 6 и § 3.6). Действительно, если $g(\omega, y)$ не зависит от \mathfrak{A} , то

$$E(g(\omega, y) | \mathfrak{A}) = Eg(\omega, y), \quad E(g(\omega, \eta) | \mathfrak{A}) = Eg(\omega, y)|_{y=\eta}, \\ Eg(\omega, \eta) = E[Eg(\omega, y)|y=\eta].$$

По форме это и есть теорема Фубини, но η здесь векторная случайная величина, а ω может иметь произвольную природу.

Мы докажем свойство 5А в упрощающем предположении, что существует последовательность простых функций η_n такая, что $g(\omega, \eta_n) \uparrow g(\omega, \eta)$, $h(\omega, \eta_n) \uparrow h(\omega, \eta)$ п. н., где $h(\omega, y) = E(g(\omega, y) | \mathfrak{A})$. Действительно, пусть $\eta_n = y_k$ при $\omega \in A_k \subset \mathfrak{A}$. Тогда

$$g(\omega, \eta_n) = \sum g(\omega, y_k) I_{A_k}.$$

В силу свойства 5 отсюда следует выполнение (19) для функций η_n . Остается воспользоваться теоремой о монотонной сходимости (свойство 4) в равенстве $E(g(\omega, \eta_n) | \mathfrak{A}) = h(\omega, \eta_n)$.

§ 9. Условные распределения

Наряду с у. м. о. можно рассматривать *условные распределения* относительно σ -подалгебр и относительно случайных величин. В этом параграфе мы остановимся лишь на последних.

Пусть ξ и η — две случайные величины на $(\Omega, \mathfrak{F}, P)$ со значениями соответственно в \mathbb{R}^s , и \mathbb{R}^k и \mathfrak{B}^s — σ -алгебра борелевских множеств из \mathbb{R}^s .

Определение 12. Функция $P(B | y)$ двух переменных $y \in \mathbb{R}^k$, $B \in \mathfrak{B}^s$ называется *условным распределением ξ при условии $\eta = y$* , если:

1. При каждом $B \in \mathfrak{B}^s$ $P(B | \eta)$ есть условная вероятность $P(\xi \in B | \eta)$ события $\{\xi \in B\}$ относительно η , т. е. $P(B | y)$ есть борелевская функция от y такая, что для любого $A \in \mathfrak{B}^k$

$$E(P(B | \eta); \eta \in A) \equiv \int_A P(B | y) P(\eta \in dy) = P(\xi \in B, \eta \in A).$$

2. При каждом $y \in \mathbb{R}^k$ $P(B | y)$ есть распределение вероятностей по B .

Функцию $P(B | y)$ будем иногда записывать в более «расшифрованной форме» в виде $P(B | y) = P(\xi \in B | \eta = y)$.

Мы знаем, что для каждого $B \in \mathfrak{B}^s$ существует борелевская функция $g_B(y)$ такая, что $g_B(\eta) = P(\xi \in B | \eta)$. Таким образом, положив $P(B | y) = g_B(y)$, мы удовлетворим условию 1 определения 1. Однако условие 2 при этом из свойств у. м. о. никак не следует и выполняться совсем не обязано — ведь условная вероятность

$P(\xi \in B | \eta)$ определена при каждом B с точностью до значений на множестве N_B меры нуль (так что существует много вариантов у. м. о.), и это множество может быть своим для каждого B . Поэтому если объединение $N = \bigcup_{B \in \mathfrak{B}^s} N_B$ имеет не нулевую вероятность, то может оказаться, что, например, равенства

$$P(\xi \in B_1 \cup B_2 | \eta) = P(\xi \in B_1 | \eta) + P(\xi \in B_2 | \eta)$$

(аддитивность вероятности) сразу для всех непересекающихся B_1, B_2 из \mathfrak{B}^s не выполняются ни для одного ω из N , т. е. на ω -множестве N положительной вероятности функция $g_B(y)$ не будет распределением как функция B .

Однако в случае, когда ξ есть случайная величина со значениями в \mathbb{R}^s с σ -алгеброй борелевских множеств \mathfrak{B}^s , $g_B(\eta) = P(\xi \in B | \eta)$ всегда можно выбрать таким образом, что $g_B(y)$ будет условным распределением²⁾.

Как и следовало ожидать, условные распределения обладают тем естественным свойством, что у. м. о. выражаются в виде интегралов по условным распределениям.

Теорема 9. Для любой измеримой функции $g(x)$, отображающей \mathbb{R}^s в \mathbb{R} , такой, что $E|g(\xi)| < \infty$, справедливо равенство

$$E(g(\xi) | \eta) = \int g(x) P(dx | \eta). \quad (20)$$

Доказательство. Достаточно рассмотреть случай, когда $g(x) \geq 0$. Если $g(x) = I_A(x)$ есть индикатор множества A , то формула (20), очевидно, верна. Стало быть, она верна для любой простой функции $g_n(x)$ (т. е. для функции, принимающей конечное число значений). Остается взять последовательность $g_n \uparrow g$ и воспользоваться монотонностью обеих частей в (20) и свойством 4 из § 8. ■

В реальных задачах для подсчета условных распределений часто можно пользоваться следующим простым правилом, которое для наглядности мы запишем в форме

$$P(\xi \in B | \eta = y) = \frac{P(\xi \in B, \eta \in dy)}{P(\eta \in dy)}. \quad (21)$$

Очевидно, что формально оба условия определения 12 будут удовлетворены.

Если ξ и η имеют плотность распределения, то этому равенству будет придан точный смысл.

Определение 13. Пусть условное распределение $P(B | y)$ при каждом y абсолютно непрерывно относительно некоторой меры μ в \mathbb{R}^s :

$$P(\xi \in B | \eta = y) = \int_B f(x|y) \mu(dx).$$

Тогда плотность $f(x|y)$ называется *условной плотностью ξ (относительно меры μ) при условии $\eta = y$* .

Другими словами, измеримая по паре переменных x, y функция $f(x|y)$ есть *условная плотность ξ при условии $\eta = y$* , если:

²⁾ Подробнее об этом см. в [7, 16, 22].

1) для любых борелевских множеств $A \subset \mathbb{R}^k$, $B \subset \mathbb{R}^s$

$$\int_{y \in A} \int_{x \in B} f(x|y) \mu(dx) P(\eta \in dy) = P(\xi \in B, \eta \in A); \quad (22)$$

2) при каждом y функция $f(x|y)$ есть плотность распределения вероятностей. Из теоремы 9 вытекает, что если существует условная плотность, то

$$E(g(\xi) | \eta) = \int g(x) f(x | \eta) \mu(dx).$$

Если предположить дополнительно, что распределение η имеет плотность $q(y)$ относительно некоторой меры λ в \mathbb{R}^k , то (22) можно записать в виде

$$\int_{y \in A} \int_{x \in B} f(x|y) q(y) \mu(dx) \lambda(dy) = P(\xi \in B, \eta \in A). \quad (23)$$

Рассмотрим теперь прямое произведение пространств \mathbb{R}^s и \mathbb{R}^k и на нем прямое произведение мер $\mu \times \lambda$ (если $C = B \times A$, $B \subset \mathbb{R}^s$, $A \in \bar{\mathbb{R}}^k$, то $\mu \times \lambda(C) = \mu(B) \lambda(A)$). В этом пространстве соотношение (23) означает, очевидно, что совместное распределение ξ и η в $\mathbb{R}^s \times \mathbb{R}^k$ имеет плотность относительно $\mu \times \lambda$, равную

$$f(x, y) = f(x|y) q(y).$$

Справедливо и обратное утверждение.

Теорема 10. Если совместное распределение ξ и η в $\mathbb{R}^s \times \mathbb{R}^k$ имеет плотность $f(x, y)$ относительно $\mu \times \lambda$, то функция

$$f(x|y) = \frac{f(x, y)}{q(y)}, \quad \text{где} \quad q(y) = \int f(x, y) \mu(dx),$$

есть условная плотность ξ при условии $\eta = y$, а функция $q(y)$ есть плотность η относительно меры λ .

Доказательство. Утверждение теоремы относительно $q(y)$ очевидно, так как $\int_A q(y) \lambda(dy) = P(\eta \in A)$. Остается заметить, что $f(x|y) = f(x, y)/q(y)$ удовлетворяет всем условиям в определении 13 условной плотности (равенство (23), эквивалентное (22), выполнено очевидным образом). ■

Теорема уточняет соотношение (21) в случае наличия плотностей у распределений ξ и η .

Пример 19. Пусть ξ_1 , ξ_2 независимы, $\xi_1 \in \Pi_{\lambda_1}$, $\xi_2 \in \Pi_{\lambda_2}$. Чему равно условное распределение ξ_1 при условии, что $\xi_1 + \xi_2 = n$? Мы могли бы без труда подсчитать искомую условную вероятность $P(\xi_1 = k | \xi_1 + \xi_2 = n)$, $k \leq n$, без использования теоремы 10, поскольку $\xi_1 + \xi_2 \in \Pi_{\lambda_1 + \lambda_2}$ и вероятность события $\{\xi_1 + \xi_2 = n\}$ положительна. Оставив этот путь для сопоставлений, воспользуемся все же теоремой 10. Здесь ξ_1 и $\eta = \xi_1 + \xi_2$ имеют плотности (равные вероятностям) относительно считающих мер, так что

$$f(k, n) = P(\xi_1 = k, \eta = n) = P(\xi_1 = k, \xi_2 = n - k) = e^{-\lambda_1 - \lambda_2} \frac{\lambda_1^k \lambda_2^{n-k}}{k!(n-k)!},$$

$$q(n) = P(\eta = n) = e^{-\lambda_1 - \lambda_2} \frac{(\lambda_1 + \lambda_2)^n}{n!}.$$

Поэтому искомая условная плотность (вероятность) равна

$$f(k|n) = P(\xi_1 = k | \eta = n) = \frac{f(k, n)}{q(n)} = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k},$$

где $p = \lambda_1 / (\lambda_1 + \lambda_2)$. Таким образом, условное распределение ξ_1 при фиксированной сумме $\xi_1 + \xi_2 = n$ есть биномиальное распределение. В частности, если ξ_1, \dots, ξ_r независимы, $\xi_j \in \Pi_{\lambda_j}$, то распределение ξ_1 при фиксированной сумме $\xi_1 + \dots + \xi_r = n$ будет равно $B_{1/r}^n$ и от λ зависеть не будет.

Ответ на тот же вопрос, что и в примере 19, для нормально распределенных случайных величин можно получить, используя следующий

Пример 20. Пусть Φ_{α, σ^2} есть двумерное нормальное распределение величин ξ_1 и ξ_2 , где $\alpha = (\alpha_1, \alpha_2)$, $\alpha_i = E\xi_i$, $\sigma^2 = \|\sigma_{ij}\|$, $\sigma_{ij} = E(\xi_i - \alpha_i)(\xi_j - \alpha_j)$, $i, j = 1, 2$. Определитель матрицы вторых моментов равен

$$|\sigma^2| = \sigma_{11}\sigma_{22} - \sigma_{12}^2 = \sigma_{11}\sigma_{22}(1 - \rho^2),$$

где ρ есть коэффициент корреляции между ξ_1 и ξ_2 . Таким образом, если $|\rho| \neq 1$, то матрица вторых моментов не вырождена и для нее существует обратная матрица

$$A = (\sigma^2)^{-1} = \frac{1}{|\sigma^2|} \begin{vmatrix} \sigma_{22} & -\sigma_{12} \\ -\sigma_{12} & \sigma_{11} \end{vmatrix} = \frac{1}{1 - \rho^2} \begin{vmatrix} \frac{1}{\sigma_{11}} & -\frac{\rho}{\sqrt{\sigma_{11}\sigma_{22}}} \\ \frac{\rho}{\sqrt{\sigma_{11}\sigma_{12}}} & \frac{1}{\sigma_{22}} \end{vmatrix}.$$

Следовательно, совместная плотность ξ_1 и ξ_2 (относительно меры Лебега) равна (см. § 3.3)

$$f(x, y) = \frac{1}{2\pi\sigma_{11}\sigma_{22}\sqrt{1-\rho^2}} \times \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[\frac{(x-\alpha_1)^2}{\sigma_{11}} - \frac{2\rho(x-\alpha_1)(y-\alpha_2)}{\sqrt{\sigma_{11}\sigma_{22}}} + \frac{(y-\alpha_2)^2}{\sigma_{22}} \right] \right\}. \quad (24)$$

Одномерные плотности ξ_1 и ξ_2 соответственно равны

$$f(x) = \frac{1}{\sqrt{2\pi\sigma_{11}}} e^{-(x-\alpha_1)^2/(2\sigma_{11})}, \quad q(y) = \frac{1}{\sqrt{2\pi\sigma_{22}}} e^{-(y-\alpha_2)^2/(2\sigma_{22})}. \quad (25)$$

Поэтому условная плотность ξ_1 при условии $\xi_2 = y$ равна

$$f(x|y) = \frac{f(x, y)}{q(y)} = \frac{1}{\sqrt{2\pi\sigma_{11}(1-\rho^2)}} \exp \left\{ -\frac{1}{2\sigma_{11}(1-\rho^2)} \left(x - \alpha_1 - \rho\sqrt{\frac{\sigma_{11}}{\sigma_{22}}}(y - \alpha_2) \right)^2 \right\},$$

это есть плотность нормального распределения со средним $\alpha_1 + \rho\sqrt{\sigma_{11}/\sigma_{22}}(y - \alpha_2)$ и дисперсией $\sigma_{11}(1 - \rho^2)$.

Отсюда следует, что $f(x|y)$ при $\rho = 0$ совпадает с безусловной плотностью $f(x)$ (и, значит, ξ_1 и ξ_2 независимы) и что у. м. о. ξ_1 относительно ξ_2 равно

$$E(\xi_1 | \xi_2) = \alpha_1 + \rho \sqrt{\frac{\sigma_{11}}{\sigma_{22}}} (\xi_2 - \alpha_2).$$

Прямую $x = \alpha_1 + \rho \sqrt{\sigma_{11}/\sigma_{22}} (y - \alpha_2)$ называют *линией регрессии* ξ_1 на ξ_2 . Она дает наилучшие среднеквадратические приближения величины ξ_1 при данной $\xi_2 = y$.

Пример 21. Рассмотрим задачу о вычислении плотности случайной величины $\xi = \varphi(\zeta, \eta)$, где ζ и η независимы. Из формулы (22) при $A = \mathbb{R}^k$ вытекает, что плотность $f(x)$ распределения ξ выражается через условную плотность $f(x|y)$ равенством

$$f(x) = \int f(x|y) P(\eta \in dy).$$

Применительно к рассматриваемой задаче под $f(x|y)$ надо понимать плотность случайной величины $\varphi(\zeta, y)$, поскольку $P(\xi \in B | \eta = y) = P(\varphi(\zeta, y) \in B)$.

Пример 22. Стрельба по цели с коррекцией. Производится стрельба из орудия по цели с известной геометрической конфигурацией. Введем систему полярных координат, в начале которых располагается орудие. Расстояние r (рис. 6) от орудия до некоторой точки, которую считают центром цели, известно расчету орудия точно; азимут неизвестен, но имеется корректировщик, который после первого пробного выстрела сообщает расчету, каково отклонение по азимуту от точки попадания до центра цели.

Предположим, что рассеивание снарядов при стрельбе (т. е. отклонение (ξ, η) точки попадания от точки прицеливания) описывается двумерным нормальным распределением с плотностью $f(x, y)$, определенной в (24) при $\alpha = 0$. В § 4 главы 8 мы сможем объяснить, почему отклонение (ξ, η) нормально. Мы пренебрегаем здесь тем обстоя-

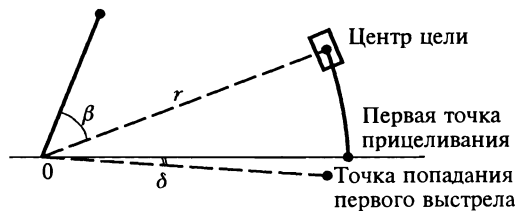


Рис. 6.

тельством, что отклонение η по азимуту не может превосходить π , а отклонение ξ по удаленности не может быть больше r . (Обычно стандартные отклонения σ_1 и σ_2 очень малы по сравнению соответственно с r и π , и отмеченное обстоятельство не имеет значения.) Если бы вместе с удалением r был бы точно известен и азимут β центра цели, то тогда вероятность попадания в цель была бы равна

$$\iint_{B(r, \beta)} f(x, y) dx dy,$$

где $B(r, \beta) = \{(x, y) : (r + x, \beta + y) \in B\}$, множество B есть цель. Но расчету орудия азимут сообщается корректировщиком по результату пробного выстрела (точку попадания он отождествляет с точкой прицеливания), т. е. сообщается с погрешностью δ , распределенной по нормальному закону с плотностью $q(y)$ (см. (25)). Какова вероятность в этих условиях события A , что вторым выстрелом

орудие поразит цель? Если $\delta = z$, то азимут сообщается с погрешностью z и

$$P(A|\delta = z) = \iint_{B(r,\beta)} f(x, y - z) dx dy \equiv \varphi(z).$$

Следовательно,

$$P(A) = E[P(A|\delta)] = E\varphi(\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{-z^2/2\sigma^2} \varphi(z) dz.$$

Пример 23. Отрезок $[0, 1]$ «наудачу» разламывается (т. е. с равномерным распределением точки излома) на две части. Затем большая часть снова «наудачу» делится на две части. Какова вероятность, что из полученных отрезков можно составить треугольник?

Эту задачу можно решать, конечно, и прямым «геометрическим» способом. Надо построить вероятностное пространство $\langle \Omega, \mathfrak{F}, P \rangle$, где в качестве Ω выбрать квадрат (как множество значений ω_1, ω_2 , равных расстоянию соответственно точек первого и второго деления от начала координат). Вероятность здесь определяется соотношением $P(A) = \iint_A f(x, y) dx dy$, где

$$f(x, y) = \begin{cases} 0, & \text{если } x < \frac{1}{2}, y < x \text{ или } x \geq \frac{1}{2}, y \geq x, \\ \frac{1}{1-x}, & \text{если } x < \frac{1}{2}, y \geq x, \\ \frac{1}{x}, & \text{если } x \geq \frac{1}{2}, y < x. \end{cases}$$

Остается вычислить интеграл от этой функции по множеству $B = B_1 \cup B_2$, где

$$B_1 = \left\{ x < \frac{1}{2}; 2(y - x) < 1, 2y > 1 \right\},$$

$$B_2 = \left\{ x > \frac{1}{2}; 2(x - y) < 1, 2y < 1 \right\},$$

так как треугольник можно построить в том и только в том случае, если $(\omega_1, \omega_2) \in B = B_1 \cup B_2$.

Использование формулы

$$P(B) = E[E(I(B) | \omega_1)] = EP(B | \omega_1)$$

дает несколько иной путь решения задачи. Здесь для сокращения записи использована буква B также для обозначения события $\{(\omega_1, \omega_2) \in B\}$. Так как ω_1 распределено равномерно на $[0, 1]$, то надо вычислить лишь $P(B | \omega_1)$. Если $\omega_1 < 1/2$, то ω_2 распределено равномерно на $[\omega_1, 1]$. Треугольник можно построить, если $1/2 < \omega_2 < 1/2 + \omega_1$. Следовательно, $P(B | \omega_1) = \omega_1(1 - \omega_1)$. Из соображений симметрии легко находим, что при $\omega_1 > 1/2$

$$P(B | \omega_1) = \frac{1 - \omega_1}{\omega_1}.$$

Следовательно,

$$P(B) = 2 \int_0^{1/2} \frac{x}{1-x} dx = -1 + 2 \int_0^{1/2} \frac{dx}{1-x} = -1 + 2 \ln 2.$$

Все приведенные выше примеры относились к у. м. о. относительно *случайных величин* (а не σ -подалгебр).

Потребность в у. м. о. относительно σ -алгебр возникает там, где с помощью у. м. о. относительно случайных величин обойтись уже трудно. Пусть, например, некоторый процесс описывается последовательностью случайных величин $\{\xi_j\}_{j=-\infty}^{\infty}$, которые не являются независимыми. Тогда описывать распределение ξ_1 , когда фиксирована вся «предыстория» (т. е. значения $\xi_0, \xi_{-1}, \xi_{-2}, \dots$), удобнее всего с помощью условного распределения ξ_1 относительно $\sigma(\xi_0, \xi_{-1}, \dots)$. Ограничиться условными распределениями относительно случайных величин здесь уже трудно. Соответствующие примеры содержатся в главах 12, 14–21.

Г л а в а 5

Последовательность независимых испытаний с двумя исходами

§ 1. Законы больших чисел

Пусть имеется последовательность испытаний, в каждом из которых некоторое событие A появляется с вероятностью p независимо от исходов других испытаний. Образует последовательность случайных величин следующим образом. Положим $\xi_k = 1$, если событие A в k -м испытании произошло, и $\xi_k = 0$ в противном случае. Тогда $\{\xi_k\}_{k=1}^{\infty}$ будет последовательностью независимых случайных величин, одинаково распределенных по закону Бернулли: $P(\xi_k = 1) = p$, $P(\xi_k = 0) = q = 1 - p$, $E\xi_k = p$, $D\xi_k = pq$. Сумма $S_n = \xi_1 + \dots + \xi_n \in B_p^n$ есть не что иное, как число появлений события A в n первых испытаниях. Очевидно, $ES_n = np$, $DS_n = npq$.

Следующее утверждение носит название *закона больших чисел* для схемы Бернулли.

Теорема 1. Для любого $\varepsilon > 0$

$$P\left(\left|\frac{S_n}{n} - p\right| > \varepsilon\right) \rightarrow 0 \quad \text{при } n \rightarrow \infty.$$

Это утверждение есть частный случай теоремы 4.7. Можно получить и более сильный результат.

Теорема 2 (усиленный закон больших чисел для схемы Бернулли). Для любого $\varepsilon > 0$ при $n \rightarrow \infty$

$$P\left(\sup_{k \geq n} \left|\frac{S_k}{k} - p\right| > \varepsilon\right) \rightarrow 0.$$

Смысл этого факта состоит в том, что введенное нами в гл. 1, 2 определение вероятности соответствует интуитивному пониманию вероятности как предела частоты появления события. Ведь S_n/n можно рассматривать как частоту появления события A , для которого $P(A) = p$. И оказалось, что в известном смысле S_n/n неограниченно сближается с p .

Доказательство теоремы 2. Имеем

$$\begin{aligned} P\left(\sup_{k \geq n} \left|\frac{S_k}{k} - p\right| > \varepsilon\right) &= P\left(\bigcup_{k=n}^{\infty} \left\{\left|\frac{S_k}{k} - p\right| > \varepsilon\right\}\right) \leq \\ &\leq \sum_{k=n}^{\infty} P\left(\left|\frac{S_k}{k} - p\right| > \varepsilon\right) \leq \sum_{k=n}^{\infty} \frac{E(S_k - kp)^4}{k^4 \varepsilon^4}. \end{aligned} \quad (1)$$

Здесь мы опять воспользовались неравенством Чебышева, но с четвертыми моментами. Раскрывая скобки, находим

$$\begin{aligned} E(S_k - kp)^4 &= E\left(\sum_{j=1}^k (\xi_j - p)\right)^4 = \\ &= \sum_{j=1}^k E(\xi_j - p)^4 + 6 \sum_{i < j} E(\xi_i - p)^2 (\xi_j - p)^2 = \\ &= k(pq^4 + qp^4) + 3k(k-1)(pq)^2 \leq k + k(k-1) = k^2. \end{aligned} \quad (2)$$

Таким образом, оцениваемая вероятность не превосходит $\varepsilon^{-4} \sum_{k=n}^{\infty} k^{-2} \rightarrow 0$ при $n \rightarrow \infty$. ■

Нетрудно видеть, что если бы мы в (1) воспользовались неравенством Чебышева со вторыми моментами, то требуемой оценки не получили бы.

Можно отметить также, что на самом деле для вероятностей $P(|S_k - kp| > \varepsilon k)$ верны и более сильные оценки, чем те, которыми мы пользовались. Они получены в § 5.

Следствие 1. Если $f(x)$ — непрерывная функция на $[0, 1]$, то при $n \rightarrow \infty$ равномерно по p

$$Ef\left(\frac{S_n}{n}\right) \rightarrow f(p). \quad (3)$$

Доказательство следует из того, что при любом $\varepsilon > 0$

$$\begin{aligned} E\left|f\left(\frac{S_n}{n}\right) - f(p)\right| &\leq E\left(\left|f\left(\frac{S_n}{n}\right) - f(p)\right|; \left|\frac{S_n}{n} - p\right| \leq \varepsilon\right) + \\ &+ E\left(\left|f\left(\frac{S_n}{n}\right) - f(p)\right|; \left|\frac{S_n}{n} - p\right| > \varepsilon\right) \leq \sup_{|x| \leq \varepsilon} |f(p+x) - f(p)| + o(1). \end{aligned} \quad \blacksquare$$

Следствие 2. Если $f(x)$ непрерывна на $[0, 1]$, то при $n \rightarrow \infty$

$$\sum_{k=0}^n f\left(\frac{k}{n}\right) \binom{n}{k} x^k (1-x)^{n-k} \rightarrow f(x)$$

равномерно по $x \in [0, 1]$.

Это соотношение есть просто другая запись (3), так как $P(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k}$ (см. главу 1). Из него следует известная *теорема Вейерштрасса* о приближении непрерывных функций полиномами. При этом нужные полиномы построены в явном виде — это *полиномы Бернштейна*.

§ 2. Локальная предельная теорема

Мы знаем, что $P(S_n = k) = \binom{n}{k} p^k q^{n-k}$, $q = 1 - p$. Однако эта формула при больших n и k становится весьма неудобной для вычислений, в связи с чем возникает вопрос об асимптотическом поведении вероятности $P(S_n = k)$ при $n \rightarrow \infty$.

В дальнейшем символ $a_n \sim b_n$, где $\{a_n\}$ и $\{b_n\}$ — две числовые последовательности, будет означать, что $a_n/b_n \rightarrow 1$ при $n \rightarrow \infty$. При этом такие последовательности $\{a_n\}$ и $\{b_n\}$ будем называть эквивалентными.

Обозначим

$$H(x) = x \ln \frac{x}{p} + (1-x) \ln \frac{1-x}{1-p}, \quad p^* = \frac{k}{n}. \quad (4)$$

Теорема 3. При $k \rightarrow \infty$, $n-k \rightarrow \infty$

$$P(S_n = k) = P\left(\frac{S_n}{n} = p^*\right) \sim \frac{1}{\sqrt{2\pi np^*(1-p^*)}} \exp\{-nH(p^*)\}. \quad (5)$$

Доказательство. Воспользуемся формулой Стирлинга, в силу которой $n! \sim \sqrt{2\pi n} n^n e^{-n}$ при $n \rightarrow \infty$. Тогда

$$\begin{aligned} P(S_n = k) &= \binom{n}{k} p^k q^{n-k} \sim \sqrt{\frac{n}{2\pi k(n-k)}} \frac{n^n}{k^k (n-k)^{n-k}} p^k (1-p)^{n-k} = \\ &= \frac{1}{\sqrt{2\pi np^*(1-p^*)}} \exp\left\{-k \ln \frac{k}{n} - (n-k) \ln \frac{n-k}{n} + k \ln p + (n-k) \ln(1-p)\right\} = \\ &= \frac{1}{\sqrt{2\pi np^*(1-p^*)}} \exp\left\{-n[p^* \ln p^* + (1-p^*) \ln(1-p^*) - \right. \\ &\quad \left. - p^* \ln p - (1-p^*) \ln(1-p)]\right\} = \frac{1}{\sqrt{2\pi np^*(1-p^*)}} \exp\{-nH(p^*)\}. \quad \blacksquare \end{aligned}$$

Если $p^* = k/n$ близко к p , то для правой части в (5) можно найти другую форму записи, представляющую значительный интерес. Заметим, что функция $H(x)$ является аналитической в интервале $(0, 1)$. Так как

$$H'(x) = \ln \frac{x}{p} - \ln \frac{1-x}{1-p}, \quad H''(x) = \frac{1}{x} + \frac{1}{1-x}, \quad (6)$$

то $H(p) = H'(p) = 0$ и при $p^* - p \rightarrow 0$ ¹⁾

$$H(p^*) = \frac{1}{2} \left(\frac{1}{p} + \frac{1}{q} \right) (p^* - p)^2 + O(|p^* - p|^3).$$

Поэтому если $p^* \sim p$ и $n(p^* - p)^3 \rightarrow 0$, то

$$P(S_n = k) \sim \frac{1}{\sqrt{2\pi npq}} \exp\left\{-\frac{n}{2pq} (p^* - p)^2\right\}.$$

Если положить $\Delta = 1/\sqrt{npq}$, $\varphi(x) = 1/\sqrt{2\pi} e^{-x^2/2}$, то получим

Следствие 3. Если $z = n(p^* - p) = k - np = o(n^{2/3})$, то

$$P(S_n = k) = P(S_n - np = z) \sim \varphi(z\Delta)\Delta. \quad (7)$$

Эта формула позволит оценивать вероятности и событий вида $\{S_n < k\}$.

Если p^* существенно отличается от p , то оценка вероятностей указанных событий может быть произведена с помощью результатов § 1.3.

¹⁾ В соответствии с общепринятыми обозначениями будем писать $a(z) = o(b(z))$ при $z \rightarrow z_0$, если $b(z) > 0$, $\lim_{z \rightarrow z_0} \frac{a(z)}{b(z)} = 0$, и $a(z) = O(b(z))$ при $z \rightarrow z_0$, если $b(z) > 0$, $\limsup_{z \rightarrow z_0} \frac{|a(z)|}{b(z)} < \infty$.

Пример 1. В жюри, состоящем из нечетного числа $n = 2m + 1$ членов, каждый человек независимо от других принимает правильное решение с вероятностью $p = 0,7$. Каково минимальное число членов жюри, при котором решение, принимаемое большинством голосов, будет справедливым с вероятностью, не меньшей, чем 0,99?

Положим в этой задаче $\xi_k = 1$, если k -й член жюри принял правильное решение, и $\xi_k = 0$ в противном случае. Нас интересуют нечетные значения n , при которых $P(S_n \leq m) \leq 0,01$. Несомненно, что высокая точность принятого решения может быть достигнута лишь при больших значениях n . В этом случае, как мы установили в § 1.3, вероятность $P(S_n \leq m)$ равна приближенно

$$\frac{(n+1-m)p}{(n+1)p-m} P(S_n = m) \approx \frac{p}{2p-1} P(S_n = m).$$

Используя теорему 3 и тот факт, что в нашей задаче $p^* \approx \frac{1}{2}$, $H(\frac{1}{2}) = -\frac{1}{2} \ln 4p(1-p)$, $H'(\frac{1}{2}) = \ln \frac{1-p}{p}$, получим

$$\begin{aligned} P(S_n \leq m) &\approx \frac{p}{2p-1} \sqrt{\frac{2}{\pi n}} \exp \left\{ -nH \left(\frac{1}{2} - \frac{1}{2n} \right) \right\} \approx \\ &\approx \frac{p}{2p-1} \sqrt{\frac{2}{\pi n}} \exp \left\{ -nH \left(\frac{1}{2} \right) + \frac{1}{2} H' \left(\frac{1}{2} \right) \right\} \approx \\ &\approx \frac{\sqrt{2p(1-p)}}{(2p-1)\sqrt{\pi n}} \left(\sqrt{4p(1-p)} \right)^n \approx 0,915 \frac{1}{\sqrt{n}} (0,84)^{n/2}. \end{aligned}$$

Справа здесь стоит монотонно убывающая функция $a(n)$. Решая уравнение $a(n) = 0,01$, в качестве ответа получим $n = 33$. Такой же результат будет, если мы воспользуемся точными формулами.

Нетрудно оценить погрешность, возникающую в формуле (5). Если в формуле Стирлинга $n! = \sqrt{2\pi n} n^n e^{-n+\theta(n)}$ воспользоваться известными неравенствами²⁾ $1/(12n+1) < \theta(n) < 1/12n$, то с помощью прежних рассуждений получим следующее уточнение теоремы 3.

Теорема 4.

$$P(S_n = k) = \frac{1}{\sqrt{2\pi n p^*(1-p^*)}} \exp\{-nH(p^*) + \theta(k, n)\}, \quad (8)$$

где

$$|\theta(k, n)| = |\theta(n) - \theta(k) \theta(n-k)| < \frac{1}{12k} + \frac{1}{12(n-k)} = \frac{1}{12np^*(1-p)^*}. \quad (9)$$

Соотношение (7) можно уточнить следующим образом.

Теорема 5. При всех k таких, что $|p^* - p| \leq \frac{1}{2} \min(p, q)$, имеет место равенство $P(S_n = k) = \varphi(z\Delta)\Delta(1 + \varepsilon(k, n))$, где

$$1 + \varepsilon(k, n) = \exp \left\{ \theta \left(\frac{|z|^3 \Delta^4}{3} + \left(|z| + \frac{1}{6} \right) \Delta^2 \right) \right\}, \quad |\theta| < 1.$$

Как это нетрудно усмотреть из свойств разложения в ряд функции e^x , порядок малости $\varepsilon(k, n)$ в приведенных формулах совпадает с порядком малости показателя

²⁾ См., например, [22, т. 1, с. 66].

экспоненты. Из теоремы 5 следует поэтому, что при $z = k - np = o(\Delta^{-4/3})$ или, что то же, при $z = o(n^{2/3})$ справедливо (7).

Доказательство. Воспользуемся утверждением теоремы 4. В дополнение к формулам (6) можем записать

$$H^{(k)}(x) = \frac{(-1)^k (k-2)!}{x^{k-1}} + \frac{(k-2)!}{(1-x)^{k-1}}, \quad k \geq 2,$$

$$H(p^*) = \frac{1}{2pq}(p^* - p)^2 + R_1,$$

где остаток $R_1 = \sum_{k=3}^{\infty} \frac{H^{(k)}(p)}{k!} (p^* - p)^k$ мы оценим. Учитывая, что

$$|H^{(k)}(p)| \leq (k-2)! \left(\frac{1}{p^{k-1}} + \frac{1}{q^{k-1}} \right), \quad k \geq 2,$$

и обозначая для краткости $|p^* - p| = \delta$, получим при $\delta \leq \frac{1}{2} \min(p, q)$

$$\begin{aligned} |R_1| &\leq \sum_{k=3}^{\infty} \frac{(k-2)!}{k!} \delta^k \left(\frac{1}{p^{k-1}} + \frac{1}{q^{k-1}} \right) \leq \\ &\leq \frac{\delta^3}{6} \left[\frac{1}{p^2} \frac{1}{1 - \delta/p} + \frac{1}{q^2} \frac{1}{1 - \delta/q} \right] \leq \frac{\delta^3}{6} \left[\frac{2}{p^2} + \frac{2}{q^2} \right] < \frac{\delta^3}{3(pq)^2}. \end{aligned}$$

Отсюда следует, что

$$-nH(p^*) = -\frac{(k-np)^2}{2npq} + \frac{\theta_1 |k-np|^3}{3(npq)^2} = -\frac{z^2 \Delta^2}{2} + \frac{\theta_1 |z|^3 \Delta^4}{3}, \quad |\theta_1| < 1. \quad (10)$$

Обратимся теперь к остальным множителям в равенстве (8) и рассмотрим произведение $p^*(1-p^*)$. Так как $-p \leq 1-p-p^* \leq 1-p$, то

$$|p^*(1-p^*) - p(1-p)| = |(p^* - p)(1-p-p^*)| \leq |p^* - p| \max(p, q).$$

Отсюда следует, в частности, что при $|p^* - p| < 1/2 \min(p, q)$ выполняется

$$|p^*(1-p^*) - pq| < \frac{1}{2}pq, \quad p^*(1-p^*) > \frac{1}{2}pq.$$

Поэтому наряду с (9) можно написать, что при значениях k , указанных в теореме 4,

$$|\theta(k, n)| < \frac{1}{6npq} = \frac{\Delta^2}{6}. \quad (11)$$

Осталось рассмотреть множитель $[p^*(1-p^*)]^{-1/2}$. Так как при $|\gamma| < 1/2$

$$|\ln(1+\gamma)| = \left| \int_1^{1+\gamma} \frac{1}{x} dx \right| < 2|\gamma|,$$

то при $\delta = |p^* - p| < 1/2 \min(p, q)$ будут справедливы соотношения

$$\begin{aligned}
\ln p^*(1-p^*) &= \ln pq + \ln \left(1 + \frac{p^*(1-p^*) - pq}{pq} \right) = \\
&= \ln pq + \ln \left(1 - \frac{\theta^* \delta}{pq} \right), \quad |\theta^*| < \max(p, q); \\
\ln \left(1 - \frac{\theta^* \delta}{pq} \right) &= -\frac{2\theta_2 \delta}{pq}, \quad |\theta_2| < \max(p, q); \\
[p^*(1-p^*)]^{-1/2} &= [pq]^{-1/2} \exp \frac{\theta_2 \delta}{pq}.
\end{aligned} \tag{12}$$

Используя представления (10)–(12) и утверждение теоремы 3, закончим доказательство. ■

Из приведенных оценок видно, что границы для числа θ в формулировке теоремы 5 могут быть сужены, если рассматривать меньшие отклонения $|p^* - p|$, скажем, до значения $\alpha \min(p, q)$, где $\alpha < 1/2$.

Найденные соотношения для $P(S_n = k)$ представляют собой так называемые *локальные теоремы для схемы Бернулли*.

Основная асимптотическая формула, изложенная в теореме 3, допускает естественное обобщение на полиномиальное распределение \mathbf{B}_p^n , $p = (p_1, \dots, p_r)$, когда в последовательности независимых испытаний при каждом испытании наступает не один из двух, а один из $r \geq 2$ возможных исходов A_1, \dots, A_r , вероятности которых равны соответственно p_1, \dots, p_r . Пусть $S_n^{(j)}$ есть число появлений события A_j в n испытаниях,

$$S_n = (S_n^{(1)}, \dots, S_n^{(r)}), \quad k = (k_1, \dots, k_r), \quad p^* = \frac{k}{n},$$

и положим $H(x) = \sum x_i \ln x_i / p_i$, $x = (x_1, \dots, x_r)$. Очевидно, что $S_n \in \mathbf{B}_p^n$. Следующее утверждение является прямым обобщением теоремы 3.

Теорема 6. Если каждое из r переменных k_1, \dots, k_r либо равно нулю, либо стремится к ∞ при $n \rightarrow \infty$, то

$$P(S_n = k) \sim (2\pi n)^{(1-r_0)/2} \left(\prod_{\substack{j=1 \\ p_j^* \neq 0}}^r p_j^* \right)^{-1/2} \exp\{-nH(p^*)\},$$

где r_0 — число переменных k_1, \dots, k_r , не равных нулю.

Доказательство. Как и в доказательстве теоремы 3, воспользуемся формулой Стирлинга

$$n! \sim \sqrt{2\pi n} e^{-n} n^n$$

при $n \rightarrow \infty$. Считая, не ограничивая общности, что все $k_j \rightarrow \infty$, $j = 1, \dots, r$, получим

$$\begin{aligned}
P(S_n = k) &\sim (2\pi)^{(1-r)/2} \left(\frac{n}{\prod_{j=1}^r k_j} \right)^{1/2} \prod_{j=1}^r \left(\frac{np_j}{k_j} \right)^{k_j} = \\
&= (2\pi n)^{(1-r)/2} \left(\prod_{j=1}^r p_j^* \right)^{-1/2} \exp \left\{ n \sum_{j=1}^r \frac{k_j}{n} \ln \frac{p_j n}{k_j} \right\}.
\end{aligned} \quad \blacksquare$$

§ 3. Теорема Муавра—Лапласа и ее уточнения

Пусть a и b — фиксированные числа, $\zeta_n = (S_n - np)/\sqrt{npq}$. Тогда

$$P(a < \zeta_n < b) = \sum_{a\sqrt{npq} < z < b\sqrt{npq}} P(S_n - np = z).$$

Если здесь вместо $P(S_n - np = z)$ подставить $\varphi(z\Delta)\Delta$, то получим интегральную сумму $\sum_{a < z\Delta < b} \varphi(z\Delta)\Delta$, соответствующую интегралу $\int_a^b \varphi(x) dx$.

Таким образом, соотношения (7) делают правдоподобным равенство

$$\lim_{n \rightarrow \infty} P(a < \zeta_n < b) = \int_a^b \varphi(x) dx = \Phi(b) - \Phi(a), \quad (13)$$

где $\Phi(x)$ есть функция нормального распределения с параметрами $(0, 1)$:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt.$$

Это есть *теорема Муавра—Лапласа*, относящаяся к *интегральным предельным теоремам*. В главе 7 мы получим более общие интегральные теоремы, из которых (13) будет следовать в качестве частного случая.

Теорема 5 позволяет получить (13) вместе с оценкой погрешности или, иначе говоря, с *оценкой скорости сходимости*.

Пусть A и B — целые числа,

$$a = \frac{A - np}{\sqrt{npq}}, \quad b = \frac{B - np}{\sqrt{npq}}. \quad (14)$$

Теорема 7. Пусть $b > a$, $c = \max(|a|, |b|)$,

$$\rho = \frac{c^3 + 3c}{3} \Delta + \frac{\Delta^2}{6}.$$

Тогда если $\Delta = 1/\sqrt{npq} \leq 1/2$, $\rho \leq 1/2$, то

$$P(A \leq S_n < B) = P(a \leq \zeta_n < b) = \int_a^b \varphi(t) dt (1 + \theta_1 \Delta c)(1 + 2\theta_2 \rho), \quad (15)$$

где $|\theta_i| \leq 1$, $i = 1, 2$.

Эта теорема показывает, что левая часть в (15) может быть эквивалентной $\Phi(b) - \Phi(a)$ и при растущих a и b . В этих условиях $\Phi(b) - \Phi(a)$ может сходиться к нулю, и знание относительной погрешности в (13) более удобно, так как ее малость всегда позволяет установить малость и абсолютной погрешности, но не наоборот.

Доказательство. Заметим прежде всего, что при всех k таких, что $|z| = |k - np| < c\sqrt{npq}$, будут выполнены условия теоремы 5. Действительно, для выполнения неравенства $|p^* - p| < 1/2 \min(p, q)$ достаточно, чтобы $|k - np| < (npq)/2 = 1/(2\Delta^2)$. Это неравенство будет выполнено, если $c < 1/(2\Delta)$. Но в силу того, что $\rho \leq 1/2$, имеем

$$\frac{c(c^3 + 3)\Delta}{3} < \frac{1}{2}, \quad c\Delta < \frac{1}{2}.$$

Таким образом, при каждом k таком, что $a\sqrt{npq} \leq z < b\sqrt{npq}$, мы можем воспользоваться теоремой 5, в силу которой

$$\begin{aligned} P(A \leq S_n < B) &= \sum_{a\sqrt{npq} \leq z < b\sqrt{npq}} P(S_n = k) = \\ &= \sum_{a \leq z\Delta < b} \varphi(z\Delta)\Delta \left[1 + \left(\exp \left\{ \theta \left(\frac{|z|^3 \Delta^4}{3} + \left(|z| + \frac{1}{6} \right) \Delta^2 \right) \right\} - 1 \right) \right], \end{aligned} \quad (16)$$

где $|\theta| < 1$. Так как при $\rho \leq 1$

$$\left| \frac{e^\rho - 1}{\rho} \right| < (e - 1) < 2,$$

то абсолютное значение поправочного члена в (16) не превосходит (подставим туда $z\Delta = c$)

$$\left| \exp \left\{ \theta \left(\frac{c^3 \Delta}{3} + c\Delta + \frac{\Delta^2}{6} \right) \right\} - 1 \right| \leq 2\theta \left(\frac{c^3 \Delta}{3} + c\Delta + \frac{\Delta^2}{6} \right) = 2\theta\rho.$$

Поэтому

$$P(A \leq S_n < B) = \sum_{a \leq z\Delta < b} \varphi(z\Delta)\Delta [1 + 2\theta_1 \rho], \quad (17)$$

где $|\theta_1| < 1$.

Преобразуем теперь сумму, стоящую в правой части полученного равенства. Для этого заметим, что для любой гладкой функции $\varphi(x)$

$$\left| \Delta\varphi(x) - \int_x^{x+\Delta} \varphi(t) dt \right| \leq \frac{\Delta^2}{2} \max_{x \leq t \leq x+\Delta} |\varphi'(t)|. \quad (18)$$

Но для функции $\varphi(x) = 1/\sqrt{2\pi} e^{-x^2/2}$ имеем $\varphi'(x) = -x\varphi(x)$, и наибольшее значение $\varphi(t)$ на отрезке $[x, x + \Delta]$, $|x| \leq c$, отличается от наименьшего не более чем на множитель $\exp \left\{ c\Delta + \Delta^2/2 \right\}$. Поэтому при $|x| \leq c$ в силу (18)

$$\left| \Delta\varphi(x) - \int_x^{x+\Delta} \varphi(t) dt \right| \leq \frac{\Delta^2 c}{2} e^{c\Delta + \Delta^2/2} \min_{x \leq t \leq x+\Delta} \varphi(t) \leq \frac{\Delta c}{2} e^{c\Delta + \Delta^2/2} \int_x^{x+\Delta} \varphi(t) dt.$$

Так как $c\Delta + \Delta^2/2 < 1/2 + 1/8$, $e^{c\Delta + \Delta^2/2} \leq 2$, то справедливо представление

$$\Delta\varphi(x) = \int_x^{x+\Delta} \varphi(t) dt (1 + \theta_1 \Delta c), \quad |\theta_1| < 1.$$

Подставляя его в (17), мы получим утверждение теоремы. ■

Таким образом, по теореме 7 разность

$$|P(x \leq \zeta_n < y) - (\Phi(y) - \Phi(x))| \quad (19)$$

эффективно, хотя и весьма грубо, оценивается сверху величиной порядка Δ , или, что то же, величиной порядка $1/\sqrt{npq}$, если только $x = a$, $y = b$ (под a и b мы понимаем значения (см. (14)), представимые в виде $(k - np)\Delta$, где k целое). Если x и y не принадлежат указанной решетке с шагом Δ , то погрешность (19) будет оцениваться величиной того же порядка, так как, например, при изменении y вероятность $P(a \leq \zeta_n < y)$ будет постоянной на полуинтервалах вида $(a + k\Delta, a + (k+1)\Delta)$, а функция $\Phi(y) - \Phi(x)$ монотонно возрастает с ограниченной производной. Аналогичные доводы справедливы и относительно левой границы x . Теперь важно отметить, что порядок оценки $1/\sqrt{npq}$ не может быть улучшен, так как именно такой порядок имеют по теореме 5 скачки функции распределения ζ_n .

Теорема 7 позволяет использовать нормальное приближение для $P(x \leq \zeta_n < y)$ и в так называемой области больших отклонений, когда x и y возрастают по абсолютной величине и имеют одинаковые знаки. В этом случае $\Phi(y) - \Phi(x)$ и приближаемая вероятность сходятся к нулю. Поэтому приближение естественно считать удовлетворительным, если только

$$P\left(\frac{x \leq \zeta_n < y}{\Phi(y) - \Phi(x)}\right) \rightarrow 1. \quad (20)$$

Как показывает теорема 7, такая сходимость будет иметь место, когда $c = \max(|x|, |y|) = o(\Delta^{-1/3})$ или, что то же, $c = o(n^{1/6})$.

При больших значениях c , как это нетрудно проверить с помощью теоремы 3, соотношение (20), вообще говоря, выполняться не будет.

В заключение отметим, что так как при $b \rightarrow \infty$

$$P(|\zeta_n| > b) \rightarrow 0,$$

то из теоремы 7 немедленно следует, что при каждом фиксированном y

$$\lim_{n \rightarrow \infty} P(\zeta_n < y) = \Phi(y).$$

Позже мы покажем, что это утверждение справедливо и в гораздо более широких условиях, когда ζ_n есть нормированная сумма произвольно распределенных случайных величин с конечной дисперсией.

§ 4. Теорема Пуассона и ее уточнения

Как было видно из оценок в предыдущем разделе, теорема Муавра—Лапласа будет обеспечивать хорошее приближение для оцениваемых вероятностей, если число npq (дисперсия S_n) велико. Это число будет расти вместе с n , если p и q положительны и фиксированы. Но как поступать в задаче, где, например, $p = 0,001$, а $n = 1000$, так что $np = 1$? Хотя n здесь велико, в этой задаче применять теорему Муавра—Лапласа бессмысленно. Оказывается, что в этом случае распределение $P(S_n = k)$ можно хорошо приближать распределением Пуассона Π_μ при подходящем значении μ (см. § 4.2). Напомним, что

$$\Pi_\mu(B) = \sum_{0 \leq k \in B} e^{-\mu} \frac{\mu^k}{k!}.$$

Обозначим $np = \mu$.

Теорема 8. Для всех множеств B

$$|P(S_n \in B) - \Pi_\mu(B)| \leq \frac{\mu^2}{n}.$$

Это утверждение мы могли бы доказывать так же, как локальную теорему, пользуясь явной формулой для $P(S_n = k)$. Однако доказательство можно получить более простым и красивым способом, который можно было бы назвать «методом одного вероятностного пространства». Этот метод нередко используется в вероятностных исследованиях и состоит в нашем случае в том, что на том же пространстве, где заданы S_n , строятся случайные величины S_n^* , по возможности близкие к S_n и такие, что распределение S_n^* совпадает с распределением Пуассона.

Важно также, что метод одного вероятностного пространства без каких бы то ни было усложнений допускает обобщение на случай сумм *разнораспределенных* случайных величин, когда вероятность p появления 1 в каждом испытании зависит от номера испытания. Пользуясь этим обстоятельством, докажем здесь более общее утверждение, из которого теорема 8 будет следовать как частный случай.

Итак, пусть имеется последовательность независимых случайных величин ξ_1, \dots, ξ_n ; $\xi_j \in \mathbf{B}_{p_j}$. Обозначим, как и прежде, $S_n = \sum_{j=1}^n \xi_j$. Формулируемая ниже теорема предназначена для оценки $P(S_n = k)$, когда p_j малы, а число $\mu = \sum_{j=1}^n p_j$ «сравнимо с 1».

Теорема 9. Для всех множеств B

$$|P(S_n \in B) - \Pi_\mu(B)| \leq \sum_{j=1}^n p_j^2.$$

Прежде чем переходить к доказательству теоремы 9, отметим одно важное свойство распределения Пуассона, которое нам понадобится: *если η_1 и η_2 независимы, $\eta_2 \in \Pi_{\mu_1}$, $\eta_2 \in \Pi_{\mu_2}$, то $\eta_1 + \eta_2 \in \Pi_{\mu_1 + \mu_2}$* ³⁾.

Действительно, по формуле полной вероятности

$$\begin{aligned} P(\eta_1 + \eta_2 = k) &= \sum_{j=0}^k P(\eta_1 = j)P(\eta_2 = k - j) = \\ &= \sum_{j=0}^k \frac{\mu_1^j e^{-\mu_1}}{j!} \cdot \frac{\mu_2^{k-j} e^{-\mu_2}}{(k-j)!} = \frac{1}{k!} e^{-(\mu_1 + \mu_2)} \sum_{j=0}^k \binom{k}{j} \mu_1^j \mu_2^{k-j} = \frac{(\mu_1 + \mu_2)^k e^{-\mu_1 - \mu_2}}{k!}. \end{aligned}$$

Доказательство теоремы 9. Пусть $\omega_1, \dots, \omega_n$ — независимые случайные величины, каждая из которых есть тождественная функция ($\xi(\omega_k) = \omega_k$) на единичном отрезке с равномерным распределением. Можно считать, следовательно, что вектор $\omega = (\omega_1, \dots, \omega_n)$ задан как тождественная функция на единичном n -мерном кубе Ω с равномерным распределением.

³⁾ Этот факт будет очень просто следовать также из свойств характеристических функций, излагаемых в главе 7.

Построим на Ω случайные величины ξ_j и ξ_j^* следующим образом:

$$\xi_j(\omega) = \begin{cases} 0, & \text{если } \omega_j < 1 - p_j, \\ 1, & \text{если } \omega_j \geq 1 - p_j, \end{cases}$$

$$\xi_j^*(\omega) = \begin{cases} 0, & \text{если } \omega_j < e^{-p_j}, \\ k \geq 1, & \text{если } \omega_j \in [\pi_{k-1}, \pi_k), \end{cases}$$

где $\pi_k = \sum_{m \leq k} e^{-p_j} \frac{(p_j)^m}{m!}$, $k = 0, 1, \dots$.

Очевидно, что $\xi_j(\omega)$ независимы, $\xi_j(\omega) \in \mathbf{B}_{p_j}$; $\xi_j^*(\omega)$ также независимы, $\xi_j^*(\omega) \in \mathbf{P}_{p_j}$. Заметим теперь, что, так как $1 - p_j \leq e^{-p_j}$, то $\xi_j(\omega) \neq \xi_j^*(\omega)$, только если $\omega_j \in [1 - p_j, e^{-p_j})$ или $\omega_j \in [e^{-p_j} + p_j e^{-p_j}, 1]$. Следовательно,

$$P(\xi_j \neq \xi_j^*) = (e^{-p_j} - 1 + p_j) + (1 - e^{-p_j} - p_j e^{-p_j}) = p_j(1 - e^{-p_j}) \leq p_j^2,$$

$$P(S_n \neq S_n^*) \leq P\left(\bigcup_j \{\xi_j \neq \xi_j^*\}\right) \leq \sum p_j^2,$$

где $S_n^* = \sum_{j=1}^n \xi_j^* \in \Pi_\mu$.

Теперь можно записать

$$\begin{aligned} P(S_n \in B) &= P(S_n \in B, S_n = S_n^*) + P(S_n \in B, S_n \neq S_n^*) = \\ &= P(S_n^* \in B) - P(S_n^* \in B, S_n \neq S_n^*) + P(S_n \in B, S_n \neq S_n^*); \\ |P(S_n \in B) - P(S_n^* \in B)| &\leq |P(S_n^* \in B, S_n \neq S_n^*) - \\ &- P(S_n \in B, S_n \neq S_n^*)| \leq P(S_n \neq S_n^*). \end{aligned} \quad (21)$$

Отсюда очевидным образом следует утверждение теоремы. ■

Замечание. Можно указать и другие конструкции одного вероятностного пространства. С помощью одной из них мы выясним сейчас существование более точного пуассоновского приближения для распределения S_n .

Именно, пусть $\xi_j^*(\omega)$ — независимые случайные величины, распределенные по закону Пуассона с параметрами $r_j = -\ln(1 - p_j) \geq p_j$, так что $P(\xi_j^* = 0) = e^{-r_j} = 1 - p_j$. Тогда $\xi_j(\omega) = \min(1, \xi_j^*(\omega)) \in \mathbf{B}_{p_j}$, при этом

$$P\left(\bigcup_{j=1}^n \{\xi_j(\omega) \neq \xi_j^*(\omega)\}\right) \leq \sum_{j=1}^n P(\xi_j^*(\omega) \geq 2) = \sum_{j=1}^n (1 - e^{-r_j} - r_j e^{-r_j}).$$

Но при $r = -\ln(1 - p)$ справедливо неравенство

$$1 - e^{-r} - r e^{-r} = p + (1 - p) \ln(1 - p) \leq p + (1 - p) \left(-p - \frac{p^2}{2}\right) = \frac{p^2}{2}(1 + p).$$

Следовательно, для построенного пуассоновского приближения

$$P(S_n^* \neq S_n) \leq \frac{1}{2} \sum_{j=1}^n p_j^2 (1 + p_j).$$

Положив $\lambda = -\sum_{j=1}^n \ln(1-p_j) \geq \sum_{j=1}^n p_j$, путем прежних рассуждений получаем оценку

$$|P(S_n \in B) - \Pi_\lambda(B)| \leq \frac{1}{2} \sum_{j=1}^n p_j^2 (1+p_j).$$

Эта оценка близости с распределением Пуассона с «чуть сдвинутым» параметром лучше, чем оценка, полученная в теореме 9. Кроме того, можно отметить, что в новом построении $\xi_j \leq \xi_j^*$, $S_n \leq S_n^*$ и, следовательно,

$$P(S_n \geq k) \leq P(S_n^* \geq k) = \Pi_\lambda([k, \infty)).$$

Вернемся к случаю одинаково распределенных ξ_k . Чтобы из теоремы 9 получить теперь предельную теорему типа Муавра—Лапласа (см. (13)), здесь требуется несколько иная, чем прежде, постановка задачи. Ведь для того, чтобы np с ростом n оставалось ограниченным, нам нужно, чтобы $p = P(\xi_k = 1)$ сходилась к нулю, чего нельзя добиться, рассматривая фиксированную последовательность случайных величин ξ_1, ξ_2, \dots .

Мы введем в рассмотрение последовательность возрастающих серий случайных величин:

$$\begin{aligned} &\xi_1^{(1)}; \\ &\xi_1^{(2)}, \xi_2^{(2)}; \\ &\xi_1^{(3)}, \xi_2^{(3)}, \xi_3^{(3)}; \\ &\dots\dots\dots \\ &\xi_1^{(n)}, \xi_2^{(n)}, \xi_3^{(n)}, \dots, \xi_n^{(n)}. \end{aligned}$$

Это так называемая *схема серий*. Индекс сверху означает номер серии. Нижний индекс означает номер случайной величины в серии.

Пусть в n -й серии случайные величины $\xi_k^{(n)}$ независимы, $\xi_k^{(n)} \in \mathbf{B}_{p_n}$, $k = 1, \dots, n$.

Следствие 4 (теорема Пуассона). Если $np_n \rightarrow \mu > 0$ при $n \rightarrow \infty$, то при каждом k

$$P(S_n = k) \rightarrow \Pi_\mu(\{k\}), \quad (22)$$

где $S_n = \xi_1^{(n)} + \dots + \xi_n^{(n)}$.

Это утверждение является очевидным следствием теоремы 8. Его легко получить также непосредственно, заметив, что из равенства $P(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k}$ следует

$$\begin{aligned} P(S_n = 0) &= e^{n \ln(1-p)} \sim e^{-\mu}, \\ \frac{P(S_n = k+1)}{P(S_n = k)} &= \frac{n-k}{k+1} \frac{p}{1-p} \sim \frac{\mu}{k+1}. \end{aligned}$$

Из теоремы 9 вытекает аналог теоремы Пуассона и в более общем случае, когда $\xi_j^{(n)}$ не обязательно одинаково распределены⁴⁾ и могут принимать значения, отличные от нуля и единицы.

⁴⁾ Обобщение теоремы Муавра—Лапласа на случай разнораспределенных величин содержится в центральной предельной теореме § 8.4.

Следствие 5. Пусть $p_{jn} = P(\xi_j^{(n)} = 1)$ зависят от n и j так, что

$$\max_j p_{jn} \rightarrow 0, \quad \sum_{j=1}^n p_{jn} \rightarrow \mu > 0,$$

$$P(\xi_j^{(n)} = 0) = 1 - p_{jn} + o(p_{jn}).$$

Тогда справедливо (22).

Для доказательства следствия надо воспользоваться теоремой 9 и тем фактом, что $P\left(\bigcup_{j=1}^n \{\xi_j^{(n)} \neq 0, \xi_j^{(n)} \neq 1\}\right) \leq \sum_{j=1}^n o(p_{jn}) = o(1)$, означаящим, что с вероятностью, сходящейся к 1, все $\xi_j^{(n)}$ принимают лишь значения нуль или единица.

Наряду с этими утверждениями из теорем 9, 8, очевидно, можно получить и несколько более сильные формулировки: для любого множества B

$$\lim_{n \rightarrow \infty} P(S_n \in B) = \Pi_\mu(B).$$

Отметим, что в условиях теоремы 8 такая сходимость будет иметь место и для значений $np \rightarrow \infty$, но таких, что $np^2 \rightarrow 0$. В то же время уточнение теоремы Муавра—Лапласа в § 3 показывает, что приближение для распределения S_n с помощью нормального закона действует, если $np \rightarrow \infty$ (мы считаем, что $p < q$, так что $npq \rightarrow 1/2np \rightarrow \infty$).

Таким образом, существуют последовательности $p \in \{p: np \rightarrow \infty, np^2 \rightarrow 0\}$, для которых применимы оба приближения — и нормальное, и пуассоновское. Другими словами, области применений нормального и пуассоновского приближений *перекрываются*.

Далее, из теоремы 8 мы видим, что скорость сходимости в следствии 4 определяется величиной порядка n^{-1} . Поскольку при $n \rightarrow \infty$

$$P(S_n = 0) - \pi_0 = e^{n \ln(1-p)} - e^\mu \sim \frac{\mu^2}{2n} e^{-\mu},$$

то эта оценка существенно улучшена быть не может. Однако при больших k (так сказать, в области больших отклонений) такая оценка для разности

$$P(S_n = k) - \Pi_\mu(\{k\})$$

становится грубой. (Ведь в (21) мы пренебрегаем не только разными знаками поправочных слагаемых, но и редкими событиями $\{S_n = k\}$ и $\{S_n^* = k\}$, которые стоят под знаком вероятностей.) Здесь мы приходим так же, как и в § 4, к необходимости иметь приближения, обладающие малой не только *абсолютной*, но и *относительной* погрешностью.

Покажем, что *соотношения асимптотической эквивалентности (22) сохраняются и при растущих (вместе с n) k и μ таких, что*

$$k = o(n^{2/3}), \quad \mu = o(n^{2/3}), \quad |k - \mu| = o(\sqrt{n}).$$

Действительно,

$$P(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k} = \frac{n(n-1) \cdots (n-k+1)}{k!} p^k (1-p)^{n-k} =$$

$$= \frac{(pn)^k}{k!} e^{-pn} \left(1 - \frac{1}{n}\right) \cdots \left(1 - \frac{k-1}{n}\right) (1-p)^{n-k} e^{pn} = \Pi_\mu(\{k\}) e^{\varepsilon(k,n)}.$$

Надо доказать, таким образом, что при указанных значениях k и μ

$$\varepsilon(k, n) = \ln \left[\left(1 - \frac{1}{n}\right) \cdots \left(1 - \frac{k-1}{n}\right) (1-p)^{n-k} e^{pn} \right] = o(1). \quad (23)$$

Мы получим это соотношение вместе с видом поправочного члена. Именно, покажем, что

$$\varepsilon(k, n) = \frac{k - (k - \mu)^2}{2n} + O\left(\frac{k^3 + \mu^3}{n^2}\right), \quad (24)$$

и, следовательно,

$$P(S_n = k) = \left(1 + \frac{k - (k - \mu)^2}{2n} + O\left(\frac{k^3 + \mu^3}{n^2}\right)\right) \Pi_\mu(\{k\}).$$

Воспользуемся тем, что при $\alpha \rightarrow 0$

$$\ln(1 - \alpha) = -\alpha - \frac{\alpha^2}{2} + O(\alpha^3).$$

Тогда соотношения (23), (24) будут следовать из равенств

$$\begin{aligned} \sum_{j=1}^{k-1} \ln\left(1 - \frac{j}{n}\right) &= -\sum_{j=1}^{k-1} \frac{j}{n} + O\left(\frac{k^3}{n^2}\right) = -\frac{k(k-1)}{2n} + O\left(\frac{k^3}{n^2}\right), \\ (n-k) \ln(1-p) + pn &= (n-k) \left(-p - \frac{p^2}{2} + O(p^3)\right) + pn = \\ &= -\frac{\mu^2}{2n} + \frac{k\mu}{n} + O\left(\frac{\mu^3}{n^2}\right). \end{aligned}$$

В заключение отметим, что приближенная формула Пуассона

$$P(S_n = k) \approx \frac{\mu^k}{k!} e^{-\mu}$$

широко используется в различных приложениях и обладает, как показывают опыт и полученные выше оценки, высокой точностью даже при умеренных значениях n .

Рассмотрим теперь примеры, где для приближенных вычислений используются теоремы Муавра—Лапласа и Пуассона.

Пример 2. Пусть имеется 10^4 пакетов с зерном. Известно, что в этих пакетах имеется 5000 меченых зерен. Какова вероятность, что в некотором фиксированном пакете есть хотя бы одно меченое зерно? Можно себе представить, что меченые зерна случайным образом распределяются среди пакетов. Тогда вероятность попадания определенного меченого зерна в выбранный пакет будет равна $p = 10^{-4}$. Так как этих зерен 5000, то это число и будет числом испытаний, т. е. $n = 5000$. Определим случайную величину ξ_k следующим образом: $\xi_k = 1$, если k -е зерно попало в выбранный пакет, $\xi_k = 0$ в противном случае. Тогда $S_{5000} = \sum_{k=1}^{5000} \xi_k$ будет числом меченых зерен в нашем пакете. По теореме 8 $P(S_{5000} = 0) \approx e^{-np} = e^{-0.5}$, так

что искомая вероятность равна приблизительно $1 - e^{-0,5}$. Точность этих соотношений оказывается весьма высокой (в силу теоремы 8 погрешность не превосходит $2^{-1} \cdot 10^{-4}$). Если бы мы пользовались теоремой Пуассона, а не теоремой 8, то должны были бы представить себе схему случайных величин, распределенных по закону Бернулли, при этом наши случайные величины составляют пятитысячную серию. Кроме того, мы считали бы, что для серии с номером n $np_n = 0,5$. Таким образом, условия теоремы Пуассона были бы выполнены, и можно было бы воспользоваться предельной теоремой для того, чтобы написать уже полученное приближенное соотношение.

Пример 3. Те же рассуждения можно использовать и в следующей задаче. Водоем объема V , из которого мы берем пробу объема $v \ll V$, содержит n опасных бактерий. Какова вероятность того, что мы обнаружим эти бактерии в контрольном сосуде?

Обычно считают, что вероятность p того, что любая заданная бактерия попадет в пробу, равна отношению объемов v/V . Кроме того, обычно также предполагают, что появление одной данной бактерии в пробе не зависит от того, где находятся остальные $n-1$ бактерий — в пробе или нет. Иными словами, обычно постулируют, что механизм переноса бактерий в контрольную пробу эквивалентен последовательности n независимых испытаний с вероятностью «успеха» в каждом испытании, равной $p = v/V$.

Вводя, как и раньше, случайные величины ξ_k , получим описание бактерий в пробе с помощью суммы $S_n = \sum_{k=1}^n \xi_k$ в схеме Бернулли. Если nv сравнимо по величине с V , то по теореме Пуассона искомая вероятность

$$P(S_n > 0) \approx 1 - e^{-nv/V}.$$

Модели, аналогичные только что описанной, используются также для описания числа видимых звезд в некоторой области неба, расположенной вдали от Млечного пути. Именно, предполагается, что если в области R имеется n видимых звезд, то вероятность того, что в подобласти $r \subset R$ будет находиться k видимых звезд, равна

$$\binom{n}{k} p^k (1-p)^{n-k},$$

где p равно отношению площадей $S(r)/S(R)$ областей r и R соответственно.

Пример 4. Допустим, что вероятность рождения мальчика постоянна и равна 0,512 (см. п. 1 из § 5 главы 3).

Рассмотрим группу новорожденных в 10^4 человек и будем считать, что ей соответствует последовательность 10^4 независимых испытаний, результатом каждого из которых является появление мальчика или девочки. Какова вероятность того, что среди этих новорожденных мальчиков будет по крайней мере на двести человек больше, чем девочек?

Случайные величины ξ_k определим следующим образом: $\xi_k = 1$, если k -й новорожденный оказался мальчиком, и $\xi_k = 0$ в противном случае. Тогда $S_n = \sum_{k=1}^{10^4} \xi_k$ есть число мальчиков в зарегистрированной группе. Число $npq \approx 0,25 \cdot 10^4$ здесь

велико, поэтому, применив интегральную теорему Муавра—Лапласа, для искомой вероятности получим

$$P(S_n \geq 5100) = 1 - P\left(\frac{S_n - np}{\sqrt{npq}} < \frac{5100 - 5120}{\sqrt{2500}}\right) \approx 1 - \Phi\left(-\frac{20}{50}\right) = 1 - \Phi(-0,4) \approx 0,66.$$

Для отыскания численных значений $\Phi(x)$ часто пользуются таблицами, имеющимися в большинстве руководств по теории вероятностей и математической статистике (см. Приложение 7). Они заложены также во многие статистические программы и в калькуляторы.

В приведенном примере $\Delta \approx 1/50$ и удовлетворительное приближение формулой Муавра—Лапласа будет заведомо обеспечено (см. теорему 7) при $c \leq 2,5$.

Если же, допустим, нам потребуется оценить вероятность того, что доля мальчиков превысит 0,55, то мы будем иметь дело с вероятностями больших уклонений, где для оценки $P(S_n > 5500)$ уже лучше воспользоваться приближенным соотношением, полученным в § 1.3, в силу которого ($k = 0,45n$, $q = 0,488$)

$$P(S_n > 5500) \approx \frac{(n+1-k)q}{(n+1)q-k} P(S_n = 5500).$$

Пользуясь теоремой 3, находим

$$P(S_n > 5500) \approx \frac{0,55q}{q-0,45} \frac{1}{\sqrt{2\pi n 0,25}} e^{-nH(0,55)} \leq \frac{1}{5} e^{-25} < 10^{-11}.$$

То есть, если допустить на минуту, что на планете ежегодно рождается 100 млн чел. и что все родившиеся разбиваются на партии по 10 тыс. в каждой, то для того, чтобы в какой-нибудь партии появилось завышение доли мальчиков всего на 3,8 %, потребуется ждать в среднем более 10^7 лет (см. пример 1 из § 4.1).

Ясно, что числовые оценки, основанные на нормальном приближении, можно проводить и в задачах примера 3, если числа np , фигурирующие там, окажутся большими.

§ 5. Неравенства для вероятностей больших уклонений в схеме Бернулли

В заключение этой главы мы получим некоторые полезные неравенства для схемы Бернулли. В § 2 мы ввели функцию

$$H(x) = x \ln \frac{x}{p} + (1-x) \ln \frac{1-x}{1-p},$$

которая играет существенную роль в теоремах 3, 4 об асимптотическом поведении вероятности $P(S_n = k)$. Там же были рассмотрены простейшие свойства этой функции.

Теорема 10. При $z \geq 0$

$$\begin{aligned} P(S_n - np \geq z) &\leq \exp \left\{ -nH \left(p + \frac{z}{n} \right) \right\}, \\ P(S_n - np \leq -z) &\leq \exp \left\{ -nH \left(p - \frac{z}{n} \right) \right\}. \end{aligned} \quad (25)$$

Кроме того, при всех p

$$H(p+x) \geq 2x^2, \quad (26)$$

так что каждая из вероятностей (25) при любом p не превосходит $\exp\{-2z^2/n\}$.

Для сравнения с утверждением (5) теоремы 3 первое неравенство теоремы 10 можно записать в виде

$$P\left(\frac{S_n}{n} \geq p^*\right) \leq \exp\{-nH(p^*)\}.$$

Неравенства (25) в какой-то мере близки и к теореме Муавра—Лапласа, поскольку при $z = o(n^{2/3})$

$$-nH\left(p + \frac{z}{n}\right) = -\frac{z^2}{2npq} + o(1).$$

Последнее утверждение вместе с (26) можно интерпретировать так: отклонение z от среднего значения np имеет максимальную вероятность при $p = 1/2$.

При $z/\sqrt{n} \rightarrow \infty$ обе вероятности (25) сходятся к нулю при $n \rightarrow \infty$, так как они соответствуют большим отклонениям суммы S_n от среднего np . В этом случае их называют *вероятностями больших отклонений*.

Доказательство теоремы 10. В следствии 4.3 предыдущей главы было установлено неравенство

$$P(\xi \geq x) \leq e^{-\lambda x} E e^{\lambda \xi}.$$

Применяя его к сумме S_n , получим

$$P(S_n \geq np + z) \leq e^{-\lambda(np+z)} E e^{\lambda S_n}.$$

Так как $e^{\lambda S_n} = \prod_{k=1}^n e^{\lambda \xi_k}$ и случайные величины $e^{\lambda \xi_k}$ независимы, то

$$E e^{\lambda S_n} = \prod_{k=1}^n E e^{\lambda \xi_k} = (p e^{\lambda} + q)^n = (1 + p(e^{\lambda} - 1))^n,$$

$$P(S_n \geq np + z) \leq \left[(1 + p(e^{\lambda} - 1)) e^{-\lambda(p+\alpha)} \right]^n, \quad \alpha = \frac{z}{n}.$$

Выражение в квадратных скобках равно $E e^{\lambda[\xi_k - (p+\alpha)]}$. Следовательно, как сумма двух выпуклых функций это есть выпуклая функция от λ . Уравнение для точки $\lambda(\alpha)$ минимума этой функции имеет вид

$$-(p+\alpha)(1 + p(e^{\lambda} - 1)) + p e^{\lambda} = 0.$$

Отсюда находим

$$e^{\lambda(\alpha)} = \frac{(p+\alpha)q}{p(q-\alpha)},$$

$$\begin{aligned} (1 + p(e^{\lambda(\alpha)} - 1)) e^{-\lambda(\alpha)(p+\alpha)} &= \frac{q}{q-\alpha} \left[\frac{p(q-\alpha)}{(p+\alpha)q} \right]^{p+\alpha} = \frac{p^{p+\alpha} q^{q-\alpha}}{(p+\alpha)^{p+\alpha} (q-\alpha)^{q-\alpha}} = \\ &= \exp \left\{ -(p+\alpha) \ln \frac{p+\alpha}{p} - (q-\alpha) \ln \frac{q-\alpha}{q} \right\} = \exp\{-H(p+\alpha)\}. \end{aligned}$$

Первое из неравенств (25) доказано. Второе неравенство следует из первого, если его рассматривать как неравенство для числа появления нулей.

Далее из (4) следует, что $H(p) = H'(p) = 0$, $H''(x) = 1/(x(1-x))$. Поскольку максимум функции $x(1-x)$ на $[0, 1]$ достигается в точке $x = 1/2$, то $H''(x) \geq 4$,

$$H(p + \alpha) \geq 4 \frac{\alpha^2}{2} = 2\alpha^2. \quad \blacksquare$$

Об аналогах теоремы 10 для сумм произвольных случайных величин см. замечание 1 из § 7.5.

Глава 6

О сходимости случайных величин и распределений

§ 1. Сходимость случайных величин

В предыдущих главах мы уже не раз сталкивались с утверждениями, где рассматривалась сходимость в том или ином смысле распределений случайных величин или самих случайных величин. Теперь введем ряд определений, описывающих разные виды сходимости, и выясним соотношения между ними.

Пусть на вероятностном пространстве $\langle \Omega, \mathfrak{F}, P \rangle$ заданы последовательность случайных величин $\{\xi_n\}$ и случайная величина ξ .

Определение 1. Последовательность $\{\xi_k\}$ *сходится по вероятности*¹⁾ к ξ , если для любого $\epsilon > 0$

$$P(|\xi_n - \xi| > \epsilon) \rightarrow 0 \quad \text{при} \quad n \rightarrow \infty.$$

Обозначается это так:

$$\xi_n \xrightarrow{p} \xi \quad \text{при} \quad n \rightarrow \infty.$$

В этих обозначениях утверждение закона больших чисел для схемы Бернулли можно было бы записать в виде

$$\frac{S_n}{n} \xrightarrow{p} p,$$

поскольку S_n/n можно рассматривать как последовательность случайных величин, заданных на одном вероятностном пространстве.

Определение 2. Мы будем говорить, что *последовательность ξ_n сходится к ξ с вероятностью 1* (или *почти наверное*: $\xi_n \xrightarrow{\text{п. н.}} \xi$), если $\xi_n(\omega) \rightarrow \xi(\omega)$

при $n \rightarrow \infty$ для всех $\omega \in \Omega$ за исключением, быть может, ω из множества $N \subset \Omega$ нулевой вероятности: $P(N) = 0$.

Такую сходимость можно назвать также сходимостью *почти всюду* (п.в.) относительно меры P .

Очевидно, что сходимость $\xi_n \xrightarrow{\text{п. н.}} \xi$ влечет за собой сходимость $\xi_n \xrightarrow{p} \xi$.

Обратное утверждение, вообще говоря, неверно, в чем мы убедимся позже. Однако, если ξ_n монотонно возрастает или убывает, то сходимость $\xi_n \xrightarrow{p} \xi$ влечет за собой $\xi_n \xrightarrow{\text{п. н.}} \xi$. Действительно, предположим, не ограничивая общности, что $\xi \equiv 0$, $\xi_n \geq 0$,

¹⁾ В теоретико-множественной терминологии сходимость по вероятности означает сходимость по мере.

$\xi_n \downarrow$, $\xi_n \xrightarrow{p} \xi$. Отсутствие сходимости $\xi_n \xrightarrow{\text{п.н.}} \xi$ означало бы, что существуют $\varepsilon > 0$ и множество A , $P(A) > \delta > 0$, такие, что $\sup_{k \geq n} \xi_k > \varepsilon$ при $\omega \in A$ и при всех n . Но $\sup_{k \geq n} \xi_k = \xi_n$, и сказанное означает, что

$$P(\xi_n > \varepsilon) \geq P(A) > \delta > 0$$

при всех n , что противоречит сходимости $\xi_n \xrightarrow{p} 0$.

В этих рассуждениях случайные величины ξ_n могли быть и несобственными.

Итак, сходимость по вероятности определяется поведением числовой последовательности $P(|\xi_n - \xi| > \varepsilon)$. Можно ли аналогичным образом характеризовать сходимость с вероятностью 1? Обозначим $\zeta_n = \sup_{k \geq n} |\xi_k - \xi|$ (эта случайная величина может быть и несобственной).

Теорема 1. $\xi_n \xrightarrow{\text{п.н.}} \xi$ тогда и только тогда, когда $\zeta_n \xrightarrow{p} 0$ или, что то же, когда для любого $\varepsilon > 0$ при $n \rightarrow \infty$

$$P(\sup_{k \geq n} |\xi_k - \xi| > \varepsilon) \rightarrow 0. \quad (1)$$

Доказательство. Ясно, что $\xi_n \rightarrow \xi$ п.н. тогда и только тогда, когда $\zeta_n \rightarrow 0$ п.н. Но последовательность ζ_n монотонно убывает $\zeta_n \downarrow$, $\zeta_n \xrightarrow{p} 0$, и нам остается воспользоваться сделанными выше замечаниями, в силу которых $\zeta_n \rightarrow 0$ п.н. ■

Во введенной терминологии утверждение усиленного закона больших чисел для схемы Бернулли можно сформулировать в силу (1) как сходимость $S_n/n \rightarrow p$ с вероятностью 1.

Мы уже отмечали, что сходимость почти наверное влечет за собой сходимость по вероятности. Приведем теперь пример, показывающий, что обратное утверждение неверно. Пусть $(\Omega, \mathfrak{F}, P)$ есть единичная окружность с σ -алгеброй борелевских множеств и равномерным распределением. Положим $\xi(\omega) \equiv 1$, $\xi_n(\omega) = 2$ на дуге $[r(n), r(n) + 1/n]$ и $\xi_n(\omega) = 1$ на остальном множестве. Здесь $r(n) = \sum_{k=1}^n \frac{1}{k}$. Очевидно, что $\xi_n \xrightarrow{p} \xi$. В то же время $r(n) \rightarrow \infty$ при $n \rightarrow \infty$, и множество сходимости ξ_n к ξ пусто (мы не можем указать ни одного ω , для которого $\xi_n(\omega) \rightarrow \xi(\omega)$).

Однако, если $P(|\xi_n - \xi| > \varepsilon)$ убывает при $n \rightarrow \infty$ достаточно быстро, то сходимость по вероятности будет означать и сходимость п.н. Именно, соотношение (1) дает следующее достаточное условие для сходимости с вероятностью 1.

Теорема 2. Если ряд $\sum_{k=1}^{\infty} P(|\xi_k - \xi| > \varepsilon)$ при любом $\varepsilon > 0$ сходится, то $\xi_n \rightarrow \xi$ п.н.

Это утверждение очевидно, поскольку

$$P\left(\bigcup_{k \geq n} \{|\xi_k - \xi| > \varepsilon\}\right) \leq \sum_{k=n}^{\infty} P(|\xi_k - \xi| > \varepsilon).$$

Именно этим критерием мы, по существу, пользовались при доказательстве усиленного закона больших чисел для схемы Бернулли.

Обратное заключение о скорости сходимости к нулю вероятности $P(|\xi_n - \xi| > \varepsilon)$ из сходимости п. н. сделать невозможно. Читатель без труда может построить пример, когда $\xi_n \rightarrow \xi$ п. н., а $P(|\xi_n - \xi| > \varepsilon)$ сходится к нулю сколь угодно медленно.

Из теоремы 2 вытекает

Следствие 1. Если $\xi_n \xrightarrow{p} \xi$, то существует подпоследовательность $\{n_k\}$ такая, что $\xi_{n_k} \rightarrow \xi$ п. н. при $k \rightarrow \infty$.

Это утверждение также очевидно, так как достаточно взять такие n_k , чтобы $P(|\xi_{n_k} - \xi| > \varepsilon) \leq 1/k^2$, и воспользоваться теоремой 2. ■

Есть еще один важный частный случай, когда из сходимости по вероятности $\xi_n \xrightarrow{p} \xi$ следует сходимость $\xi_n \rightarrow \xi$ п. н. Это случай, когда ξ_n образованы последовательными суммами независимых случайных величин. Именно, справедливо следующее утверждение. Если $\xi_n = \sum_{k=1}^n \eta_k$, η_k независимы, то сходимость ξ_n по вероятности влечет за собой сходимость с вероятностью 1. Это утверждение будет доказано в § 10.2.

Рассмотрим, наконец, третий вид сходимости случайных величин.

Определение 3. Будем говорить, что ξ_n сходится к ξ в среднем r -го порядка (в среднем, если $r = 1$; в среднеквадратическом, если $r = 2$), если при $n \rightarrow \infty$

$$E|\xi_n - \xi|^r \rightarrow 0.$$

Такую сходимость будем обозначать $\xi_n \xrightarrow{(r)} \xi$.

Очевидно, что из $\xi_n \xrightarrow{(r)} \xi$ в силу неравенства Чебышева следует $\xi_n \xrightarrow{p} \xi$. Из п. н. сходимости (и тем более из сходимости по вероятности) сходимости $\xrightarrow{(r)}$ не следует. Таким образом, сходимость по вероятности является слабой из введенных трех видов сходимости.

Отметим, что при наличии дополнительных условий из сходимости $\xi_n \xrightarrow{p} \xi$ может, конечно, следовать $\xi_n \xrightarrow{(r)} \xi$. Например, в следствии 4 будет показано, что если $\xi_n \xrightarrow{p} \xi$, $E|\xi_n|^{r+\alpha} < c$ при некотором $\alpha > 0$ и всех n , то $\xi_n \xrightarrow{(r)} \xi$.

Определение 4. Будем говорить, что последовательность ξ_n является фундаментальной по вероятности (п. н., в среднем), если для любого $\varepsilon > 0$

$$P(|\xi_n - \xi_m| > \varepsilon) \rightarrow 0 \quad (P(\sup_{n \geq m} |\xi_n - \xi_m| > \varepsilon) \rightarrow 0, E|\xi_n - \xi_m|^r \rightarrow 0)$$

при $n \rightarrow \infty, m \rightarrow \infty$.

Теорема 3 (признак сходимости Коши). $\xi_n \rightarrow \xi$ в каком-нибудь смысле (\xrightarrow{p} , $\xrightarrow{(r)}$, п. н.) тогда и только тогда, когда ξ_n фундаментальна в соответствующем смысле.

Доказательство. Фундаментальность очевидным образом следует из сходимости в силу неравенств

$$\begin{aligned} |\xi_n - \xi_m| &\leq |\xi_n - \xi| + |\xi_m - \xi|, \\ \sup_{n \geq m} |\xi_n - \xi_m| &\leq \sup_{n \geq m} |\xi_n - \xi| + |\xi_m - \xi| \leq 2 \sup_{n \geq m} |\xi_n - \xi|, \\ |\xi_n - \xi_m|^r &\leq C_r (|\xi_n - \xi|^r + |\xi_m - \xi|^r) \end{aligned}$$

при некотором C_r .

Пусть теперь последовательность ξ_n фундаментальна по вероятности. Выберем последовательность $\{n_k\}$ так, чтобы

$$P(|\xi_n - \xi_m| > 2^{-k}) < 2^{-k}$$

при $n \geq n_k, m \geq n_k$. Обозначим

$$\xi_{n_k} = \xi'_k, \quad A_k = \{|\xi'_k - \xi'_{k+1}| > 2^{-k}\}, \quad \eta = \sum_{k=1}^{\infty} I(A_k).$$

Тогда $P(A_k) \leq 2^{-k}$, $E\eta = \sum P(A_k) \leq 1$. Это, конечно же, означает, что число η появлений событий A_k есть собственная случайная величина: $P(\eta < \infty) = 1$ и, стало быть, с вероятностью 1 произойдет лишь конечное число событий A_k . Это означает далее, что для каждого $\omega \in \Omega - A$ найдется $k_0(\omega)$ такое, что $|\xi'_k(\omega) - \xi'_{k+1}(\omega)| \leq 2^{-k}$ при всех $k \geq k_0(\omega)$. Отсюда получаем неравенство $|\xi'_k(\omega) - \xi'_l(\omega)| \leq 2^{-k+1}$ при всех $k \geq k_0(\omega), l \geq k_0(\omega)$, означающее, что последовательность чисел $\xi'_k(\omega)$ фундаментальна и что, следовательно, существует значение $\xi(\omega)$ такое, что $|\xi'_k(\omega) - \xi(\omega)| \rightarrow 0$ при $k \rightarrow \infty$. Это означает, в свою очередь, что $\xi'_k \xrightarrow{\text{п.н.}} \xi$,

$$P(|\xi_n - \xi| \geq \varepsilon) \leq P\left(|\xi_n - \xi_{n_k}| \geq \frac{\varepsilon}{2}\right) + P\left(|\xi_{n_k} - \xi| > \frac{\varepsilon}{2}\right) \rightarrow 0$$

при $n \rightarrow \infty, k \rightarrow \infty$.

Пусть теперь ξ_n фундаментальна в среднем. Тогда в силу неравенства Чебышева она будет фундаментальна по вероятности, и, следовательно, существуют случайная величина ξ и подпоследовательность $\{n_k\}$ такие, что $\xi_{n_k} \xrightarrow{\text{п.н.}} \xi$. Покажем, что $E|\xi_n - \xi|^r \rightarrow 0$. Для заданного $\varepsilon > 0$ выберем n так, чтобы $E|\xi_k - \xi_l|^r < \varepsilon$ при $k \geq n, l \geq n$. Тогда в силу леммы Фату (см. Приложение 3)

$$E|\xi_n - \xi|^r = E \lim_{n_k \rightarrow \infty} |\xi_n - \xi_{n_k}|^r = E \lim_{n_k \rightarrow \infty} \inf |\xi_n - \xi_{n_k}|^r \leq \lim_{n_k \rightarrow \infty} \inf E|\xi_n - \xi_{n_k}|^r \leq \varepsilon.$$

Это и означает, что $E|\xi_n - \xi|^r \rightarrow 0$.

Осталось проверить утверждение теоремы для сходимости п.н. Мы уже знаем, что если ξ_n фундаментальна по вероятности (и тем более п.н.), то существуют ξ и подпоследовательность ξ_{n_k} такие, что $\xi_{n_k} \xrightarrow{\text{п.н.}} \xi$. Поэтому если обозначить $n_{k(n)} = \min\{n_k : n_k \geq n\}$, то

$$P(\sup_{k \geq n} |\xi_k - \xi| > \varepsilon) \leq P\left(\sup_{k \geq n} |\xi_k - \xi_{n_{k(n)}}| > \frac{\varepsilon}{2}\right) + P\left(|\xi_{n_{k(n)}} - \xi| > \frac{\varepsilon}{2}\right) \rightarrow 0$$

при $n \rightarrow \infty$. Теорема доказана. ■

Замечание 1. Если в пространстве L_r всех случайных величин ξ на $(\Omega, \mathfrak{F}, P)$, для которых $E|\xi|^r < \infty$, ввести норму $\|\xi\| = (E|\xi|^r)^{1/r}$ (неравенство $\|\xi_1 + \xi_2\| \leq \|\xi_1\| + \|\xi_2\|$ есть не что иное, как неравенство Минковского, см. теорему 4.4), то утверждение теоремы 3 применительно к сходимости $\xrightarrow{(r)}$ (это есть сходимость по

норме в L_r) означает, что пространство L_r полно и, стало быть, является банаховым пространством.

Пространство *всех* случайных величин на $\langle \Omega, \mathfrak{F}, P \rangle$ можно метризовать так, чтобы сходимость во введенной метрике была эквивалентна сходимости по вероятности. Например, можно положить $\rho(\xi_1, \xi_2) = E \frac{|\xi_1 - \xi_2|}{1 + |\xi_1 - \xi_2|}$. Так как всегда

$$\frac{|x + y|}{1 + |x + y|} \leq \frac{|x|}{1 + |x|} + \frac{|y|}{1 + |y|},$$

то $\rho(\xi_1, \xi_2)$ удовлетворяет всем аксиомам метрики. Нетрудно видеть, что соотношения $\rho(\xi_n, \xi) \rightarrow 0$ и $\xi_n \xrightarrow{p} \xi$ эквивалентны. Утверждение теоремы 3 применительно к сходимости \xrightarrow{p} будет означать полноту введенного метрического пространства.

В заключение этого параграфа получим две «теоремы непрерывности».

Теорема 4. Пусть $\xi_n \xrightarrow{\text{п. н.}} \xi$ ($\xi_n \xrightarrow{p} \xi$) и $H(s)$ — функция, непрерывная почти всюду относительно распределения случайной величины ξ (т. е. $H(s)$ непрерывна в каждой точке множества B такого, что $P(\xi \in B) = 1$). Тогда

$$H(\xi_n) \xrightarrow{\text{п. н.}} H(\xi) \quad (H(\xi_n) \xrightarrow{p} H(\xi)).$$

Доказательство. Так как вероятности множеств $A = \{\omega: \xi_n(\omega) \rightarrow \xi(\omega)\}$ и $C = \{\omega: \xi(\omega) \in B\}$ равны единице, то $P(AC) = P(A) + P(C) - P(A + C) = 1$. Но на множестве AC имеет место сходимость $H(\xi_n) \rightarrow H(\xi)$. Сходимость п. н. доказана.

Пусть теперь $\xi_n \xrightarrow{p} \xi$. Если допустить, что сходимость $H(\xi_n) \xrightarrow{p} H(\xi)$ не имеет места, то найдутся $\varepsilon > 0$, $\delta > 0$ и подпоследовательность $\{n'\}$ такие, что

$$P(|H(\xi_{n'}) - H(\xi)| > \varepsilon) > \delta.$$

Но $\xi_{n'} \xrightarrow{p} \xi$ и, значит, найдется подпоследовательность $\{n''\}$ такая, что $\xi_{n''} \xrightarrow{\text{п. н.}} \xi$, $H(\xi_{n''}) \xrightarrow{\text{п. н.}} H(\xi)$. Это противоречит сделанному допущению, из которого вытекает, что

$$P(|H(\xi_{n''}) - H(\xi)| > \varepsilon) > \delta.$$

■

Вторая теорема непрерывности касается моментов и связи со сходимостью в среднем.

Определение 5. Последовательность $\{\xi_n\}$ называется *равномерно интегрируемой*, если

$$\sup_n E(|\xi_n|; |\xi_n| > N) \rightarrow 0 \quad \text{при} \quad N \rightarrow \infty.$$

Если $\{\xi_n\}$ равномерно интегрируема, то $\sup_n E|\xi_n| \leq c < \infty$.

Действительно, выберем N так, чтобы $\sup_n E(|\xi_n|; |\xi_n| > N) \leq 1$. Тогда $\sup_n E|\xi_n| = \sup_n [E(|\xi_n|; |\xi_n| \leq N) + E(|\xi_n|; |\xi_n| > N)] \leq N + 1$. ■

Обратное утверждение неверно. Например, для последовательности $\xi_n: P(\xi_n = n) = 1/n = 1 - P(\xi_n = 0)$ справедливо $E|\xi_n| = 1$, но она не является равномерно интегрируемой.

Теорема 5. Пусть $\xi_n \xrightarrow{p} \xi$, $\{\xi_n\}$ равномерно интегрируема. Тогда существует $E\xi$ и

$$E|\xi_n - \xi| \rightarrow 0.$$

Если, кроме того, $|\xi_n|^r$ равномерно интегрируема, то $\xi_n \xrightarrow{(r)} \xi$.

Обратно, если $\xi_n \xrightarrow{p} \xi$, $E|\xi|^r < \infty$ и $E|\xi_n - \xi|^r \rightarrow 0$, $r \geq 1$, то $|\xi_n|^r$ равномерно интегрируема.

Доказательство. Покажем, что $E\xi$ существует. В силу свойств интеграла $E(\zeta; A_n) \rightarrow 0$ при $P(A_n) \rightarrow 0$, $E|\zeta| < \infty$ (см. лемму 3 в Приложении 3). Поэтому при любых N и ε

$$E \min(|\xi|, N) = \lim_{n \rightarrow \infty} E[\min(|\xi|, N); |\xi_n - \xi| < \varepsilon] \leq \lim_{n \rightarrow \infty} E \min(|\xi_n| + \varepsilon, N) \leq c + \varepsilon.$$

Отсюда следует, что $E|\xi| \leq c$.

Далее положим для краткости $|\xi_n - \xi| = \eta_n$. Тогда $\eta_n \xrightarrow{p} 0$ и η_n вместе с ξ_n равномерно интегрируема. Имеем для любых N, ε

$$\begin{aligned} E\eta_n &= E(\eta_n; \eta_n \leq \varepsilon) + E(\eta_n; N \geq \eta_n > \varepsilon) + E(\eta_n; \eta_n > N) \leq \\ &\leq \varepsilon + NP(\eta_n \geq \varepsilon) + E(\eta_n; \eta_n > N). \end{aligned} \quad (3)$$

Выберем N так, чтобы $\sup_n E(\eta_n; \eta_n > N) \leq \varepsilon$. Тогда для такого N

$$\lim_{n \rightarrow \infty} \sup E\eta_n \leq 2\varepsilon.$$

Так как ε произвольно, то $E\eta_n \rightarrow 0$ при $n \rightarrow \infty$.

Соотношение $E|\xi_n - \xi|^r \rightarrow 0$ доказывается точно так же, как (3), поскольку $\eta_n^r = |\xi_n - \xi|^r \rightarrow 0$ и η_n^r равномерно интегрируемы вместе с $|\xi_n|^r$.

Докажем теперь обратное утверждение. Пусть для простоты $r = 1$. Имеем

$$\begin{aligned} E(|\xi_n|; |\xi_n| > N) &\leq E(|\xi_n - \xi|; |\xi_n| > N) + E(|\xi|; |\xi_n| > N) \leq \\ &\leq E|\xi_n - \xi| + E(|\xi|; |\xi_n| > N) \leq \\ &\leq E|\xi_n - \xi| + E(|\xi|; |\xi_n - \xi| > 1) + E(|\xi|; |\xi| > N - 1). \end{aligned}$$

Здесь первое слагаемое в правой части сходится к нулю по условию, второе — в силу уже упоминавшейся леммы 3 Приложения 3 и того, что $P(|\xi_n - \xi| > 1) \rightarrow 0$. Последнее слагаемое от n не зависит и может быть выбором N сделано сколь угодно малым. Теорема доказана. ■

Следствие 2. Если $\xi_n \xrightarrow{p} \xi$, $H(s)$ удовлетворяет условиям теоремы 4 и $H(\xi_n)$ равномерно интегрируема, то

$$EH(\xi_n) \rightarrow EH(\xi).$$

Это утверждение вытекает из теорем 4, 5, так как по теореме 4 $H(\xi_n) \xrightarrow{p} H(\xi)$. ■

Иногда бывает удобно различать равномерную интегрируемость *справа и слева*. Мы будем говорить, что $\{\xi_n\}$ равномерно интегрируема *справа (слева)*, если

$$\sup E(\xi_n; \xi_n \geq N) \rightarrow 0 \quad (\sup E(|\xi_n|; \xi_n \leq -N) \rightarrow 0)$$

при $N \rightarrow \infty$. Ясно, что последовательность $\{\xi_n\}$ равномерно интегрируема тогда и только тогда, когда она равномерно интегрируема *справа и слева*.

Лемма 1. Для равномерной интегрируемости $\{\xi_n\}$ справа достаточно выполнения хотя бы одного из следующих условий.

1. Для любой последовательности $N(n) \rightarrow \infty$ при $n \rightarrow \infty$

$$E(\xi_n; \xi_n > N(n)) \rightarrow 0.$$

(Очевидно, это условие является и необходимым для равномерной интегрируемости.)

2. $\xi_n \leq \eta$, $E\eta < \infty$.

3. $E(\xi_n^+)^{1+\alpha} < c < \infty$ при некотором $\alpha > 0$ ($x^+ = \max(0, x)$).

4. ξ_n равномерно интегрируема слева, $\xi_n \rightarrow \xi$, $E\xi_n \rightarrow E\xi < \infty$.

Доказательство. 1. Отсутствие равномерной интегрируемости справа означает существование $\varepsilon > 0$ и подпоследовательностей $n' \rightarrow \infty$, $N' = N'(n') \rightarrow \infty$ таких, что $E(\xi_{n'}; \xi_{n'} > N') > \varepsilon$. Но это противоречит условию 1.

2. $E(\xi_n; \xi_n > N) \leq E(\eta; \eta > N) \rightarrow 0$ при $N \rightarrow \infty$.

3. $E(\xi_n; \xi_n > N) \leq E(\xi_n^{1+\alpha} N^{-\alpha}; \xi_n > N) \leq N^{-\alpha} c \rightarrow 0$ при $N \rightarrow \infty$.

4. Положим, не ограничивая общности, $\xi \equiv 0$. Тогда $E(\xi_n; \xi_n > N) = E\xi_n - E(\xi_n; \xi_n \leq -N) - E(\xi_n; |\xi_n| \leq N)$.

Первые два слагаемые сходятся к нулю при $n \rightarrow \infty$ и любых $N = N(n) \rightarrow \infty$. В последнем слагаемом при любом $\varepsilon > 0$ $|E(\xi_n; |\xi_n| \leq N)| \leq |E(\xi_n; |\xi_n| \leq \varepsilon)| + |E(\xi_n; \varepsilon < |\xi_n| \leq N)| \leq \varepsilon + NP(|\xi_n| > \varepsilon)$. Для заданного $\varepsilon > 0$ выберем $n(\varepsilon)$ так, чтобы при всех $n \geq n(\varepsilon)$ выполнялось $P(|\xi_n| > \varepsilon) < \varepsilon$, и положим $N(\varepsilon) = [1/\sqrt{\varepsilon}]$. Это будет означать, что при всех $n \geq n(\varepsilon)$, $N \leq N(\varepsilon)$ выполняется $E(\xi_n; |\xi_n| \leq N) < \varepsilon + \sqrt{\varepsilon}$ и, стало быть, для $E(\xi_n; \xi_n > N)$ выполнены условия п. 1 настоящей леммы. ■

На основании сказанного мы можем теперь сформулировать еще три полезных следствия.

Следствие 3 (теорема о сходимости мажорируемой последовательности). Если $\xi_n \xrightarrow{p} \xi$ и $|\xi_n| < \eta$, $E\eta < \infty$, то существует $E\xi$ и $E\xi_n \rightarrow E\xi$.

Следствие 4. Если $\xi_n \xrightarrow{p} \xi$, $E|\xi_n|^{1+\alpha} < c$, $\alpha > 0$, то $\xi_n \xrightarrow{(r)} \xi$.

Следствие 5. Если $\xi_n \xrightarrow{p} \xi$ и $H(x)$ — непрерывная ограниченная функция, то $EH(\xi_n) \rightarrow EH(\xi)$.

В заключение этого параграфа получим еще одно вспомогательное утверждение, которое может быть полезным (обобщение леммы 3 Приложения 3).

Лемма 2 (об интегралах по множеству малой вероятности). Если $\{\xi_n\}$ равномерно интегрируема и $\{A_n\}$ — произвольная последовательность событий такая, что $P(A_n) \rightarrow 0$, то

$$E(|\xi_n|; A_n) \rightarrow 0$$

при $n \rightarrow \infty$.

Доказательство. Обозначим $B_n = \{|\xi_n| \leq N\}$. Тогда

$$E(|\xi_n|; A_n) = E(|\xi_n|; A_n B_n) + E(|\xi_n|; A_n \bar{B}_n) \leq NP(A_n) + E(|\xi_n|; |\xi_n| > N).$$

Для заданного $\varepsilon > 0$ выберем сначала N так, чтобы второе слагаемое не превосходило $\varepsilon/2$, а затем n так, чтобы первое слагаемое не превосходило $\varepsilon/2$. Мы получили, что $E(|\xi_n|; A_n)$ выбором n может быть сделана меньше ε . ■

§ 2. Сходимость распределений

Мы ввели в § 1 три способа характеризовать близость случайных величин, заданных на одном вероятностном пространстве. Но как быть, если случайные величины заданы на разных вероятностных пространствах (или вообще не известно, где заданы), а их распределения очень похожи? (Вспомним, например, теоремы Пуассона или Муавра—Лапласа.) В этих случаях надо уметь характеризовать близость самих распределений. При удачном определении такой близости мы сможем во многих задачах приближать нужные, но труднодоступные распределения с помощью распределений известных и, как правило, более простых.

Итак, какие распределения считать близкими? Речь идет, очевидно, об определении сходимости последовательности функций распределения $F_n(x)$ к функции распределения $F(x)$. Было бы естественным, например, считать, что распределения величин $\xi_n = \xi + 1/n$ и ξ сходятся при $n \rightarrow \infty$. Поэтому требовать малости $\sup_x |F_n(x) - F(x)|$ в определении сходимости было бы неразумно, так как этому условию не удовлетворяют распределения $\xi + 1/n$ и ξ , если $F(x) = P(\xi < x)$ имеет хотя бы один разрыв.

Мы определим сходимость F_n к F как такую, которая возникает при рассмотрении сходимости по вероятности.

Определение 6. Будем говорить, что F_n слабо сходится к F , и обозначать это через $F_n \Rightarrow F$, если для любой непрерывной и ограниченной функции $f(x)$

$$\int f(x) dF_n(x) \rightarrow \int f(x) dF(x). \quad (4)$$

Если речь идет о распределениях $P_n(B)$ и $P(B)$ (B — борелевские множества), соответствующих F_n и F , то говорят, что P_n слабо сходится к P , и обозначают это через $P_n \Rightarrow P$. Очевидно, что (4) можно записать также в виде

$$\int f(x) P_n(dx) \rightarrow \int f(x) P(dx) \quad \text{или} \quad Ef(\xi_n) \rightarrow Ef(\xi), \quad (5)$$

(ср. со следствием 5), где $\xi_n \in F_n$, $\xi \in F$.

Другое возможное определение слабой сходимости возникает из следующего утверждения.

Теорема 6.²⁾ $F_n \Rightarrow F$ тогда и только тогда, когда $F_n(x) \rightarrow F(x)$ в каждой точке x , являющейся точкой непрерывности F .

Доказательство. Пусть выполнено (4). Рассмотрим $\varepsilon > 0$ и непрерывную функцию $f_\varepsilon(t)$, равную единице при $t < x$, нулю при $t \geq x + \varepsilon$ и меняющуюся линейно на $[x, x + \varepsilon]$. Так как

$$F_n(x) = \int_{-\infty}^x f_\varepsilon(t) dF_n(t) \leq \int f_\varepsilon(t) dF_n(t),$$

²⁾ Во многих учебниках по теории вероятностей высказывание теоремы принимается за определение слабой сходимости. Однако определение $P_n \Rightarrow P$ в форме

$$\int f(x) P_n(dx) \rightarrow \int f(x) P(dx)$$

для всех ограниченных непрерывных функций f , по-видимому, целесообразнее, поскольку оно сохраняется и для распределений в произвольных топологических пространствах (см., например, [15, 1]).

то в силу (4)

$$\limsup_{n \rightarrow \infty} F_n(x) \leq \int f_\varepsilon(t) dF(t) \leq F(x + \varepsilon).$$

Если x — точка непрерывности F , то в силу произвольности ε получаем $\limsup_{n \rightarrow \infty} F_n(x) \leq F(x)$.

Совершенно аналогично с помощью функции $f_\varepsilon^*(t) = f_\varepsilon(t + \varepsilon)$ получаем неравенство

$$\liminf_{n \rightarrow \infty} F_n(x) \geq F(x).$$

Получим теперь обратное утверждение теоремы. Пусть $-M$ и N — такие точки непрерывности F , что $F(-M) < \varepsilon/5$, $1 - F(N) < \varepsilon/5$. Тогда $F_n(-M) < \varepsilon/4$, $1 - F_n(N) < \varepsilon/4$ при всех достаточно больших n . Поэтому, считая для простоты $|f| \leq 1$, получим, что

$$\int f dF_n \quad \text{и} \quad \int f dF \tag{6}$$

будут отличаться соответственно от

$$\int_{-M}^N f dF_n \quad \text{и} \quad \int_{-M}^N f dF$$

меньше, чем на $\varepsilon/2$. Построим на полуотрезке $(-M, N]$ ступенчатую функцию f_ε со скачками в точках непрерывности F , которая отличалась бы от f меньше, чем на $\varepsilon/2$. Вне $(-M, N]$ функцию f_ε положим равной нулю. Можно считать,

например, что $f_\varepsilon(x) = \sum_{j=1}^k f(x_j) \delta_j(x)$, где $x_0 = -M$, $x_1, \dots, x_k = N$ есть выбранные

нужным образом точки непрерывности F , функция $\delta_j(x)$ — индикатор полуотрезка $(x_{j-1}, x_j]$. Тогда $\int f_\varepsilon dF_n$ и $\int f_\varepsilon dF$ будут отличаться от интегралов (6) при достаточно больших n меньше, чем на ε . В то же время

$$\int f_\varepsilon dF_n = \sum_{j=1}^k f(x_j) [F_n(x_j) - F_n(x_{j-1})] \rightarrow \int f_\varepsilon dF.$$

Отсюда в силу произвольности $\varepsilon > 0$ следует сходимость (4) (надо воспользоваться неравенством

$$\limsup \int f dF_n \leq \varepsilon + \limsup \int f_\varepsilon dF_n = \varepsilon + \int f_\varepsilon dF \leq 2\varepsilon + \int f dF$$

и аналогичным неравенством для $\liminf \int f dF_n$). ■

Замечания по поводу другого, в известном смысле более простого доказательства второго утверждения теоремы 6 см. в конце § 3 и в § 7.4.

Замечание 2. Если повторить с очевидными изменениями приведенное доказательство, то мы получим несколько иной эквивалент сходимости (4): *сходимость разностей $F_n(y) - F_n(x) \rightarrow F(y) - F(x)$ для любых x и y , являющихся точками непрерывности F .*

Замечание 3. Если $F(x)$ непрерывна, то сходимость $F_n \Rightarrow F$ эквивалентна равномерной сходимости $\sup_x |F_n(x) - F(x)| \rightarrow 0$.

Доказательство этого утверждения предоставляем читателю. Оно следует из того, что сходимость $F_n(x) \rightarrow F(x)$ при каждом x влечет за собой в силу непрерывности F равномерную сходимость на любом конечном отрезке. Равномерная малость $F_n(x) - F(x)$ на «краях» обеспечивается малостью $F(x)$ или $1 - F(x)$.

Замечание 4. Если распределения F_n и F дискретны и имеют скачки в одних и тех же точках x_1, x_2, \dots , то $F_n \Rightarrow F$ будет эквивалентной, очевидно, сходимости вероятностей значений x_1, x_2, \dots ($F_n(x_k + 0) - F_n(x_k) \rightarrow F(x_k + 0) - F(x_k)$).

Введем теперь некоторые удобные для дальнейшего обозначения. Пусть ξ_n, ξ — некоторые случайные величины (заданные, вообще говоря, на разных вероятностных пространствах) такие, что $\xi_n \in P_n, \xi \in P$.

Определение 7. Если $P_n \Rightarrow P$, то будем говорить, что ξ_n сходится к ξ по распределению, и обозначать это через $\xi_n \Rightarrow \xi$.

Мы используем здесь тот же символ \Rightarrow , что и для обозначения слабой сходимости, но это нигде к недоразумениям не приведет.

Ясно, что $\xi_n \xrightarrow{p} \xi$ влечет за собой $\xi_n \Rightarrow \xi$, но не наоборот.

В то же время справедливо следующее утверждение.

Лемма 3. Если $\xi_n \Rightarrow \xi$ ($F_n \Rightarrow F$), то можно на одном вероятностном пространстве построить случайные величины ξ'_n и ξ' такие, что $P(\xi'_n < x) = P(\xi_n < x) = F_n(x)$, $P(\xi' < x) = P(\xi < x) = F(x)$,

$$\xi'_n \xrightarrow{\text{п.н.}} \xi'.$$

Доказательство. Определим обратные функции $F_n^{-1}(t) = \sup(x: F_n(x) \leq t)$, $F^{-1}(t) = \sup(x: F(x) \leq t)$. (Если $F(x)$ непрерывна и строго возрастает, то $F^{-1}(t)$ совпадает с решением уравнения $F(v) = t$.) Пусть $\eta \in U_{0,1}$. Положим

$$\xi'_n = F_n^{-1}(\eta) \in F_n, \quad \xi' = F^{-1}(\eta) \in F$$

(см. § 3.2) и покажем, что $\xi'_n \xrightarrow{\text{п.н.}} \xi'$. Для этого достаточно доказать, что $F_n^{-1}(t) \rightarrow F^{-1}(t)$ для п. в. $t \in [0, 1]$.

Функции F и F^{-1} являются монотонными, и, стало быть, число точек разрыва для каждой из них не более чем счетно. Это означает, что для всех $t \in [0, 1]$ за исключением, быть может, счетного множества T функция $F^{-1}(t)$ будет непрерывной и точка $F^{-1}(t)$ будет точкой непрерывности F .

Допустим противное: требуемая сходимость $F_n^{-1}(t) \rightarrow F^{-1}(t)$ отсутствует. Это будет означать, что найдутся $t \notin T$, $\varepsilon > 0$ и подпоследовательность n' такие, что

$$F_{n'}^{-1}(t) < F^{-1}(t) - \varepsilon \quad (\text{или} \quad > F^{-1}(t) + \varepsilon).$$

Применение $F_{n'}$ дает эквивалентное неравенство

$$t < F_{n'}(F^{-1}(t) - \varepsilon).$$

При достаточно малом ε $F^{-1}(t) - \varepsilon$ также будет точкой непрерывности F и, следовательно, правая часть последнего неравенства будет сходиться к $F(F^{-1}(t) - \varepsilon)$.

Это означает, что

$$t \leq F(F^{-1}(t) - \varepsilon), \quad F^{-1}(t) \leq F^{-1}(t) - \varepsilon.$$

Мы получили противоречие, доказывающее требуемое утверждение. ■

Лемма 3 останется справедливой и для векторнозначных случайных величин.

Иногда удобно иметь также простой символ для записи выражения «распределения ξ_n слабо сходятся к P ». Мы будем записывать это соотношение в виде

$$\xi_n \Leftrightarrow P, \quad (7)$$

так что символ \Leftrightarrow выражает тот же факт, что и \Rightarrow , но соединяет объекты разной природы так же, как это делает символ \in в соотношении $\xi \in P$ (слева в (7) стоят случайные величины, а справа — распределение).

В этих терминах утверждение теоремы Пуассона можно записать в виде $S_n \Leftrightarrow \Pi_p$, а утверждение закона больших чисел для схемы Бернулли в виде $S_n/n \Leftrightarrow I_p$.

Совпадение распределений ξ и η будем обозначать символом $\xi =_p \eta$.

Замечание 5. Ясно, что если $\xi_n \Rightarrow \xi$, $\varepsilon_n \xrightarrow{p} 0$, то $\xi_n + \varepsilon_n \Rightarrow \xi$.

Действительно, при любых t и $\delta > 0$ таких, что t , $t \pm \delta$ являются точками непрерывности $P(\xi < t)$, будем иметь

$$\begin{aligned} \limsup_{n \rightarrow \infty} P(\xi_n + \varepsilon_n < t) &= \limsup_{n \rightarrow \infty} P(\xi_n + \varepsilon_n < t, \varepsilon_n > -\delta) \leq \\ &\leq \limsup_{n \rightarrow \infty} P(\xi_n < t + \delta) = P(\xi < t + \delta). \end{aligned}$$

Аналогично

$$\liminf_{n \rightarrow \infty} P(\xi_n + \varepsilon_n < t) \geq P(\xi < t - \delta).$$

Так как $P(\xi < t \pm \delta)$ могут быть сделаны выбором δ сколь угодно близкими к $P(\xi < t)$, то отсюда следует требуемая сходимость.

Приведем аналоги теорем 4, 5 в терминах распределений.

Теорема 4А. Если $\xi_n \Rightarrow \xi$ и $H(s)$ удовлетворяют условиям теоремы 4, то $H(\xi_n) \Rightarrow H(\xi)$.

Теорема 5А. Если $\xi_n \Rightarrow \xi$ и $\{\xi_n\}$ равномерно интегрируема, то существует $E\xi$ и $E\xi_n \rightarrow E\xi$.

Есть два пути доказательства этих теорем. Один из них состоит в сведении их к теоремам 4, 5. Для этого надо построить случайные величины $\xi'_n = F_n^{-1}(\eta)$ и $\xi' = F^{-1}(\eta)$, где $\eta \in U_{0,1}$, F_n^{-1} , F^{-1} — обратные функции к F_n , F , и доказать, что $\xi'_n \xrightarrow{p} \xi'$ (мы уже знаем, что $F^{-1}(\eta) \in F$; если F разрывна или не строго монотонна, то F^{-1} должна быть определена, как в лемме 3).

Другой путь состоит в проведении доказательств заново на языке распределений. При наличии незначительных дополнительных предположений это оказывается иногда даже проще. Чтобы проиллюстрировать это, предположим, например, в теореме 4А, что функция H непрерывна. Надо доказать, что $Eg(H(\xi_n)) \rightarrow Eg(H(\xi))$ для любой непрерывной и ограниченной функции g . Но это есть прямое следствие (4), (5) при $f = g \circ H$ (f есть суперпозиция функций g и H).

В теореме 5А предположим, что $\xi_n \geq 0$ (это не ограничивает общности). Тогда, интегрируя по частям, получим

$$E\xi_n = - \int_0^{\infty} x dP(\xi_n \geq x) = \int_0^{\infty} P(\xi_n \geq x) dx. \quad (8)$$

Так как в силу равномерной интегрируемости

$$\sup_n \int_N^{\infty} P(\xi_n \geq x) dx \leq \sup_n E(\xi_n; \xi_n \geq N) \rightarrow 0$$

при $N \rightarrow \infty$, то интеграл в (8) сходится равномерно. Так как, кроме того, $P(\xi_n \geq x) \rightarrow P(\xi \geq x)$ п. в., то

$$\lim_{n \rightarrow \infty} E\xi_n = \lim_{n \rightarrow \infty} \int_0^{\infty} P(\xi_n \geq x) dx = \int_0^{\infty} P(\xi \geq x) dx = E\xi.$$

Условия, обеспечивающие равномерную интегрируемость, содержатся в лемме 1. Приведем здесь модификацию п. 4 этой леммы на случай слабой сходимости.

Лемма 1А. Если ξ_n равномерно интегрируемы слева, $\xi_n \Rightarrow \xi$, $E\xi_n \rightarrow E\xi$, то ξ_n равномерно интегрируемы.

Мы предлагаем читателю построить примеры, показывающие, что все три условия леммы существенны.

Из леммы 1А следует, в частности, что если $\xi_n \geq 0$, $\xi_n \Rightarrow \xi$, $E\xi_n \rightarrow E\xi$, то ξ_n равномерно интегрируемы.

Как и для теорем 4А, 5А здесь также возможны два пути доказательства. Один из них состоит в использовании леммы 1. Мы приведем здесь иное и несколько более простое доказательство.

Доказательство леммы 1А. Предположим для простоты, что $\xi_n \geq 0$ и допустим, что лемма не верна. Тогда найдутся $\varepsilon > 0$, подпоследовательности $n' \rightarrow \infty$ и $N(n') \rightarrow \infty$ такие, что

$$E(\xi_{n'}; \xi_{n'} > N(n')) > \varepsilon.$$

Так как $E\xi_{n'} = E(\xi_{n'}; \xi_{n'} \leq N) + E(\xi_{n'}; \xi_{n'} > N)$, то для любого N , являющегося точкой непрерывности распределения ξ ,

$$E\xi = \lim E\xi_n \geq E\xi; \xi \leq N) + \varepsilon.$$

Выберем N так, чтобы первое слагаемое в правой части превосходило $E\xi - \varepsilon/2$. Тогда мы получим противоречие $E\xi \geq E\xi + \varepsilon/2$, доказывающее лемму.

Переход к произвольным ξ_n , равномерно интегрируемым слева, мы предоставляем читателю.

Бывает полезной также следующая

Теорема 7. Пусть $\xi_n \Rightarrow \xi$, $H(s)$ дифференцируема в точке a , $b_n \rightarrow 0$ при $n \rightarrow \infty$. Тогда

$$\frac{H(a + b_n \xi_n) - H(a)}{b_n} \Rightarrow \xi H'(a).$$

Если $H'(a) = 0$ и существует $H''(a)$, то

$$\frac{H(a + b_n \xi_n) - H(a)}{b_n^2} \Rightarrow \frac{\xi^2}{2} H''(a).$$

Доказательство. Рассмотрим функцию

$$h(x) = \begin{cases} \frac{H(a+x) - H(a)}{x} & \text{при } x \neq 0, \\ H'(a) & \text{при } x = 0, \end{cases}$$

которая будет непрерывной в точке $x = 0$. Так как $b_n \xi_n \Rightarrow 0$, то по теореме 4А $h(b_n \xi_n) \Rightarrow h(0) = H'(a)$. Пользуясь снова этой теоремой (для двумерных распределений), получим

$$\frac{H(a + b_n \xi_n) - H(a)}{b_n} = h(b_n \xi_n) \xi_n \Rightarrow H'(a) \xi.$$

Второе утверждение доказывается точно так же. ■

Многомерный аналог этой теоремы будет выглядеть несколько сложнее. При необходимости читатель может получить ее самостоятельно, следуя рассуждениям теоремы 7.

§ 3. Условия слабой сходимости

Вернемся к понятию слабой сходимости. Мы имеем два критерия этой сходимости: (4) и теорема 6. Однако с точки зрения применений (проверки их в конкретных задачах) оба эти критерия неудобны. Значительно проще было бы, например, доказывать сходимость $Ef(\xi_n) \rightarrow Ef(\xi)$ не для всех ограниченных непрерывных f , а для функций f из какого-нибудь достаточно узкого и просто устроенного класса. Ясно, что совсем узким такой класс быть не может.

Прежде чем формулировать основные утверждения, введем в рассмотрение некоторые понятия.

Расширим класс \mathcal{F} всех функций распределения до класса \mathcal{G} функций G , удовлетворяющих условиям F1, F2 из § 3.2 и условиям $G(-\infty) \geq 0$, $G(\infty) \leq 1$. Функции G из \mathcal{G} можно называть обобщенными распределениями. Их можно представлять себе как функции распределения несобственных случайных величин ξ , принимающих с положительной вероятностью и бесконечные значения, так что $G(-\infty) = P(\xi = -\infty)$, $1 - G(\infty) = P(\xi = \infty)$. Мы будем писать $G_n \Rightarrow G$, $G_n \in \mathcal{G}$, $G \in \mathcal{G}$, если $G_n(x) \rightarrow G(x)$ во всех точках непрерывности $G(x)$.

Теорема 8 (Хелли). *Класс \mathcal{G} есть компакт относительно сходимости \Rightarrow , т.е. из любой последовательности $\{G_n\}$, $G_n \in \mathcal{G}$, можно выделить сходящуюся подпоследовательность $G_{n_k} \Rightarrow G \in \mathcal{G}$.*

Доказательство этой теоремы см. в Приложении 4.

Следствие 6. *Если любая сходящаяся подпоследовательность $\{G_{n_k}\}$ из $\{G_n\}$, $G_n \in \mathcal{G}$, сходится к G , то $G_n \Rightarrow G$.*

Доказательство. Если $G_n \not\Rightarrow G$, то существует точка x_0 непрерывности G такая, что $G_n(x_0) \not\rightarrow G(x_0)$. Так как $G_n(x_0) \in [0, 1]$, то должна существовать подпоследовательность $G_{n_k}(x_0) \rightarrow g \neq G(x_0)$. Однако по предположению это невозможно, так как $G_{n_k}(x_0) \rightarrow G(x_0)$. ■

Причина расширения класса \mathcal{F} всех функций распределения состоит в том, что этот класс не является компактом (в смысле теоремы 8), а сходимость $F_n \Rightarrow G$,

$F_n \in \mathcal{F}$, еще не означает, что $G \in \mathcal{F}$. Например, последовательность

$$F_n(x) = \begin{cases} 0, & x \leq -n, \\ 1/2, & -n < x \leq n, \\ 1, & x > n, \end{cases} \quad (9)$$

всюду сходится к функции $G(x) \equiv 1/2 \notin \mathcal{F}$, соответствующей несобственной случайной величине, принимающей значения $\pm\infty$ с вероятностями $1/2$.

Однако иметь дело с классом \mathcal{G} тоже не совсем удобно. Дело в том, что сходимость в точках непрерывности $G_n \Rightarrow G$ в классе \mathcal{G} вовсе не эквивалентна сходимости $\int f dG_n \rightarrow \int f dG$ (см. пример (9) для $f \equiv 1$), да и сами интегралы $\int f dG$ не определяют однозначно G (они определяют *приращения* G , но не значения $G(-\infty)$ и $G(\infty)$). Введем теперь два понятия, которые помогут избежать отмеченных неудобств.

Определение 8. Последовательность распределений $\{P_n\}$ (или функций распределения $\{F_n\}$) называется *плотной*, если для любого $\varepsilon > 0$ существует N такое, что

$$\inf_n P_n([-N, N]) > 1 - \varepsilon. \quad (10)$$

Определение 9. Мы будем говорить, что класс \mathcal{L} непрерывных и ограниченных функций f *определяет распределение*, если равенство

$$\int f(x) dF(x) = \int f(x) dG(x), \quad F \in \mathcal{F}, \quad G \in \mathcal{G},$$

для всех $f \in \mathcal{L}$ влечет за собой $F = G$ (или $E f(\xi) = E f(\eta)$ для всех $f \in \mathcal{L}$, где одна из случайных величин ξ, η собственная, влечет $\xi = \eta$).

Основной теоремой этого параграфа является

Теорема 9. Пусть класс \mathcal{L} определяет распределение. Тогда для существования распределения $F \in \mathcal{F}$ такого, что $F_n \Rightarrow F$, необходимо и достаточно, чтобы³⁾

- 1) последовательность $\{F_n\}$ была плотной;
- 2) существовал $\lim_{n \rightarrow \infty} \int f dF_n$ для всех $f \in \mathcal{L}$.

Доказательство теоремы 9. Необходимость очевидна.

Достаточность. В силу теоремы 8 существует подпоследовательность $F_{n_k} \Rightarrow F \in \mathcal{G}$. Но по условию 1) $F \in \mathcal{F}$. Действительно, если $x \geq N$ есть точка непрерывности F , то по определению 8 $F(x) = \lim F_{n_k}(x) \geq 1 - \varepsilon$. Аналогично устанавливается, что для $x \leq -N$ выполняется $F(x) < \varepsilon$. Так как ε произвольно, то $F(-\infty) = 0$, $F(\infty) = 1$.

Далее, возьмем любую другую сходящуюся подпоследовательность $F_{n'_k} \Rightarrow G \in \mathcal{F}$. Тогда для любой $f \in \mathcal{L}$

$$\lim \int f dF_{n_k} = \int f dF, \quad \lim \int f dF_{n'_k} = \int f dG. \quad (11)$$

Но по условию 2)

$$\int f dF = \int f dG \quad (12)$$

³⁾ В таком виде теорема о сходимости распределений сохраняется и для пространств более общей природы. Роль отрезков $[-N, N]$ в (10) в этом случае играют компакты (ср. с [1, 7, 15, 25]).

и, следовательно, $F = G$. По следствию 6 теорема доказана. ■

Если доказывать сходимость к «известному» распределению $F \in \mathcal{F}$, то условие плотности в теореме 9 становится ненужным.

Следствие 7. Пусть класс \mathcal{L} определяет распределение и

$$\int f dF_n \rightarrow \int f dF, \quad F \in \mathcal{G}, \quad (13)$$

для любой $f \in \mathcal{L}$. Пусть, кроме того, выполнено хотя бы одно из следующих трех условий:

- 1) последовательность $\{F_n\}$ является плотной;
- 2) $F \in \mathcal{F}$;
- 3) $f \equiv 1 \in \mathcal{L}$.

Тогда $F \in \mathcal{F}$ и $F_n \Rightarrow F$.

Доказательство следствия почти очевидно. Утверждение при условии 1) сразу вытекает из теоремы 9. Условие 3) и сходимость (13) влекут за собой условие 2). Если выполнено 2), то в соотношениях (11), (12) $F \in \mathcal{F}$ и, значит, $G = F$. ■

Так как, как правило (увидим это ниже), хотя бы одно из условий 1)–3) бывает выполненным, то основной задачей становится проверка сходимости (13) для класса \mathcal{L} .

Отметим также, что в случае, когда доказывается сходимость к «известному» заранее распределению $F \in \mathcal{F}$, вся схема рассуждений может быть иной и более простой. Один из таких альтернативных подходов изложен в § 7.4.

Приведем теперь несколько примеров классов \mathcal{L} , определяющих распределение.

Пример 1. Класс \mathcal{L}_0 функций, имеющих вид

$$f_{a,\varepsilon}(x) = \begin{cases} 1, & x \leq a, \\ 0, & x \geq a + \varepsilon. \end{cases}$$

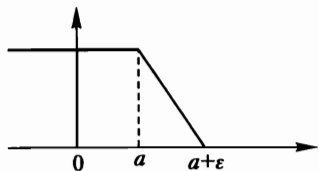


Рис. 7.

На отрезке $[a, a + \varepsilon]$ функции $f_{a,\varepsilon}$ определяются как линейные и непрерывные (график $f_{a,\varepsilon}(x)$ см. на рис. 7). Это двупараметрическое семейство функций.

Покажем, что \mathcal{L}_0 определяет распределение. Пусть $\int f dF = \int f dG$ для всех $f \in \mathcal{L}_0$. Тогда

$$F(a) \leq \int f_{a,\varepsilon} dF = \int f_{a,\varepsilon} dG \leq G(a + \varepsilon),$$

и наоборот,

$$G(a) \leq F(a + \varepsilon)$$

для любого $\varepsilon > 0$. Если взять в качестве a точку непрерывности F и G , мы получим

$$F(a) = G(a), \quad F = G.$$

Аналогичным образом легко убедиться, что класс $\widehat{\mathcal{L}}_0$ «трапецеобразных» функций $f(x) = \min(f_{a,\varepsilon}, 1 - f_{b,\varepsilon})$ также определяет распределение.

Пример 2. Класс \mathcal{L}_1 непрерывных ограниченных функций таких, что для каждой $f \in \mathcal{L}_0$ (или $f \in \widehat{\mathcal{L}}_0$) существует последовательность $f_n \in \mathcal{L}_1$, $\sup_x |f_n(x)| < M$, для которой $\lim_{n \rightarrow \infty} f_n(x) = f(x)$ при каждом $x \in \mathbb{R}$.

Пусть $\int f dF = \int f dG$ для всех $f \in \mathcal{L}_1$. По теореме о сходимости мажорируемой последовательности

$$\lim \int f_n dF = \int f dF, \quad \lim \int f_n dG = \int f dG, \quad f \in \mathcal{L}_0.$$

Поэтому $\int f dF = \int f dG$, $f \in \mathcal{L}_0$; $F = G$ и, следовательно, \mathcal{L}_1 определяет распределения.

Пример 3. Класс C_k всех ограниченных функций $f(x)$ с ограниченной равномерно непрерывной производной $f^{(k)}(x)$ k -го порядка ($\sup_x |f^{(k)}(x)| < \infty$).

Очевидно, C_k определяет распределение, так как обладает свойствами \mathcal{L}_1 .

Точно так же убеждаемся, что будет определять распределение подкласс $C_k^0 \subset C_k$ финитных функций (обращающихся в нуль вне конечного интервала). Это следует из того, что C_k^0 обладает свойством класса \mathcal{L}_1 относительно класса $\widehat{\mathcal{L}}_0$ трапещиеобразных (и, следовательно, финитных) функций.

Очевидно, что класс C_k удовлетворяет условию 3) следствия 7 и, стало быть, для доказательства сходимости $F_n \Rightarrow F \in \mathcal{F}$ достаточно проверять сходимость (13) лишь для $f \in C_k$.

Если в качестве \mathcal{L} взять класс C_k^0 финитных дифференцируемых функций, то соотношение (13) вместе с условием 2) следствия 7 можно записать в виде

$$\int F_n f' dx \rightarrow \int F f' dx, \quad F \in \mathcal{F}. \quad (14)$$

(Нужно провести интегрирование по частям в (13), где f' снова есть финитная дифференцируемая функция.) Критерий сходимости (14) иногда бывает полезным. Из него сразу видно, например, что из сходимости $F_n(x) \rightarrow F(x)$ в точках непрерывности (т. е. почти всюду) следует (13), так как из названной сходимости и теоремы о мажорируемой сходимости вытекает соотношение (14), эквивалентное (13).

Пример 4. Одним из наиболее важных классов, определяющих распределение, является однопараметрическое семейство комплекснозначных функций $\{e^{itz}\}$, $t \in \mathbb{R}$.

Изучению свойств $\int e^{itz} dF(x)$ будет посвящена следующая глава.

Все сказанное в этой главе может быть после очевидных изменений перенесено на многомерный случай.

Характеристические функции

§ 1. Определение и свойства характеристических функций

Предварительно отметим, что наряду с вещественными случайными величинами $\xi(\omega)$ мы с таким же успехом можем рассматривать и комплекснозначные случайные величины, понимая под этим функцию $\xi_1(\omega) + i\xi_2(\omega)$, где (ξ_1, ξ_2) — случайный вектор. Естественно положить $E(\xi_1 + i\xi_2) = E\xi_1 + iE\xi_2$. Комплекснозначные случайные величины $\xi = \xi_1 + i\xi_2$ и $\eta = \eta_1 + i\eta_2$ *независимы*, если σ -алгебры $\sigma(\xi_1, \xi_2)$ и $\sigma(\eta_1, \eta_2)$, порожденные соответственно векторами (ξ_1, ξ_2) и (η_1, η_2) , независимы. Без труда проверяется, что для таких величин

$$E\xi\eta = E\xi E\eta.$$

Определение 1. *Характеристической функцией (х. ф.) вещественной случайной величины ξ называется комплекснозначная функция*

$$\varphi_\xi(t) = Ee^{it\xi} = \int e^{itx} dF(x),$$

где t — действительное число.

Если функция распределения имеет плотность $f(x)$, то х. ф. равна $\varphi_\xi(t) = \int e^{itx} f(x) dx$ и есть просто преобразование Фурье функции $f(x)$. В общем случае х. ф. есть преобразование Фурье—Стилтьеса над функцией $F(x)$. Х. ф. существует для любой случайной величины ξ . Это сразу следует из того, что

$$|\varphi_\xi(t)| = \left| \int e^{itx} dF(x) \right| \leq \int 1 dF(x) = 1.$$

Х. ф. представляют собой прекрасный аппарат для исследования свойств сумм независимых случайных величин.

Свойства характеристических функций.

1. Для любой случайной величины ξ

$$\varphi_\xi(0) = 1 \quad \text{и} \quad |\varphi_\xi(t)| \leq 1 \quad \text{для всех } t.$$

Это свойство очевидно.

2. Для любой случайной величины ξ

$$\varphi_{a\xi+b}(t) = e^{itb} \varphi_\xi(ta).$$

Действительно,

$$\varphi_{a\xi+b}(t) = Ee^{it(a\xi+b)} = e^{itb} Ee^{iat\xi} = e^{itb} \varphi_\xi(ta).$$

3. Если ξ_1, \dots, ξ_n — независимые случайные величины, то х. ф. суммы $S_n = \xi_1 + \dots + \xi_n$ равна

$$\varphi_{S_n}(t) = \varphi_{\xi_1}(t) \cdots \varphi_{\xi_n}(t).$$

Доказательство следует из свойств математического ожидания произведения независимых случайных величин. Действительно,

$$\begin{aligned} \varphi_{S_n}(t) &= E e^{it(\xi_1 + \dots + \xi_n)} = E e^{it\xi_1} e^{it\xi_2} \dots e^{it\xi_n} = \\ &= E e^{it\xi_1} E e^{it\xi_2} \dots E e^{it\xi_n} = \varphi_{\xi_1}(t) \varphi_{\xi_2}(t) \cdots \varphi_{\xi_n}(t). \quad \blacksquare \end{aligned}$$

Так что свертке $F_{\xi_1} * F_{\xi_2}$ соответствует произведение $\varphi_{\xi_1} \varphi_{\xi_2}$.

4. Х. ф. $\varphi_{\xi}(t)$ является равномерно непрерывной функцией.

Действительно,

$$|\varphi(t+h) - \varphi(t)| = |E(e^{i(t+h)\xi} - e^{it\xi})| \leq E|e^{ih\xi} - 1| \rightarrow 0$$

по теореме о сходимости мажорируемой последовательности (см. следствие 6.3), так как $|e^{ih\xi} - 1| \xrightarrow{p} 0$ при $h \rightarrow 0$, $|e^{ih\xi} - 1| \leq 2$.

5. Если существует k -й момент $E|\xi|^k < \infty$, $k \geq 1$, то существует непрерывная k -я производная функции $\varphi_{\xi}(t)$ и $\varphi^{(k)}(0) = i^k E\xi^k$.

Так как $|\int i x e^{itx} dF(x)| \leq \int |x| dF(x) = E|\xi| < \infty$, то интеграл $\int i x e^{itx} dF(x)$ равномерно сходится относительно t . Поэтому возможно дифференцирование под знаком интеграла:

$$\varphi'(t) = i \int x e^{itx} dF(x), \quad \varphi'(0) = i E\xi.$$

Дальнейшие рассуждения проводятся по индукции. Если для $l < k$

$$\varphi^{(l)}(t) = i^l \int x^l e^{itx} dF(x),$$

то

$$\varphi^{(l+1)}(t) = i^{l+1} \int x^{l+1} e^{itx} dF(x)$$

в силу равномерной сходимости интеграла в правой части. Значит, $\varphi^{(l+1)}(0) = i^{l+1} E\xi^{l+1}$. \blacksquare

Следовательно, если $E|\xi|^k < \infty$, то в окрестности точки $t = 0$ справедливо разложение

$$\varphi(t) = 1 + \sum_{j=1}^k \frac{(it)^j}{j!} E\xi^j + o(|t|^k). \quad (1)$$

Утверждение в обратную сторону верно лишь частично:

Если производная четного порядка $\varphi^{(2k)}$ существует, то

$$E|\xi|^{2k} < \infty, \quad \varphi^{(2k)}(0) = (-1)^k E\xi^{2k}.$$

Докажем это свойство для $k = 1$ (далее можно использовать индукцию).

Достаточно убедиться в конечности $E|\xi|^2$. Имеем

$$-\frac{2\varphi(0) - \varphi(2h) - \varphi(-2h)}{4h^2} = E \left(\frac{e^{ih\xi} - e^{-ih\xi}}{2h} \right)^2 = E \frac{\sin^2 h\xi}{h^2}.$$

Так как $(\sin^2 h\xi)/h^2 \rightarrow \xi^2$ при $h \rightarrow 0$, то по лемме Фату

$$-\varphi''(0) = \lim_{h \rightarrow 0} \left(\frac{2\varphi(0) - \varphi(2h) - \varphi(-2h)}{4h^2} \right) = \lim_{h \rightarrow 0} \frac{E \sin^2 h\xi}{h^2} \geq E \lim_{h \rightarrow 0} \frac{\sin^2 h\xi}{h^2} = E\xi^2.$$

6. Если $\xi \geq 0$, то $\varphi_\xi(\lambda)$ определена в плоскости комплексного переменного λ при $\operatorname{Im} \lambda \geq 0$. При этом $|\varphi_\xi(\lambda)| \leq 1$, $\varphi_\xi(\lambda)$ есть аналитическая функция в области $\operatorname{Im} \lambda > 0$, непрерывная, включая границу $\operatorname{Im} \lambda = 0$.

Аналитичность следует из возможности при $\operatorname{Im} \lambda > 0$ дифференцирования под знаком интеграла

$$\varphi_\xi(\lambda) = \int_0^\infty e^{i\lambda x} dF(x).$$

Непрерывность доказывается так же, как в свойстве 4. Сказанное означает, что для неотрицательных ξ х.ф. $\varphi_\xi(\lambda)$ однозначно определяет функцию $\beta(s)$ вещественного переменного $s \geq 0$, равную $\beta(s) = \varphi_\xi(is) = Ee^{-s\xi}$ и называемую преобразованием Лапласа (или Лапласа—Стилтьеса) над распределением ξ .

Из свойств аналитических функций вытекает и обратное утверждение: преобразование Лапласа $\beta(s)$ на полуоси $s \geq 0$ однозначно определяет х.ф. $\varphi_\xi(\lambda)$.

7. $\overline{\varphi}_\xi(t) = \varphi_\xi(-t) = \varphi_{-\xi}(t)$, где черта означает комплексную сопряженность.

Доказательство следует из равенств

$$\overline{\varphi}_\xi(t) = \overline{Ee^{it\xi}} = E\overline{e^{it\xi}} = Ee^{-it\xi}.$$

Отсюда вытекает, что если ξ симметрична (распределена так же, как $-\xi$), то ее х.ф. вещественна. Можно показать и обратное, если воспользоваться теоремой единственности, о которой речь будет идти ниже.

Найдем теперь х.ф. основных законов распределений.

Пример 1. Если $\xi = a$ с вероятностью 1, т.е. $\xi \in \mathbf{I}_a$, то $\varphi_\xi(t) = e^{ita}$.

Пример 2. Если $\xi \in \mathbf{B}_p$, то $\varphi_\xi(t) = pe^{it} + (1-p) = 1 + p(e^{it} - 1)$.

Пример 3. Если $\xi \in \Phi_{0,1}$, то

$$\varphi(t) = \varphi_\xi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{itx - x^2/2} dx.$$

Дифференцируя по t и интегрируя по частям ($xe^{-x^2/2} dx = -de^{-x^2/2}$), получим

$$\varphi'(t) = \frac{1}{\sqrt{2\pi}} \int ixe^{itx - x^2/2} dx = -\frac{1}{\sqrt{2\pi}} \int te^{itx - x^2/2} dx = -t\varphi(t),$$

$$(\ln \varphi(t))' = -t, \quad \ln \varphi(t) = -\frac{t^2}{2} + c.$$

Так как $\varphi(0) = 1$, то $c = 0$, $\varphi(t) = e^{-t^2/2}$.

Пусть теперь η — нормально распределенная случайная величина с параметрами (a, σ) . Тогда ее можно представить в виде $\eta = \sigma\xi + a$, где ξ нормально распределена с параметрами $(0, 1)$. Х.ф. η определяется по свойству 2:

$$\varphi_\eta(t) = e^{ita} e^{-(t\sigma)^2/2} = e^{ita - t^2\sigma^2/2}.$$

Дифференцируя $\varphi_\eta(t)$ для $\eta \in \Phi_{0,\sigma^2}$, получим $E\eta^k = 0$ при нечетных k и $E\eta^k = \sigma^k(k-1)(k-3)\dots 1$ при $k = 2, 4, \dots$.

Пример 4. Если $\xi \in \Pi_\mu$, то

$$\varphi_\xi(t) = Ee^{it\xi} = \sum_k e^{itk} \frac{\mu^k}{k!} e^{-\mu} = e^{-\mu} \sum_k \frac{(\mu e^{it})^k}{k!} = e^{-\mu} e^{\mu e^{it}} = \exp[\mu(e^{it} - 1)].$$

Пример 5. Если ξ имеет показательное распределение Γ_α с плотностью $\alpha e^{-\alpha x}$ при $x \geq 0$, то

$$\varphi_\xi(t) = \alpha \int_0^\infty e^{itx - \alpha x} dx = \frac{\alpha}{\alpha - it}.$$

Значит, если ξ имеет «двустороннее» показательное распределение с плотностью $1/2 e^{-|x|}$, $-\infty < x < \infty$, то

$$\varphi_\xi(t) = \frac{1}{2} \left(\frac{1}{1 - it} + \frac{1}{1 + it} \right) = \frac{1}{1 + t^2}.$$

Если ξ имеет геометрическое распределение, $P(\xi = k) = (1 - p)p^k$, $k = 0, 1, \dots$, то

$$\varphi_\xi(t) = \frac{1 - p}{1 - pe^{it}}.$$

Пример 6. Если $\xi \in K_{0,1}$ (имеет плотность $1/(\pi(1+x^2))$), то $\varphi_\xi(t) = e^{-|t|}$. Читатель может легко доказать это чуть позже, пользуясь формулой обращения и примером 5.

Пример 7. Если $\xi \in U_{0,1}$, то

$$\varphi_\xi(t) = \int_0^1 e^{itx} dx = \frac{e^{it} - 1}{it}.$$

Х. ф. сумм $\xi_1 + \xi_2$, $\xi_1 + \xi_2 + \xi_3$, \dots , рассматривавшихся в примере 3.5, будут равны в силу свойств 3

$$\varphi_{\xi_1 + \xi_2}(t) = -\frac{(e^{it} - 1)^2}{t^2}, \quad \varphi_{\xi_1 + \xi_2 + \xi_3}(t) = -\frac{(e^{it} - 1)^3}{t^3}, \dots$$

Как установить, является ли та или иная функция φ характеристической? Иногда это можно сделать с помощью указанных свойств. Предлагаем читателю установить, являются ли характеристическими функции $(1+t)^{-1}$, $1+t$, $\sin t$, $\cos t$ и, если являются, то каким распределениям соответствуют.

В общем случае ответ на поставленный вопрос является сложным. Сформулируем здесь без доказательства один из известных результатов.

Теорема Бохнера—Хинчина. Для того, чтобы непрерывная функция $\varphi(t)$, обладающая свойством $\varphi(0) = 1$, была характеристической, необходимо и достаточно, чтобы она была неотрицательно определенной, т. е. чтобы для любых действительных t_1, \dots, t_n и любых комплексных чисел $\lambda_1, \dots, \lambda_n$ выполнялось $(\bar{\lambda} - \text{сопряженное к } \lambda)$

$$\sum_{k,j=1}^n \varphi(t_k - t_j) \lambda_k \bar{\lambda}_j \geq 0.$$

Отметим, что необходимость этого условия почти очевидна, так как если $\varphi(t) = E e^{it\xi}$, то

$$\sum_{k,j=1}^n \varphi(t_k - t_j) \lambda_k \bar{\lambda}_j = E \sum_{k,j=1}^n e^{i(t_k - t_j)\xi} \lambda_k \bar{\lambda}_j = E \left| \sum_{k=1}^n \lambda_k e^{it_k \xi} \right|^2 \geq 0.$$

Укажем теперь еще два свойства х. ф.

8. Если распределение ξ имеет плотность, то $\varphi_\xi(t) \rightarrow 0$ при $t \rightarrow \infty$.

Это есть прямое следствие теоремы Лебега о преобразованиях Фурье. Обратное утверждение неверно. В целом же скорость стремления к нулю функции $\varphi_\xi(t)$ при $|t| \rightarrow \infty$ тем выше, чем выше гладкость $F(x)$. Формулы примера 7 в этом смысле типичны. Если плотность $f(x)$ имеет интегрируемую k -ю производную, то интегрированием по частям получаем

$$\varphi_\xi(t) = \int e^{itx} f(x) dx = \frac{1}{it} \int e^{itx} f'(x) dx = \dots = \frac{1}{(it)^k} \int e^{itx} f^{(k)}(x) dx.$$

Отсюда следует, что

$$|\varphi_\xi(t)| \leq \frac{c}{|t|^k}.$$

9. Будем называть распределение ξ *решетчатым с шагом решетки h* , если существуют a и h такие, что

$$\sum_{k=-\infty}^{\infty} P(\xi = a + kh) = 1.$$

Случайная величина ξ имеет решетчатое распределение с шагом $h > 0$ тогда и только тогда, когда

$$\left| \varphi_\xi \left(\frac{2\pi}{h} \right) \right| = 1.$$

Действительно, если ξ решетчата, то

$$\varphi_\xi(t) = e^{ita} \sum_k P(\xi = a + kh) e^{ikh}$$

и, следовательно,

$$\varphi_\xi \left(\frac{2\pi}{h} \right) = e^{2\pi ia/h}. \quad (2)$$

Обратно, если верно (2), то при $t = 2\pi/h$

$$\varphi_{\xi-a}(t) = E \cos t(\xi - a) + iE \sin t(\xi - a) = 1.$$

Следовательно, $E(1 - \cos t(\xi - a)) = 0$ и по свойству E4 $\cos t(\xi - a) = 1$, $(\xi - a)2\pi/h = 2\pi k$, $k = k(\omega)$, с вероятностью 1.

Среди решетчатых распределений важное место занимают распределения целочисленных случайных величин. Для них a и h — целые числа, и мы можем положить для простоты $a = 0$, $h = 1$. Тогда $\varphi_\xi(t)$ будет функцией переменной $z = e^{it}$ и будет периодической по t с периодом 2π . Значит, в этом случае достаточно знать поведение х. ф. на отрезке $[-\pi, \pi]$, или, что то же, знать поведение функции

$$\psi_\xi(z) = E z^\xi = \sum z^k P(\xi = k)$$

на единичной окружности $|z| = 1$. Функция $\psi_\xi(z)$ называется *производящей функцией случайной величины* ξ (или распределения ξ). Так как $\psi_\xi(e^{it}) = \varphi_\xi(t)$ есть х. ф., то для производящих функций остаются справедливыми все свойства х. ф. с теми лишь изменениями, которые вытекают из замены аргумента. Подробнее о применении аппарата производящих функций см. § 8.

§ 2. Формула обращения

Итак, для каждой случайной величины определена соответствующая ей х. ф. Покажем теперь, что класс \mathcal{L} функций e^{itx} определяет распределение, т. е. что по х. ф. однозначно восстанавливается распределение. Этот факт доказывает формула обращения.

Теорема 1 (формула обращения). Если $F(x)$ — функция распределения случайной величины ξ , а $\varphi(t)$ — х. ф. ξ , то для любых точек непрерывности x и y функции $F(x)$

$$F(y) - F(x) = \frac{1}{2\pi} \lim_{\sigma \rightarrow 0} \int \frac{e^{-itz} - e^{-ity}}{it} \varphi(t) e^{-t^2 \sigma^2} dt. \quad (3)$$

Если функция $\varphi(t)/t$ интегрируема на бесконечности, то становится законным предельный переход под знаком интеграла, и можно записать

$$F(y) - F(x) = \frac{1}{2\pi} \int \frac{e^{-itz} - e^{-ity}}{it} \varphi(t) dt. \quad (4)$$

Доказательство. Предположим сначала, что существует плотность $f(x)$, а х. ф. $\varphi(t)$ интегрируема. Тогда в силу известной формулы обращения для преобразования Фурье

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \varphi(t) dt. \quad (5)$$

Утверждение теоремы в этом случае получится, если проинтегрировать обе части этого равенства в пределах от x до y и поменять в правой части (в силу абсолютной сходимости) порядок интегрирования¹⁾.

Пусть теперь $\varphi(t)$ — характеристическая функция ξ с произвольным распределением $F(x)$. Рассмотрим на одном вероятностном пространстве с ξ случайную величину η , независимую от ξ и распределенную нормально с параметрами $(0, 2\sigma^2)$. Как мы уже знаем, характеристическая функция η будет равна $e^{-t^2 \sigma^2}$.

¹⁾ В литературе формула обращения часто приводится также в виде $F(y) - F(x) = \frac{1}{2\pi} \lim_{A \rightarrow \infty} \int_{-A}^A \frac{e^{-itz} - e^{-ity}}{it} \varphi(t) dt$, эквивалентном (3).

²⁾ Формулу (4) можно получить из (5) и без интегрирования, заметив, что $\frac{F(y) - F(x)}{y - x}$ есть значение в нуле плотности, равной свертке двух плотностей: $f(x)$ и плотности равномерного на $[-y, -x]$ распределения (замечание по этому поводу см. также в конце § 3.6). Х. ф. этой свертки равна $\frac{e^{-itz} - e^{-ity}}{(y-x)it} \varphi(t)$.

Это означает, что характеристическая функция $\xi + \eta$, равная $\varphi(t)e^{-t^2\sigma^2}$, интегрируема. Поэтому будет справедливо равенство

$$F_{\xi+\eta}(y) - F_{\xi+\eta}(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{-itx} - e^{-ity}}{it} \varphi(t) e^{-t^2\sigma^2} dt. \quad (6)$$

Так как $P(|\eta| > \delta) \rightarrow 0$ при любом $\delta > 0$ и $\sigma \rightarrow 0$, то в силу главы 6 $F_{\xi+\eta} \Rightarrow F$ (если бы была последовательность $\sigma_k \rightarrow 0$ и каждому σ_k соответствовала бы своя случайная величина η_k , то, очевидно, мы имели бы $\xi + \eta_k \xrightarrow{p} \xi$; $F_{\xi+\eta_k} \Rightarrow F$).

Значит, если x и y есть точки непрерывности F , то $F(y) - F(x) = \lim_{\sigma \rightarrow 0} (F_{\xi+\eta}(y) - F_{\xi+\eta}(x))$. Это вместе с (6) доказывает утверждение теоремы. ■

В доказательстве этой теоремы использован прием, который можно назвать «сглаживанием» распределений. Он часто применяется для преодоления технических трудностей, связанных с формулой обращения.

Следствие 1 (теорема единственности). *Х.ф. случайной величины однозначно определяет ее функцию распределения.*

Доказательство следует из формулы обращения и из того, что разности $F(y) - F(x)$ однозначно определяют $F(x)$. ■

Для решетчатых распределений формула обращения упрощается. Пусть для простоты ξ — целочисленная случайная величина и $\psi_{\xi}(z) = Ez^{\xi}$ есть ее производящая функция. Тогда

$$P(\xi = k) = \frac{1}{2\pi i} \int_{|z|=1} \psi_{\xi}(z) z^{-k-1} dz. \quad (7)$$

Действительно, переходя к х.ф. $\varphi_{\xi}(t) = \sum_j e^{itj} P(\xi = j)$ и делая замену в (7) $z = it$, получим, что интеграл в правой части (7) равен

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-itk} \varphi_{\xi}(t) dt = \frac{1}{2\pi} \sum_j P(\xi = j) \int_{-\pi}^{\pi} e^{it(j-k)} dt.$$

Здесь все интегралы в правой части, очевидно, равны нулю, кроме интеграла при $j = k$, равного 2π . ■

Формула (7) есть не что иное, как формула для коэффициентов ряда Фурье, и имеет простой геометрический смысл.

Функции $\{e_k = e^{ik}\}$ образуют ортонормированный базис в гильбертовом пространстве $\mathcal{L}_2(-\pi, \pi)$ комплекснозначных функций, интегрируемых с квадратом, со скалярным произведением

$$(f, g) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t) \bar{g}(t) dt$$

(\bar{g} — комплексно-сопряженное к g). Если $\varphi_\xi = \sum e_k P(\xi = k)$, то из равенства $\varphi_\xi = \sum e_k (\varphi_\xi, e_k)$ немедленно следует, что

$$P(\xi = k) = (\varphi_\xi, e_k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-itk} \varphi_\xi(t) dt.$$

§ 3. Теорема непрерывности (сходимости)

Пусть $\{\varphi_n(t)\}_{n=1}^{\infty}$ — последовательность х. ф., а $\{F_n\}_{n=1}^{\infty}$ — последовательность соответствующих функций распределения. Напомним, что знак \Rightarrow означает слабую сходимость распределений, определенную нами в главе 6.

Теорема 2 (теорема непрерывности). *Для сходимости $F_n \Rightarrow F$ необходимо и достаточно, чтобы $\varphi_n(t) \rightarrow \varphi(t)$ при каждом t , где $\varphi(t)$ — х. ф., соответствующая F .*

Теорема очевидным образом вытекает из следствия 7 предыдущей главы (здесь выполнены сразу два из достаточных условий следствия 6.7: условия 2) и 3)). Ее доказательство будет получено также более простым способом в § 4.

В § 1 для неотрицательных случайных величин ξ было введено понятие преобразования Лапласа $\beta(s) = E e^{-s\xi}$. Пусть $\beta_n(s)$, $\beta(s)$ — преобразования Лапласа, соответствующие F_n и F . Справедливо аналогичное теореме 2 утверждение:

Для $F_n \Rightarrow F$ необходимо и достаточно, чтобы $\beta_n(s) \rightarrow \beta(s)$ при каждом $s \geq 0$.

Как и теорема 2, это утверждение вытекает из следствия 6.7, так как класс функций $\{f(x) = e^{-sx}\}$, $s \geq 0$ (как и $\{e^{itz}\}$), определяет распределение (см. свойство 6 из § 1) и, кроме того, выполнены достаточные условия 2), 3) следствия 6.7.

Теорема 2 обладает одним недостатком — в ней заранее требуется знать, что функция $\varphi(t)$, к которой сходится х. ф. $\varphi_n(t)$, является характеристической. Между тем таких сведений может и не быть (см., например, § 8.9). В связи с этим возникает естественный вопрос, при каких условиях предельная функция $\varphi(t)$ будет характеристической.

Ответ на этот вопрос содержится в следующей теореме.

Теорема 3. *Пусть $\varphi_n(t) = \int e^{itz} dF_n(x)$ есть последовательность характеристических функций и $\varphi_n(t) \rightarrow \varphi(t)$ при $n \rightarrow \infty$ и при каждом t .*

Тогда следующие условия эквивалентны:

- $\varphi(t)$ является х. ф.,
- $\varphi(t)$ непрерывна в точке $t = 0$,
- последовательность $\{F_n\}$ является плотной.

Таким образом, если мы установим, что $\varphi_n(t) \rightarrow \varphi(t)$ и выполнено одно из трех перечисленных свойств, то можно утверждать, что существует распределение F такое, что φ является х. ф. F и $F_n \Rightarrow F$.

Доказательство. Эквивалентность условий а) и в) вытекает из теоремы 6.9. То, что из а) следует б), нам известно. Осталось установить, что из б) следует в). Для этого покажем сначала, что справедлива

Лемма 1. *Если φ есть х. ф. ξ , то при любом $u > 0$*

$$P\left(|\xi| > \frac{2}{u}\right) \leq \frac{1}{u} \int_{-u}^u [1 - \varphi(t)] dt.$$

Доказательство. Правая часть этого неравенства равна

$$\frac{1}{u} \int_{-u}^u \int_{-\infty}^{\infty} (1 - e^{-itx}) dF(x) dt,$$

где F — функция распределения ξ . Меняя порядок интегрирования и замечая, что

$$\int_{-u}^u (1 - e^{-itx}) dt = \left(t + \frac{e^{-itx}}{ix} \right) \Big|_{-u}^u = 2u \left(1 - \frac{\sin ux}{ux} \right),$$

получим

$$\begin{aligned} \frac{1}{u} \int_{-u}^u [1 - \varphi(t)] dt &= 2 \int_{-\infty}^{\infty} \left(1 - \frac{\sin ux}{ux} \right) dF(x) \geq \\ &\geq 2 \int_{|x| \geq 2/u} \left(1 - \left| \frac{\sin ux}{ux} \right| \right) dF(x) \geq \\ &\geq 2 \int_{|x| \geq 2/u} \left(1 - \frac{1}{|ux|} \right) dF(x) \geq \int_{|x| \geq 2/u} dF(x). \end{aligned}$$

Пусть теперь выполнено условие б). В силу леммы 1

$$\limsup_{n \rightarrow \infty} \int_{|x| \geq 2/u} dF_n(x) \leq \limsup_{n \rightarrow \infty} \frac{1}{u} \int_{-u}^u [1 - \varphi_n(t)] dt = \frac{1}{u} \int_{-u}^u [1 - \varphi(t)] dt.$$

Так как $\varphi(t)$ непрерывна в точке 0 и $\varphi(0) = 1$, то среднее значение интеграла, стоящего в правой части, может быть сделано выбором u сколь угодно малым. Это означает, очевидно, выполнение условия в). ■

С помощью х.ф. можно не только устанавливать факт сходимости распределений, но и оценивать также скорость этой сходимости в тех случаях, когда удастся оценить скорость убывания $\varphi_n - \varphi$. С соответствующими примерами мы столкнемся в § 5.

Основные применения аппарата х.ф. встречаются в главах 8, 11, 16. В этой главе мы также коснемся применений х.ф., однако более иллюстративного характера. Прежде чем переходить к ним, рассмотрим еще один подход (возможно, он покажется читателю более простым) к доказательству теоремы 2 о сходимости распределений.

§ 4*. Другой подход к доказательству теорем сходимости к известному распределению

При доказательстве с помощью х.ф. сходимости $F_n \Rightarrow F$ к известному распределению F вместо схемы, изложенной в § 6.3, 6.2, 7.3, можно предложить другую, которая может оказаться в данном случае предпочтительнее. Она состоит в следующем.

1. После определения 6.6 слабой сходимости следует доказать возможность сужения класса функций f , для которых надо проверять сходимость

$$\int f dF_n \rightarrow \int f dF. \quad (8)$$

Однако сделать это нужно «непосредственно», а не опираясь на теорему 6.6, как это было в § 6.3. Таким классом может быть любой класс, определяющий распределение, например, класс C_k^0 финитных (отличных от нуля на конечном отрезке) k раз непрерывно дифференцируемых функций при $k \geq 2$.

Схема доказательства возможности сужения такова. Пусть C — класс всех непрерывных ограниченных функций на прямой, C^0 — класс непрерывных финитных функций.

Утверждение 1. Если соотношение (8) верно для любой $f \in C^0$, то оно верно для любой $f \in C$.

Доказательство. Выберем $g_\epsilon \in C^0$, $0 \leq g_\epsilon \leq 1$, так, чтобы

$$\int (1 - g_\epsilon(x)) dF(x) \leq \epsilon.$$

Это всегда возможно, так как для любой функции $g \in C^0$, $0 \leq g \leq 1$, $g(0) = 1$, выполняется $1 - g(\delta x) \rightarrow 0$ при $\delta \rightarrow 0$, и по теореме о мажорируемой сходимости

$$\int (1 - g(\delta x)) dF(x) \rightarrow 0.$$

Пусть $f \in C$, $\|f\| = \sup_x |f(x)|$ и (8) справедливо для $f \in C^0$. Тогда $fg_\epsilon \in C^0$,

$$\left| \int (1 - g_\epsilon) f dF_n \right| \leq \|f\| \left(1 - \int g_\epsilon dF_n \right) \rightarrow \|f\| \left(1 - \int g_\epsilon dF \right) \leq \epsilon \|f\|. \quad (9)$$

Следовательно, в силу равенства $f = fg_\epsilon + f(1 - g_\epsilon)$

$$\begin{aligned} \limsup \left| \int f dF_n - \int f dF \right| &\leq \limsup \left[\left| \int (1 - g_\epsilon) f dF_n \right| + \right. \\ &\quad \left. + \left| \int fg_\epsilon dF_n - \int fg_\epsilon dF \right| + \left| \int (1 - g_\epsilon) f dF \right| \right] \leq 2\epsilon \|f\|. \end{aligned}$$

Так как ϵ здесь произвольно, то для $f \in C$ справедливо (8). ■

В (9) использован тот факт, что $\int dF = 1$.

Утверждение 2. Если соотношение (8) верно для любой $f \in C_k^0$, $k \geq 1$, то оно верно и для любой $f \in C$.

Доказательство в силу утверждения 1 очевидно, так как $f \in C^0$ можно равномерно приблизить функцией $f_{(\delta)} \in C_k^0$. Можно рассмотреть, например, функции

$$f_{(\delta)}(x) = \frac{1}{2\delta} \int_{x-\delta}^{x+\delta} f(t) dt,$$

равномерно близкие к $f \in C$ и дифференцируемые. Повторив эту операцию, получим дважды дифференцируемую функцию и т. д. ■

Для финитной функции $f \in C_k^0$

$$\int f dF = - \int F df = - \int F f' dx, \quad f' \in C_{k-1}^0.$$

Следовательно, критерий (8) можно записать также в форме

$$\int f F_n dx \rightarrow \int f F dx, \quad f \in C_k^0, \quad k \geq 0. \quad (10)$$

Отметим, что доказательство теоремы 6.9 после этого можно упростить (мы уже отмечали это), так как из сходимости $F_n(x) \rightarrow F(x)$ в точках непрерывности F (следовательно, всюду, за исключением, быть может, счетного множества) вытекает сходимость (10) по теореме о мажорируемой сходимости.

Далее вводится понятие х.ф., как это сделано в § 1, и в числе ее свойств нужно доказать равенство Парсеваля.

Утверждение 3. Если $g \in C_k^0$, $k \geq 2$, $\xi \in F$, то

$$Eg(\xi) = \frac{1}{2\pi} \int \varphi_\xi(-t) \tilde{g}(t) dt,$$

где $\varphi_\xi(t) = \int e^{itx} dF(x)$, $\tilde{g}(t) = \int e^{itx} g(x) dx$.

Если F имеет плотность h , то равенство Парсеваля записывается в более привычной в математическом анализе форме:

$$\int g(x) h(x) dx = \frac{1}{2\pi} \int \tilde{g}(t) \tilde{h}(-t) dt.$$

Доказательство. Так как $g \in C_k^0$, $k \geq 2$, то в силу свойства 8 характеристических функций \tilde{g} абсолютно интегрируема. Применяя к $\tilde{g}(t)$ формулу обращения, получим

$$Eg(\xi) = E \frac{1}{2\pi} \int e^{-it\xi} \tilde{g}(t) dt = \frac{1}{2\pi} \int E e^{-it\xi} \tilde{g}(t) dt = \frac{1}{2\pi} \int \varphi_\xi(-t) \tilde{g}(t) dt. \quad \blacksquare$$

Мы можем предложить теперь новый вариант доказательства теоремы непрерывности 2.

Доказательство теоремы 2. Пусть $\varphi_n \rightarrow \varphi$ (φ_n и φ суть х.ф. F_n и F соответственно). Тогда для $g \in C_k^0$

$$\int g dF_n = \frac{1}{2\pi} \int \varphi_n(-t) \tilde{g}(t) dt, \quad (11)$$

где функция $|\varphi_n \tilde{g}| \leq |\tilde{g}|$ интегрируема. Поэтому по теореме о мажорируемой сходимости правая часть в (11) сходится к

$$\frac{1}{2\pi} \int \varphi(-t) \tilde{g}(t) dt = \int g dF. \quad \blacksquare$$

§ 5. Применение характеристических функций в теореме Пуассона

Пусть ξ_1, \dots, ξ_n — независимые целочисленные случайные величины, $S_n = \sum_{k=1}^n \xi_k$, $P(\xi_k = 1) = p_k$, $P(\xi_k = 0) = 1 - p_k - q_k$. Формулируемая ниже теорема представляет собой некоторое обобщение³⁾ теорем, полученных в § 5.4.

Теорема 4.

$$|P(S_n = k) - \Pi_\mu(\{k\})| \leq \sum_{k=1}^n p_k^2 + 2 \sum_{k=1}^n q_k,$$

где $\mu = \sum_{k=1}^n p_k$.

Таким образом, если заданы серии

$$\xi_{1n}, \xi_{2n}, \dots, \xi_{nn}, \quad n = 1, 2, \dots,$$

независимых целочисленных случайных величин

$$S_n = \sum_{k=1}^n \xi_{kn}, \quad P(\xi_{kn} = 1) = p_{kn},$$

$$P(\xi_{kn} = 0) = 1 - p_{kn} = q_{kn}, \quad \mu = \sum_{k=1}^n p_{kn},$$

то для сходимости разности $P(S_n = k) - \Pi_\mu(\{k\})$ к нулю достаточно, чтобы

$$\sum_{k=1}^n q_{kn} \rightarrow 0, \quad \sum_{k=1}^n p_{kn}^2 \rightarrow 0.$$

Так как $\sum_{k=1}^n p_{kn}^2 \leq \mu \max_{k \leq n} p_{kn}$, то последнее условие всегда выполнено, если $\max_{k \leq n} p_{kn} \rightarrow 0$, $\mu \leq \mu_0 = \text{const.}$

³⁾ Это обобщение не очень существенно, так как результаты, близкие к теореме 4, можно получить и пользуясь теоремой 5.4, в которой ξ_k может принимать лишь значения нуль и единица. Достаточно заметить, что вероятность события $A = \bigcup_k \{\xi_k \neq 0, \xi_k \neq 1\}$ оценивается значением $\sum q_k$ и, следовательно,

$$P(S_n = k) = \theta_1 \sum q_k + \left(1 - \theta_2 \sum q_k\right) P(S_n = k | \bar{A}), \quad \theta_i \leq 1, \quad i = 1, 2,$$

где $P(S_n = k | \bar{A}) = P(S_n^* = k)$, S_n^* образованы случайными величинами ξ_k^* ,

$$P(\xi_k^* = 1) = p_k^* = \frac{p_k}{1 - q_k}, \quad P(\xi_k^* = 0) = 1 - p_k^*.$$

Лемма 2. Если $\operatorname{Re} \beta \leq 0$, то

$$|e^\beta - 1| \leq |\beta|, \quad |e^\beta - 1 - \beta| \leq \frac{|\beta|^2}{2},$$

$$\left| e^\beta - 1 - \beta - \frac{\beta^2}{2} \right| \leq \frac{|\beta|^3}{6}.$$

Доказательство. Требуемые неравенства вытекают из соотношений (в них используется замена $t = \beta v$ и тот факт, что $|e^s| \leq 1$ при $\operatorname{Re} s \leq 0$)

$$|e^\beta - 1| = \left| \int_0^\beta e^t dt \right| = \left| \beta \int_0^1 e^{\beta v} dv \right| \leq |\beta|,$$

$$|e^\beta - 1 - \beta| = \left| \int_0^\beta (e^t - 1) dt \right| = \left| \beta \int_0^1 (e^{\beta v} - 1) dv \right| \leq |\beta|^2 \int_0^1 v dv = \frac{|\beta|^2}{2}.$$

Последнее неравенство доказывается аналогично. ■

Лемма 3. Если $|a_k| \leq 1$, $|b_k| \leq 1$; $k = 1, \dots, n$, то

$$\left| \prod_{k=1}^n a_k - \prod_{k=1}^n b_k \right| \leq \sum_{k=1}^n |a_k - b_k|.$$

Таким образом, если $\varphi_k(t)$, $\psi_k(t)$ — х. ф., то при любом t

$$\left| \prod_{k=1}^n \varphi_k(t) - \prod_{k=1}^n \psi_k(t) \right| \leq \sum_{k=1}^n |\varphi_k(t) - \psi_k(t)|.$$

Доказательство. Обозначим $A_n = \prod_{k=1}^n a_k$, $B_n = \prod_{k=1}^n b_k$. Тогда $|A_n| \leq 1$, $|B_n| \leq 1$,

$$|A_n - B_n| = |A_{n-1}a_n - B_{n-1}b_n| =$$

$$= |(A_{n-1} - B_{n-1})a_n + (a_n - b_n)B_{n-1}| \leq |A_{n-1} - B_{n-1}| + |a_n - b_n|.$$

Применяя это неравенство n раз, получим требуемое соотношение. ■

Доказательство теоремы 4. Имеем

$$\varphi_k(t) \equiv \mathbb{E} e^{it\xi_k} = 1 + p_k(e^{it} - 1) + q_k(\gamma_k(t) - 1),$$

где $\gamma_k(t)$ есть х. ф. некоторой целочисленной случайной величины. В силу независимости случайных величин ξ_k

$$\varphi_{s_n}(t) = \prod_{k=1}^n \varphi_k(t).$$

Пусть далее $\zeta \in \Pi_\mu$. Тогда

$$\varphi_\zeta(t) = \mathbb{E} e^{it\zeta} = e^{\mu(e^{it}-1)} = \prod_{k=1}^n \psi_k(t),$$

где $\psi_k(t) = e^{p_k(e^{it}-1)}$. Следовательно, разность х. ф. φ_{S_n} и φ_ζ по лемме 3 можно оценить следующим образом:

$$|\varphi_{S_n}(t) - \varphi_\zeta(t)| = \left| \prod_{k=1}^n \varphi_k - \prod_{k=1}^n \psi_k \right| \leq \sum_{k=1}^n |\varphi_k - \psi_k|,$$

где по лемме 2 ($\operatorname{Re}(e^{it} - 1) \leq 0$)

$$\begin{aligned} |\psi_k(t) - 1 - p_k(e^{it} - 1)| &\leq \frac{p_k^2 |e^{it} - 1|^2}{2} = \\ &= \frac{p_k^2}{2} (\sin^2 t + (1 - \cos t)^2) = p_k^2 \left(\frac{\sin^2 t}{2} + 2 \sin^4 \frac{t}{2} \right), \end{aligned} \quad (12)$$

$$\sum_{k=1}^n |\varphi_k - \psi_k| \leq 2 \sum_{k=1}^n q_k + \sum_{k=1}^n p_k^2 \left(\frac{\sin^2 t}{2} + 2 \sin^4 \frac{t}{2} \right).$$

Остается воспользоваться формулой обращения (7), в силу которой

$$\begin{aligned} |P(S_n = k) - \Pi_\mu(\{k\})| &= \left| \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-itk} (\varphi_{S_n}(t) - \varphi_\zeta(t)) dt \right| \leq \\ &\leq \frac{1}{\pi} \int_0^{\pi} \left[2 \sum_{k=1}^n q_k + \sum_{k=1}^n p_k^2 \left(\frac{\sin^2 t}{2} + 2 \sin^4 \frac{t}{2} \right) \right] dt = 2 \sum_{k=1}^n q_k + \sum_{k=1}^n p_k^2, \end{aligned}$$

так как $\frac{1}{2\pi} \int_0^{\pi} \sin^2 t dt = \frac{1}{4}$, $\frac{2}{\pi} \int_0^{\pi} \sin^4 \frac{t}{2} dt = \frac{3}{4}$. ■

Если в (12) воспользоваться неравенством $|e^{it} - 1| \leq 2$, то выкладки упростятся, необходимость в вычислении двух последних интегралов отпадет, но оценки будут несколько хуже:

$$\begin{aligned} \sum |\varphi_k - \psi_k| &\leq 2 \left(\sum q_k + \sum p_k^2 \right), \\ |P(S_n = k) - \Pi_\mu(\{k\})| &\leq 2 \left(\sum q_k + \sum p_k^2 \right). \end{aligned}$$

§ 6. Характеристические функции многомерных распределений.

Многомерное нормальное распределение

Определение 2. Если дан случайный вектор $\xi = (\xi_1, \xi_2, \dots, \xi_n)$, то его х. ф. (х. ф. распределения ξ) называется функция от вектора $t = (t_1, \dots, t_n)$, равная

$$\begin{aligned} \varphi_\xi(t) &= E e^{it\xi^T} = E e^{i(t, \xi)} = E \exp \left\{ i \sum_{k=1}^n t_k \xi_k \right\} = \\ &= \int \exp \left\{ i \sum_{k=1}^n t_k x_k \right\} P_{\xi_1, \dots, \xi_n}(dx_1, \dots, dx_n), \end{aligned}$$

здесь ξ^T — вектор-столбец, транспонированный к ξ , (t, ξ) — скалярное произведение.

Х. ф. многомерных распределений обладают всеми свойствами (с очевидными изменениями в формулировках), которые излагались выше в § 1-3.

Ясно, что $\varphi_\xi(0) = 1$ и всегда $|\varphi_\xi(t)| \leq 1$, $\varphi_\xi(-t) = \overline{\varphi_\xi(t)}$. Далее, $\varphi_\xi(t)$ всюду непрерывна. Если существует смешанный момент $E\xi_1^{k_1} \cdots \xi_n^{k_n}$, то φ_ξ имеет производную порядка $k_1 + \dots + k_n$:

$$\left. \frac{\partial \varphi_\xi^{k_1 + \dots + k_n}(t)}{\partial t_1^{k_1} \cdots \partial t_n^{k_n}} \right|_{t=0} = i^{k_1 + \dots + k_n} E\xi_1^{k_1} \cdots \xi_n^{k_n}.$$

Если существуют все моменты некоторого порядка, то в окрестности точки $t = 0$ справедливо разложение функции $\varphi_\xi(t)$, аналогичное (1).

Если известна $\varphi_\xi(t)$, то х. ф. любой подсовокупности случайных величин $(\xi_{k_1}, \dots, \xi_{k_j})$, очевидно, можно получить, положив все t_k , за исключением t_{k_1}, \dots, t_{k_j} , равными нулю.

Следующие теоремы являются простыми обобщениями своих одномерных аналогов.

Формула обращения. Если Δ есть «интервал», определенный неравенствами $a_k < x_k < b_k$, $k = 1, \dots, n$, и вероятность $P(\xi \in \Delta)$ непрерывна на гранях этого интервала, то

$$P(\xi \in \Delta) = \lim_{\sigma \rightarrow 0} \frac{1}{(2\pi)^n} \int \cdots \int \left(\prod_{k=1}^n \frac{e^{-it_k a_k} - e^{-it_k b_k}}{it_k} e^{-t_k^2 \sigma^2} \right) \varphi_\xi(t) dt_1 \cdots dt_n.$$

Если существует плотность $f(x)$ случайного вектора ξ , а х. ф. $\varphi_\xi(t)$ интегрируема, то формула обращения может быть записана в форме

$$f(x) = \frac{1}{(2\pi)^n} \int e^{-i(t,x)} \varphi_\xi(t) dt.$$

Если функция $g(x)$ такова, что ее преобразование Фурье

$$\bar{g}(t) = \int e^{i(t,x)} g(x) dx$$

интегрируемо (это всегда так для достаточно гладких $g(x)$), то справедливо равенство Парсеваля

$$Eg(\xi) = E \frac{1}{(2\pi)^n} \int e^{-i(t,\xi)} \bar{g}(t) dt = \frac{1}{(2\pi)^n} \int \varphi_\xi(-t) \bar{g}(t) dt.$$

Из формулы обращения следует, как и раньше, теорема о взаимно однозначном соответствии между х. ф. и функциями распределения и вместе с ней тот факт, что класс функций $\{e^{i(t,x)}\}$ в многомерном случае определяет распределение (ср. с определением 6.9).

Слабая сходимость распределений $P_n(B)$ в n -мерном пространстве к распределению $P(B)$ определяется так же, как в одномерном случае: $P_n \Rightarrow P$, если

$$\int f(x) dP_n(dx) \rightarrow \int f(x) dP(dx)$$

для любой непрерывной и ограниченной функции $f(x)$.

Обозначим $\varphi_n(t)$ и $\varphi(t)$ х. ф. соответственно распределений P_n и P .

Теорема непрерывности 2А. Для слабой сходимости $P_n \Rightarrow P$ необходимо и достаточно, чтобы при каждом t и $n \rightarrow \infty$ $\varphi_n(t) \rightarrow \varphi(t)$.

Если удастся установить сходимость $\varphi_n(t)$ к некоторой функции $\varphi(t)$, то возникает вопрос, будет ли $\varphi(t)$ х. ф. какого-нибудь распределения или, что то же, будет ли последовательность P_n слабо сходиться к некоторому распределению P ? Ответы на эти вопросы содержатся в следующем утверждении. Пусть Δ_N — куб, определяемый неравенством $\max_k |x_k| < N$.

Теорема непрерывности 3А. Пусть последовательность х. ф. $\varphi_n(t)$ сходится к функции $\varphi(t)$ при каждом t . Тогда следующие три условия эквивалентны:

- φ является х. ф.,
- φ непрерывна в точке $t = 0$,
- $\lim_{n \rightarrow \infty} \sup_{x \notin \Delta_N} \int dP_n(x) \rightarrow 0$ при $N \rightarrow \infty$.

Доказательство всех трех теорем проводится так же, как в одномерном случае.

Пример 8. Многомерное нормальное распределение определено как распределение с плотностью (см. § 3.3)

$$f_{\xi}(x) = \frac{\sqrt{|A|}}{(2\pi)^{n/2}} e^{-1/2 Q(x)},$$

где $Q(x) = xAx^T = \sum_{i,j=1}^n a_{i,j} x_i x_j$, $|A|$ есть определитель положительно определенной матрицы $A = \|a_{i,j}\|$.

Это *центрированное* нормальное распределение, для которого $E\xi = 0$. Распределение вектора $\xi + a$ при любом постоянном векторе a также называется нормальным.

Найдем х. ф. ξ . Покажем, что

$$\varphi_{\xi}(t) = e^{-1/2 t M t^T}, \quad (13)$$

где $M = A^{-1}$ есть матрица, обратная к A , совпадающая с матрицей $\|m_{i,j}\|$ вторых моментов распределения ξ :

$$m_{i,j} = E\xi_i \xi_j.$$

В самом деле,

$$\varphi_{\xi}(t) = \frac{\sqrt{|A|}}{(2\pi)^{n/2}} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} e^{itx^T - 1/2 xAx^T} dx_1 \dots dx_n. \quad (14)$$

Выберем ортогональную матрицу C так, чтобы $CAC^T = D$ была бы диагональной матрицей, и обозначим через μ_1, \dots, μ_n значения диагональных элементов. Сделаем замену переменных, положив $x = yC$ и $t = vC$. Тогда

$$|A| = |D| = \prod_{k=1}^n \mu_k,$$

$$itx^T - \frac{1}{2} xAx^T = ivy^T - \frac{1}{2} yDy^T = i \sum_{k=1}^n v_k y_k - \frac{1}{2} \sum_{k=1}^n \mu_k y_k^2,$$

и в силу свойств х. ф. одномерных нормальных распределений

$$\begin{aligned}\varphi_{\xi}(t) &= \frac{\sqrt{|A|}}{(2\pi)^{n/2}} \prod_{k=1}^n \int_{-\infty}^{\infty} e^{i v_k y_k - 1/2 \mu_k y_k^2} dy_k = \\ &= \sqrt{|A|} \prod_{k=1}^n \frac{1}{\sqrt{\mu_k}} e^{-1/2 v_k^2 / \mu_k} = e^{-1/2 v D^{-1} v^T} = e^{-1/2 C^T D^{-1} C t^T} = e^{-1/2 t A^{-1} t^T}.\end{aligned}$$

С другой стороны, так как все моменты ξ существуют, то в окрестности точки $t = 0$

$$\varphi_{\xi}(t) = 1 - \frac{1}{2} t A^{-1} t^T + o\left(\sum t_k^2\right) = 1 + i t E \xi^T + \frac{t^2}{2} t M t^T + o\left(\sum t_k^2\right).$$

Отсюда следует, что $E \xi = 0$, $A^{-1} = M$.

Из доказанной формулы (13) вытекает следующее свойство нормальных распределений: *компоненты вектора (ξ_1, \dots, ξ_n) независимы тогда и только тогда, когда коэффициенты корреляции $\rho(\xi_i, \xi_j)$ равны нулю при всех $i \neq j$* . Действительно, если M — диагональная матрица, то $A = M^{-1}$ также диагональна и $f_{\xi}(x)$ равна произведению плотностей. Обратно, если (ξ_1, \dots, ξ_n) независимы, то A диагональна и, стало быть, диагональна M .

§ 7. Другие применения х. ф.

Свойства гамма-распределения

1. Свойство устойчивости распределений Φ_{α, σ^2} , $K_{\alpha, \sigma}$. Свойство устойчивости, грубо говоря, состоит в сохранении типа распределения при суммировании случайных величин (это описание устойчивости не точно, подробнее об этом см. § 8.6).

Сумма независимых случайных величин, распределенных по нормальному закону, снова имеет нормальное распределение. Действительно, пусть ξ_1 и ξ_2 независимы и нормально распределены соответственно с параметрами (α_1, σ_1^2) , (α_2, σ_2^2) . Тогда х. ф. $\xi_1 + \xi_2$ равна

$$\varphi_{\xi_1 + \xi_2}(t) = \varphi_{\xi_1}(t) \varphi_{\xi_2}(t) = e^{i t \alpha_1 - t^2 \sigma_1^2 / 2} e^{i t \alpha_2 - t^2 \sigma_2^2 / 2} = e^{i t (\alpha_1 + \alpha_2) - t^2 (\sigma_1^2 + \sigma_2^2) / 2}.$$

Таким образом, сумма $\xi_1 + \xi_2$ снова является нормально распределенной случайной величиной с параметрами $(\alpha_1 + \alpha_2, \sigma_1^2 + \sigma_2^2)$.

Нормальность распределения сохраняется и при суммировании *зависимых* величин (компонент произвольного нормально распределенного вектора). Это немедленно следует из вида х. ф. многомерного нормального закона, найденного в § 6. Надо лишь заметить, что для получения х. ф. $\xi_1 + \dots + \xi_n$ достаточно в выражении для

$$\varphi_{(\xi_1, \dots, \xi_n)}(t_1, \dots, t_n) = E \exp\{i t_1 \xi_1 + \dots + i t_n \xi_n\}$$

положить $t_1 = \dots = t_n = t$.

Сумма независимых случайных величин, распределенных по закону Пуассона, снова имеет распределение Пуассона.

Рассмотрим две независимые случайные величины $\xi_1 \in \Pi_{\lambda_1}$, $\xi_2 \in \Pi_{\lambda_2}$. Х. ф. их суммы равна

$$\varphi_{\xi_1+\xi_2}(t) = \exp[\lambda_1(e^{it} - 1)] \exp[\lambda_2(e^{it} - 1)] = \exp[(\lambda_1 + \lambda_2)(e^{it} - 1)].$$

Следовательно, $\xi_1 + \xi_2 \in \Pi_{\lambda_1+\lambda_2}$.

Сумма независимых случайных величин, распределенных по закону Коши, также имеет распределение Коши. Действительно, если $\xi_1 \in K_{\alpha_1, \sigma_1}$, $\xi_2 \in K_{\alpha_2, \sigma_2}$, то

$$\varphi_{\xi_1+\xi_2}(t) = e^{i\alpha_1 t - \sigma_1 |t|} e^{i\alpha_2 t - \sigma_2 |t|} = e^{i(\alpha_1+\alpha_2)t - (\sigma_1+\sigma_2)|t|},$$

$$\xi_1 + \xi_2 \in K_{\alpha_1+\alpha_2, \sigma_1+\sigma_2}.$$

Высказанные утверждения тесно связаны с тем, что нормальный закон и закон Пуассона выступают, как мы видели, в качестве предельных законов для сумм независимых слагаемых (тем же свойством обладает и закон Коши, см § 8.9).

Ведь если $S_{2n}/\sqrt{2n}$ сходится по распределению к нормальному закону, то ясно, что S_n/\sqrt{n} и $(S_{2n} - S_n)/\sqrt{n}$ также будут сходиться к нормальному закону, так что сумма двух асимптотически нормальных величин обязана снова быть асимптотически нормальной.

Равномерное распределение, например, этим свойством сохранения вида распределения при суммировании не обладает. Если ξ_1 и ξ_2 распределены равномерно на $[0, 1]$ и независимы, то F_{ξ_1} и $F_{\xi_1+\xi_2}$ оказываются существенно разными функциями (см. пример 3.5).

2. Г-распределение и его свойства. В этом разделе рассмотрим еще один весьма распространенный вид распределений, тесно связанных с нормальным распределением и часто используемых в приложениях. Речь идет о так называемом *гамма-распределении Пирсона* $\Gamma_{\alpha, \lambda}$. Будем писать, что $\xi \in \Gamma_{\alpha, \lambda}$, если ξ имеет плотность

$$f(x; \alpha, \lambda) = \begin{cases} \frac{\alpha^\lambda}{\Gamma(\lambda)} x^{\lambda-1} e^{-\alpha x}, & x \geq 0, \\ 0, & x < 0, \end{cases}$$

зависящую от двух параметров $\alpha > 0$ и $\lambda > 0$, где $\Gamma(\lambda)$ есть гамма-функция

$$\Gamma(\lambda) = \int_0^\infty x^{\lambda-1} e^{-x} dx, \quad \lambda > 0.$$

Из этого равенства следует, что $\int f(x; \alpha, \lambda) dx = 1$ (нужно сделать замену $\alpha x = y$). Если х. ф.

$$\varphi(t) = \varphi(t; \alpha, \lambda) = \frac{\alpha^\lambda}{\Gamma(\lambda)} \int_0^\infty x^{\lambda-1} e^{itz - \alpha x} dx$$

продифференцировать по t и затем произвести интегрирование по частям, то получим

$$\varphi'(t) = \frac{\alpha^\lambda}{\Gamma(\lambda)} \int_0^\infty i x^\lambda e^{itz - \alpha x} dx = \frac{\alpha^\lambda}{\Gamma(\lambda)} \frac{i\lambda}{\alpha - it} \int_0^\infty x^{\lambda-1} e^{itz - \alpha x} dx = \frac{i\lambda}{\alpha - it} \varphi(t);$$

$$(\ln \varphi(t))' = (-\lambda \ln(\alpha - it))', \quad \varphi(t) = c(\alpha - it)^{-\lambda}.$$

Так как $\varphi(0) = 1$, то $c = \alpha^\lambda$, $\varphi(t) = (1 - it/\alpha)^{-\lambda}$.

Из вида х. ф. следует, что подсемейство распределений $\Gamma_{\alpha, \lambda}$ при фиксированном α также обладает известным свойством устойчивости: если $\xi_1 \in \Gamma_{\alpha, \lambda_1}$, $\xi_2 \in \Gamma_{\alpha, \lambda_2}$ независимы, то $\xi_1 + \xi_2 \in \Gamma_{\alpha, \lambda_1 + \lambda_2}$.

К Γ -распределениям относится, в частности, распределение случайной величины

$$\chi_n^2 = \sum_{i=1}^n \xi_i^2,$$

где ξ_i независимы и нормально распределены с параметрами $(0, 1)$. Это так называемое χ^2 -распределение с n степенями свободы, играющее важную роль в математической статистике.

Чтобы найти распределение χ_n^2 , достаточно заметить, что в силу равенства

$$P(\chi_1^2 < x) = P(|\xi_1| < \sqrt{x}) = \frac{2}{\sqrt{2\pi}} \int_0^{\sqrt{x}} e^{-u^2/2} du$$

плотность χ_1^2 равна

$$\frac{1}{\sqrt{2\pi}} e^{-x/2} x^{-1/2} = f\left(x; \frac{1}{2}, \frac{1}{2}\right), \quad \chi_1^2 \in \Gamma_{1/2, 1/2}.$$

Это означает, что х. ф. χ_n^2 равна

$$\varphi^n\left(t; \frac{1}{2}, \frac{1}{2}\right) = (1 - 2it)^{-n/2} = \varphi(t; 1/2, n/2)$$

и соответствует плотности $f(x; 1/2, n/2)$.

К Γ -распределениям относится также *показательное распределение* $\Gamma_\alpha = \Gamma_{\alpha, 1}$ с плотностью

$$f(x; \alpha, 1) = \alpha e^{-\alpha x}, \quad x \geq 0,$$

и характеристической функцией

$$\varphi(t; \alpha, 1) = \left(1 - \frac{it}{\alpha}\right)^{-1}.$$

Предоставляем читателю с помощью х. ф. убедиться, что если $\xi_i \in \Gamma_{\alpha_i}$ независимы, $\alpha_j \neq \alpha_l$ при $j \neq l$, то

$$P\left(\sum_{j=1}^n \xi_j > x\right) = \sum_{j=1}^n e^{-\alpha_j x} \prod_{\substack{l=1 \\ l \neq j}}^n \left(1 - \frac{\alpha_j}{\alpha_l}\right)^{-1}.$$

В различных приложениях, например, в теории массового обслуживания (ср. с § 11.4), важную роль играет также так называемое *распределение Эрланга*. Это распределение с плотностью $f(x; \alpha, \lambda)$ при целом λ . Ясно, что распределение Эрланга есть λ -кратная свертка показательного распределения.

Найдем математическое ожидание и дисперсию случайной величины ξ , имеющей гамма-распределение с параметрами (α, λ) :

$$\begin{aligned} E\xi &= -i\varphi'(0; \alpha, \lambda) = \frac{\lambda}{\alpha}, \\ E\xi^2 &= -\varphi''(0; \alpha, \lambda) = \frac{\lambda(\lambda + 1)}{\alpha^2}, \\ D\xi &= \frac{\lambda(\lambda + 1)}{\alpha^2} - \left(\frac{\lambda}{\alpha}\right)^2 = \frac{\lambda}{\alpha^2}. \end{aligned}$$

Распределения из гамма-семейства и особенно показательное распределение часто используются с достаточными основаниями для аппроксимации распределений в разного рода прикладных задачах. В связи с этим приведем три примера.

Пример 9. Рассмотрим какой-нибудь сложный механизм. Выход из строя хотя бы одной из n деталей или узлов, составляющих этот механизм, означает выход из строя всего механизма. Распределение срока службы каждого из узлов обычно хорошо описывается показательным законом. (Причины этого явления можно понять с помощью теоремы Пуассона о редких событиях; см. также пример 2.5 и главу 18.)

Таким образом, если сроки службы узлов ξ_j независимы и для узла с номером j выполняется

$$P(\xi_j > x) = e^{-\alpha_j x},$$

то срок службы всего механизма будет равен $\eta_n = \min(\xi_1, \dots, \xi_n)$, и мы получим

$$P(\eta_n > x) = P\left(\bigcap_{j=1}^n \{\xi_j > x\}\right) = \prod_{j=1}^n P(\xi_j > x) = e^{-x \sum_{i=1}^n \alpha_i}.$$

Это означает, что η_n снова будет иметь показательное распределение, и так как

$$E\xi_j = \frac{1}{\alpha_j},$$

то среднее время службы всего механизма будет равно

$$E\eta_n = \left(\sum_{i=1}^n \frac{1}{E\xi_i}\right)^{-1}.$$

Пример 10. Обратимся теперь к распределению случайной величины $\zeta_n = \max(\xi_1, \dots, \xi_n)$, где ξ_i независимы и имеют Γ -распределение с параметрами (α, λ) . Мы можем рассматривать, например, некоторую систему обслуживания, имеющую n каналов. (Это может быть, скажем, электронно-вычислительная машина, решающая некоторую задачу методом перебора и состоящая из большого числа малых машин, каждая из которых просматривает свой вариант.) Канал с номером i занят случайное время ξ_i . Через какое время освободится вся система? Это время, очевидно, будет распределено так же, как случайная величина ζ_n .

Так как ξ_i независимы, то

$$P(\zeta_n < x) = P\left(\bigcap_{j=1}^n \{\xi_j < x\}\right) = [P(\xi_1 < x)]^n. \quad (15)$$

Если n велико, то для приближенных вычислений представляет интерес отыскание предельного при $n \rightarrow \infty$ распределения ζ_n (если $n \rightarrow \infty$, то $P(\zeta_n < x) \rightarrow 0$ при каждом фиксированном x).

Считая для простоты $\alpha = 1$ (общий случай сводится к этому изменением масштаба), по правилу Лопиталя убеждаемся, что при $x \rightarrow \infty$

$$P(\xi_j > x) = \int_x^{\infty} \frac{1}{\Gamma(\lambda)} y^{\lambda-1} e^{-y} dy \sim \frac{x^{\lambda-1}}{\Gamma(\lambda)} e^{-x}.$$

Полагая при $n \rightarrow \infty$ $x = x(n) = \ln \left[\frac{n(\ln n)^{\lambda-1}}{\Gamma(\lambda)} \right] + u$, $u = \text{const}$, получим

$$P(\xi_j > x) \sim \frac{(\ln n)^{\lambda-1}}{\Gamma(\lambda)} \frac{\Gamma(\lambda)}{n(\ln n)^{\lambda-1}} e^{-u} = \frac{e^{-u}}{n}.$$

Поэтому при таких x и $n \rightarrow \infty$ в силу (15)

$$P(\zeta_n < x) = \left(1 - \frac{e^{-u}}{n} (1 + o(1)) \right)^n \rightarrow e^{-e^{-u}}.$$

Таким образом, мы получили, что существует

$$\lim_{n \rightarrow \infty} P \left(\zeta_n - \ln \left[\frac{n(\ln n)^{\lambda-1}}{\Gamma(\lambda)} \right] < u \right) = e^{-e^{-u}},$$

или, что то же,

$$\zeta_n - \ln \left[\frac{n(\ln n)^{\lambda-1}}{\Gamma(\lambda)} \right] \Rightarrow F_0, \quad F_0(u) = e^{-e^{-u}}.$$

Другими словами, величину ζ_n при больших n можно представить в виде

$$\zeta_n \approx \ln \left[\frac{n(\ln n)^{\lambda-1}}{\Gamma(\lambda)} \right] + \zeta^0,$$

где $\zeta^0 \in F_0$.

Пример 11. Пусть ξ_1 и ξ_2 независимы, $\xi_1 \in \Gamma_{\alpha, \lambda_1}$, $\xi_2 \in \Gamma_{\alpha, \lambda_2}$. Чему равно распределение $\xi_1/(\xi_1 + \xi_2)$? Воспользуемся теоремой 4.10. Так как совместная плотность $f(x, y)$ для ξ_1 и $\eta = \xi_1 + \xi_2$ равна

$$f(x, y) = f(x; \alpha, \lambda_1) f(y - x; \alpha, \lambda_2),$$

а плотность для η есть

$$q(y) = f(y; \alpha, \lambda_1 + \lambda_2),$$

то условная плотность $f(x|y)$ для ξ_1 при условии $\eta = y$ равна

$$f(x|y) = \frac{f(x, y)}{q(y)} = \frac{\Gamma(\lambda_1 + \lambda_2)}{\Gamma(\lambda_1)\Gamma(\lambda_2)} \frac{x^{\lambda_1-1}(y-x)^{\lambda_2-1}}{y^{\lambda_1+\lambda_2-1}}.$$

В силу формул § 3.2 условная плотность для $\xi_1/y = \xi_1/(\xi_1 + \xi_2)$ (при том же условии $\xi_1 + \xi_2 = y$) равна

$$yf(yx|y) = \frac{\Gamma(\lambda_1 + \lambda_2)}{\Gamma(\lambda_1)\Gamma(\lambda_2)} x^{\lambda_1-1}(1-x)^{\lambda_2-1}, \quad x \in [0, 1].$$

Это распределение не зависит от y (и от α). Поэтому такой же будет и безусловная плотность $\xi_1/(\xi_1 + \xi_2)$. Мы получили так называемое **бета-распределение** $\mathbf{B}_{\lambda_1, \lambda_2}$ с параметрами λ_1, λ_2 , определенное на отрезке $[0, 1]$. В частности, при $\lambda_1 = \lambda_2 = 1$ $\mathbf{B}_{1,1} = \mathbf{U}_{0,1}$.

§ 8. Производящие функции.

Применение к изучению ветвящегося процесса.

Задача о вырождении

1. Производящие функции. Мы уже знаем, что если случайная величина ξ целочисленна, т. е. $P\left(\bigcup_k \{\xi = k\}\right) = 1$, то х. ф. $\varphi_\xi(t)$ будет на самом деле функцией от $z = e^{it}$, и наряду с х. ф. распределение ξ можно характеризовать производящей функцией

$$\psi_\xi(z) = \mathbf{E} z^\xi = \sum z^k P(\xi = k).$$

Формула обращения здесь может быть записана в виде

$$P(\xi = k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-itk} \varphi_\xi(t) dt = \frac{1}{2\pi i} \int_{|z|=1} z^{-k-1} \psi_\xi(z) dz. \quad (16)$$

Читателю полезно отметить, что (16) представляет собой не что иное, как формулу для коэффициентов ряда Фурье (ведь $e^{itk} = \cos tk + i \sin tk$).

Если ξ и η — независимые целочисленные случайные величины, то распределение $\xi + \eta$ будет определяться сверткой последовательностей $P(\xi = k)$ и $P(\eta = k)$:

$$P(\xi + \eta = n) = \sum_{k=-\infty}^{\infty} P(\xi = k) P(\eta = n - k)$$

(формула полной вероятности). Свертке этих последовательностей отвечает произведение производящих функций

$$\psi_{\xi+\eta}(z) = \psi_\xi(z) \psi_\eta(z).$$

Из примеров, рассмотренных в § 1, ясно, что производящие функции случайных величин, распределенных соответственно по законам Бернулли и Пуассона, равны

$$\psi_\xi(z) = 1 + p(z - 1), \quad \psi_\eta(z) = \exp\{\mu(z - 1)\}.$$

Из определения производящей функции видно, что для неотрицательной случайной величины $\xi \geq 0$ функция $\psi_\xi(z)$ определена при $|z| \leq 1$ и аналитична в области $|z| < 1$.

2. Простейшие ветвящиеся процессы. Обратимся теперь к последовательностям случайных величин, которыми описываются так называемые *ветвящиеся процессы*. С простейшим примером таких процессов мы уже сталкивались при описании схемы цепной реакции в примере 4.10. Рассмотрим более общую схему ветвящегося процесса. Представим себе частицы, которые могут производить другие частицы того же вида; это могут быть нейтроны в цепных реакциях, бактерии, размножающиеся по определенным законам, и др. Пусть в начальный момент

времени имеется одна частица («нулевое поколение»), которая в результате акта «деления» с вероятностью f_k , $k = 0, 1, 2, \dots$, переходит в k частиц того же типа, $\sum_{k=0}^{\infty} f_k = 1$. Полученные частицы образуют «первое поколение». Каждая из частиц этого поколения ведет себя точно так же, как исходная частица, независимо от предыстории и судьбы других частиц. В результате получаем «второе поколение» и т. д. Обозначим ζ_n число частиц в n -м поколении. Для описания последовательности ζ_n введем в рассмотрение, как и в примере 4.10, независимые между собой последовательности независимых одинаково распределенных случайных величин $\{\xi_j^{(1)}\}_{j=1}^{\infty}, \{\xi_j^{(2)}\}_{j=1}^{\infty}, \dots$, где $\xi_j^{(n)}$ имеют распределение

$$P(\xi_j^{(n)} = k) = f_k, \quad k = 0, 1, \dots$$

Тогда последовательность ζ_n можно представить в виде

$$\begin{aligned} \zeta_0 &= 1, \\ \zeta_1 &= \xi_1^{(1)}, \\ \zeta_2 &= \xi_1^{(2)} + \dots + \xi_{\zeta_1}^{(2)}, \\ &\dots \dots \dots \\ \zeta_n &= \xi_1^{(n)} + \dots + \xi_{\zeta_{n-1}}^{(n)}. \end{aligned}$$

Это есть суммы случайного числа случайных величин. Так как $\xi_1^{(n)}, \xi_2^{(n)}, \dots$ от ζ_{n-1} не зависят, то по формуле полной вероятности для производящей функции $f_{(n)}(z) = Ez^{\zeta_n}$ получим

$$f_{(n)}(z) = \sum_{k=0}^{\infty} P(\zeta_{n-1} = k) Ez^{\xi_1^{(n)} + \dots + \xi_k^{(n)}} = \sum_{k=0}^{\infty} P(\zeta_{n-1} = k) f^k(z) = f_{(n-1)}(f(z)), \quad (17)$$

где $f(z) = f_{(1)}(z) = Ez^{\xi_1^{(1)}} = \sum_{k=0}^{\infty} f_k z^k$. Пусть $f_n(z)$ означает n -ю итерацию функции $f(z)$, т. е. $f_1(z) = f(z)$, $f_2(z) = f(f(z))$, $f_3(z) = f(f_2(z))$ и т. д. Тогда с помощью индукции из соотношений (17) заключаем, что производящая функция ζ_n равна n -й итерации $f(z)$:

$$Ez^{\zeta_n} = f_n(z).$$

Отсюда с помощью дифференцирования в точке $z = 1$ легко получить рекуррентные соотношения для моментов ζ_n .

Как найти вероятность вырождения процесса? Под вырождением мы будем понимать событие, состоящее в том, что все ζ_n , начиная с некоторого n , равны нулю. (Если $\zeta_n = 0$, то, очевидно, $\zeta_{n+1} = \zeta_{n+2} = \dots = 0$, так как $P(\zeta_{n+1} = 0 | \zeta_n = 0) = 1$.)

Обозначим $A_k = \{\zeta_k = 0\}$. Тогда вырождение представляет собой событие $\bigcup_{k=1}^{\infty} A_k$.

Так как $A_n \subset A_{n+1}$, то вероятность вырождения q равна $q = \lim_{n \rightarrow \infty} P(A_n)$.

Теорема 5. Вероятность вырождения q равна наименьшему неотрицательному корню уравнения $q = f(q)$.

Доказательство. Имеем $P(A_n) = f_n(0) \leq 1$, и эта последовательность не убывает. Переходя в равенстве

$$f_{n+1}(0) = f(f_n(0)) \quad (18)$$

к пределу при $n \rightarrow \infty$, получим $q = f(q)$, $q \leq 1$. Это уравнение для вероятности вырождения. Исследуем поведение его решения. $f(z)$ есть выпуклая ($f''(z) > 0$) неубывающая функция в области $z \geq 0$, $f'(1) = m$ есть средняя величина потомства одной частицы. Предположим сначала, что $P(\xi_1^{(1)} = 1) < 1$. Тогда если $m \leq 1$, то $f(z) > z$ при $z < 1$ и, следовательно, $q = 1$. Если $m > 1$, то в силу выпуклости f уравнение $q = f(q)$ будет иметь на отрезке $[0, 1]$ ровно два решения: $q_1 < 1$ и $q_2 = 1$ (см. рис. 8). Допустим, что $q = q_2 = 1$. Тогда последовательность $\delta_n = 1 - f_n(0)$ монотонно сходится к нулю и $f(1 - \delta_n) < 1 - \delta_n$ при достаточно больших n . Следовательно, для таких n

$$\delta_{n+1} = 1 - f(1 - \delta_n) > \delta_n,$$

что противоречит монотонности убывания δ_n . Это означает, что $q = q_1 < 1$. Наконец, в случае $P(\xi_1^{(1)} = 1) = f_1 = 1$ имеем $f(z) \equiv z$ и, очевидно, $q = 0$. ■

Рассмотрим теперь более внимательно случай $m = 1$, который называют *критическим*. Мы знаем, что в этом случае при $f_1 < 1$ вероятность вырождения q равна 1. Пусть $q_n = P(A_n) = f_n(0)$ есть вероятность вырождения к моменту n . Как быстро q_n сходится к 1? В силу (18) $q_{n+1} = f(q_n)$. Отсюда следует, что вероятность продолжения (не вырождения) процесса $p_n = 1 - q_n$ (к моменту n) удовлетворяет соотношению

$$p_{n+1} = g(p_n), \quad g(x) = 1 - f(1 - x).$$

Ясно также, что $\gamma_n = p_n - p_{n+1}$ есть вероятность того, что вырождение процесса совершится на шаге n .

Теорема 6. Если $m = f'(1) = 1$, $0 < b \equiv f''(1) < \infty$, то $\gamma_n \sim 2/(bn^2)$, $p_n \sim 2/(bn)$.

Доказательство. Если второй момент числа потомков одной частицы конечен ($b < \infty$), то существует $g''(0) = -b$ и, следовательно ($g(0) = 0$, $g'(0) = f'(0) = 1$),

$$g(x) = x - \frac{b}{2}x^2 + o(x^2).$$

Полагая здесь $x = p_n \rightarrow 0$, для последовательности $a_n = 1/p_n$ находим

$$a_{n+1} - a_n = \frac{p_n - p_{n+1}}{p_n p_{n+1}} = \frac{bp_n^2(1 + o(1))}{2p_n^2(1 - bp_n/2 + o(p_n))} \rightarrow \frac{b}{2},$$

$$a_n = a_1 + \sum_{k=1}^{n-1} (a_{k+1} - a_k) \sim \frac{bn}{2}, \quad p_n \sim \frac{2}{bn}.$$

Рассмотрим теперь вопрос о распределении числа частиц ζ_n при условии $\zeta_n > 0$.

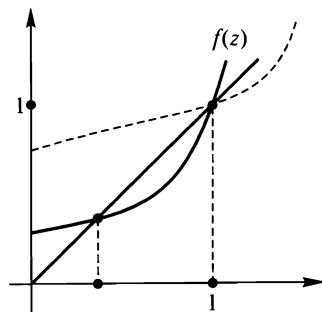


Рис. 8.

Теорема 7. Пусть выполнены условия теоремы 6. Тогда условное распределение $p_n \zeta_n$ (или $2\zeta_n/(bn)$) при условии $\zeta_n > 0$ сходится при $n \rightarrow \infty$ к показательному:

$$P(p_n \zeta_n > x | \zeta_n > 0) \rightarrow e^{-x}.$$

Сформулированное утверждение означает, в частности, что при $n \rightarrow \infty$ и условию $\zeta_n > 0$ число частиц ζ_n имеет величину порядка n .

Доказательство. Рассмотрим преобразование Лапласа (см. свойство 6 из § 1) над условным распределением $p_n \zeta_n$ (при условии $\zeta_n > 0$), равное

$$E(e^{-sp_n \zeta_n} | \zeta_n > 0) = \frac{1}{p_n} \sum_{k=1}^{\infty} e^{-skp_n} P(\zeta_n = k). \quad (19)$$

Так как $p_n \rightarrow 0$, то при любом фиксированном $s > 0$

$$e^{-sp_n} - 1 \sim -sp_n \sim -\frac{2s}{bn}.$$

Ясно, что всегда можно выбрать $N \sim n/s$, $s_n \sim s$, $s_n \downarrow s$ так, чтобы $e^{-s_n p_n} - 1 = -p_N$. Поэтому $e^{-sp_n k} = (1 - p_N)^k$, и правая часть в (19) при $s = s_n$ может быть записана в виде

$$\begin{aligned} \frac{1}{p_n} \sum_{k=1}^{\infty} P(\zeta_n = k)(1 - p_N)^k &= \frac{1}{p_n} P(\zeta_n > 0, \zeta_{n+N} = 0) = \\ &= \frac{p_n - p_{n+N}}{p_n} = 1 - \frac{p_{n+N}}{p_n} \sim 1 - \frac{n}{n+N} = \frac{N}{n+N} \rightarrow \frac{1}{1+s}. \end{aligned}$$

Заметим теперь, что

$$E(e^{-sp_n \zeta_n} | \zeta_n > 0) - E(e^{-s_n p_n \zeta_n} | \zeta_n > 0) = E[e^{-sp_n \zeta_n} (1 - e^{-(s_n - s)p_n \zeta_n}) | \zeta_n > 0].$$

Так как $e^{-\alpha} \leq 1$, $1 - e^{-\alpha} \leq \alpha$ при $\alpha \geq 0$, $E\zeta_n = 1$, $E(\zeta_n | \zeta_n > 0) = 1/p_n$, то оцениваемая положительная разность ($s_n > s$) не превосходит

$$(s_n - s)p_n E(\zeta_n | \zeta_n > 0) = s_n - s \rightarrow 0.$$

Следовательно, преобразование Лапласа (19) сходится при $n \rightarrow \infty$ к $1/(1+s)$. Так как $1/(1+s)$ есть преобразование Лапласа показательного распределения

$$\int_0^{\infty} e^{-sx-x} dx = \frac{1}{1+s},$$

то в силу теоремы непрерывности (см. замечание к теореме 2 из § 3) рассматриваемое условное распределение сходится к показательному⁴⁾. ■

Утверждения, касающиеся поведения ζ_n при $n \rightarrow \infty$ для ветвящихся процессов в случае $\mu > 1$ (так называемых *надкритических* процессов) будут получены в § 14.4 (пример 8) в качестве следствий теорем о сходимости мартингалов.

⁴⁾ Приведенные простые доказательства теорем 6, 7 сообщил нам К. А. Боровков.

Глава 8

Последовательности независимых случайных величин. Предельные теоремы

§ 1. Закон больших чисел

Теорема 1 (закон больших чисел в форме Хинчина). Пусть $\{\xi_n\}_{n=1}^{\infty}$ — последовательность независимых одинаково распределенных случайных величин, имеющих математическое ожидание $E\xi_n = a$, и пусть $S_n = \xi_1 + \dots + \xi_n$. Тогда $S_n/n \xrightarrow{p} a$ при $n \rightarrow \infty$.

Отметим, что здесь как условие независимости ξ_k , так и само утверждение теоремы предполагают, что все случайные величины ξ_k заданы на одном вероятностном пространстве.

По своему физическому существу сформулированный закон больших чисел представляет собой простейшую эргодическую теорему, означающую, грубо говоря, что для случайных величин их средние значения «по времени» и «по пространству» совпадают. Еще в большей мере это замечание можно будет отнести к усиленному закону больших чисел, в силу которого $S_n/n \rightarrow p$ с вероятностью 1.

При более жестких предположениях (существование дисперсии) теорема 1 была получена нами в § 7 главы 4 в качестве следствия неравенства Чебышева.

Доказательство теоремы 1. Надо доказать, что при любом $\varepsilon > 0$

$$P\left(\left|\frac{S_n}{n} - a\right| > \varepsilon\right) \rightarrow 0$$

при $n \rightarrow \infty$. Выписанное соотношение эквивалентно слабой сходимости распределений $S_n/n \Leftrightarrow I_a$. Поэтому в силу теоремы непрерывности и примера 7.1 достаточно доказать, что для каждого фиксированного t

$$\varphi_{S_n/n}(t) \rightarrow e^{iat}.$$

Х. ф. $\varphi(t)$ случайной величины ξ_k в некоторой окрестности нуля обладает свойством $|\varphi(t) - 1| < 1/2$. Поэтому для таких t можно определить функцию $l(t) = \ln \varphi(t)$ (берется главное значение логарифма). Так как существует математическое ожидание ξ_n , то существует

$$l'(0) = \frac{\varphi'(0)}{\varphi(0)} = ia.$$

При каждом фиксированном t и достаточно больших n значение $l(t/n)$ определено и

$$\varphi_{S_n/n}(t) = \varphi^n\left(\frac{t}{n}\right) = e^{l(t/n)n}.$$

Так как $l(0) = 0$, то при $n \rightarrow \infty$

$$e^{l(t/n)n} = \exp \left\{ t \frac{l(t/n) - l(0)}{t/n} \right\} \rightarrow e^{l'(0)t} = e^{iat}.$$

§ 2. Центральная предельная теорема для одинаково распределенных случайных величин

Пусть по-прежнему $\{\xi_n\}$ — последовательность независимых одинаково распределенных случайных величин. Но теперь, кроме математического ожидания $E\xi_n = a$, будем предполагать существование дисперсии $D\xi_n = \sigma^2$. Как и раньше, $S_n = \xi_1 + \dots + \xi_n$, а $\Phi(x)$ — функция распределения нормального закона с параметрами $(0, 1)$. Введем еще последовательность случайных величин

$$\zeta_n = \frac{S_n - an}{\sigma\sqrt{n}}.$$

Теорема 2. Если $0 < \sigma^2 < \infty$, то при $n \rightarrow \infty$ $P(\zeta_n < x) \rightarrow \Phi(x)$ равномерно относительно x ($-\infty < x < \infty$).

В этом случае последовательность $\{\zeta_n\}$ называется *асимптотически нормальной*.

Из того, что $\zeta_n \Rightarrow \zeta \in \Phi_{0,1}$, $\zeta_n^2 \geq 0$, $E\zeta_n^2 = E\zeta^2 = 1$ и из леммы 6.1А вытекает равномерная интегрируемость ζ_n^2 . Поэтому наряду со слабой сходимостью $\zeta_n \Rightarrow \zeta$, $\zeta \in \Phi_{0,1}$ ($Ef(\zeta_n) \rightarrow Ef(\zeta)$ для любой непрерывной ограниченной f) имеет место также сходимость $Ef(\zeta_n) \rightarrow Ef(\zeta)$ для любой непрерывной f такой, что $|f(x)| < c(1 + x^2)$ (см. теорему 6.5А).

Доказательство теоремы 2. Равномерная сходимость является следствием слабой сходимости и непрерывности $\Phi(x)$. Далее, без ограничения общности можно считать $a = 0$, так как иначе можно было бы рассмотреть последовательность $\{\xi'_n = \xi_n - a\}_{n=1}^\infty$, при этом последовательность $\{\zeta_n\}$ не изменилась бы. Стало быть, для доказательства требуемой сходимости достаточно показать, что $\varphi_{\zeta_n}(t) \rightarrow e^{-t^2/2}$, когда $a = 0$. Имеем

$$\varphi_{\zeta_n}(t) = \varphi^n \left(\frac{t}{\sigma\sqrt{n}} \right), \quad \text{где} \quad \varphi(t) = \varphi_{\xi_1}(t).$$

Так как существует $E\xi_1^2$, то существует $\varphi''(t)$, и справедливо разложение

$$\varphi(t) = \varphi(0) + t\varphi'(0) + \frac{t^2}{2}\varphi''(0) + o(t^2) = 1 - \frac{t^2\sigma^2}{2} + o(t^2).$$

Следовательно, при $n \rightarrow \infty$

$$\begin{aligned} \ln \varphi_{\zeta_n}(t) &= n \ln \left[1 - \frac{\sigma^2}{2} \left(\frac{t}{\sigma\sqrt{n}} \right)^2 + o \left(\frac{t^2}{n} \right) \right] = \\ &= n \left(-\frac{t^2}{2n} + o \left(\frac{t^2}{n} \right) \right) = -\frac{t^2}{2} + o(1) \rightarrow -\frac{t^2}{2}. \end{aligned}$$

§ 3. Закон больших чисел для произвольных независимых случайных величин

Перейдем теперь к выяснению условий, при которых справедливы закон больших чисел и центральная предельная теорема в случае, когда ξ_k независимы, но не обязательно одинаково распределены. Мы не добавим трудностей, если сразу будем рассматривать более общую задачу, когда даны произвольные серии $\xi_{1,n}, \dots, \xi_{n,n}$, $n = 1, 2, \dots$, независимых случайных величин, где распределения $\xi_{k,n}$ могут зависеть от n .

Обозначим $\zeta_n = \sum_{k=1}^n \xi_{k,n}$. С точки зрения последующих результатов можно считать, не ограничивая общности, что

$$E\xi_{k,n} = 0. \quad (1)$$

Предположим, что выполнено условие:

$$D_1 = \sum_{k=1}^n E \min(|\xi_{k,n}|, |\xi_{k,n}|^2) \rightarrow 0 \quad (D_1)$$

при $n \rightarrow \infty$.

Теорема 3 (закон больших чисел) При выполнении условий (1), (D_1)

$$\zeta_n \xrightarrow{p} 0, \quad \text{или, что то же,} \quad \zeta_n \xrightarrow{p} 0.$$

Замечание 1. Применительно к одинаково распределенным случайным величинам ξ_k и их суммам S_n теорема 3 является более сильным в известном смысле утверждением по сравнению с теоремой 1. Тот факт, что теорема 1 вытекает из теоремы 3, будет установлен в теореме 3А. Здесь же мы проиллюстрируем высказанное утверждение на примере случайных величин ξ_k , $E\xi_k = 0$ с $E|\xi_k|^s \leq c < \infty$, $1 < s \leq 2$. Пусть $b(n) = o(n)$ — любая последовательность, растущая быстрее, чем $n^{1/s}$, $\xi_{k,n} = \xi_k/b(n)$. Тогда $D_1 \leq nc/b(n)^s \rightarrow 0$ и, следовательно, $S_n/b(n) \xrightarrow{p} 0$.

Рассмотрим внимательнее условие (D_1) . Очевидно, что его можно записать в также в форме

$$D_1 = \sum_{k=1}^n E(|\xi_{k,n}|; |\xi_{k,n}| > 1) + \sum_{k=1}^n E(|\xi_{k,n}|^2; |\xi_{k,n}| \leq 1) \rightarrow 0.$$

Если

$$M_1 \equiv \sum_{k=1}^n E|\xi_{k,n}| \leq c < \infty \quad (2)$$

при $n \rightarrow \infty$, то для выполнения (D_1) необходимо и достаточно выполнения условия

$$M_1(\tau) \equiv \sum_{k=1}^n E(|\xi_{k,n}|; |\xi_{k,n}| > \tau) \rightarrow 0 \quad (M_1)$$

при любом $\tau > 0$ и $n \rightarrow \infty$, которое можно назвать условием типа Линдберга (условие Линдберга (M_2) будет введено в § 4). Действительно, $(\tau \leq 1, g_1(x) = \min(|x|, |x|^2))$,

$$\begin{aligned}
D_1 &= \sum_{k=1}^n E g_1(\xi_{k,n}) \leq \sum_{k=1}^n E(|\xi_{k,n}|; |\xi_{k,n}| > \tau) + \sum_{k=1}^n E(|\xi_{k,n}|^2; |\xi_{k,n}| \leq \tau) \leq \\
&\leq M_1(\tau) + \tau \sum_{k=1}^n E(|\xi_{k,n}|; |\xi_{k,n}| \leq \tau) \leq M_1(\tau) + \tau M_1(0).
\end{aligned} \quad (3)$$

Так как $M_1(0) = M_1 \leq c$, а τ может быть произвольно малым, то $D_1 \rightarrow 0$ при $n \rightarrow \infty$.

Обратно (здесь условие (2) не требуется), при $\tau \leq 1$

$$\begin{aligned}
M_1(\tau) &\leq \sum_{k=1}^n E(|\xi_{k,n}|; |\xi_{k,n}| > 1) + \\
&+ \frac{1}{\tau} \sum_{k=1}^n E(|\xi_{k,n}|^2; \tau < |\xi_{k,n}| \leq 1) \leq \frac{1}{\tau} D_1 \rightarrow 0
\end{aligned} \quad (4)$$

при $n \rightarrow \infty$ и при любом $\tau > 0$.

Каждое из условий (D_1) , (M_1) влечет за собой равномерную малость $E|\xi_{k,n}|$:

$$\max_{1 \leq k \leq n} E|\xi_{k,n}| \rightarrow 0 \quad \text{при } n \rightarrow \infty. \quad (5)$$

Действительно, (M_1) означает, что существует достаточно медленно убывающая последовательность $\tau_n \rightarrow 0$ такая, что $M_1(\tau_n) \rightarrow 0$. Поэтому

$$\max_{k \leq n} E|\xi_{k,n}| \leq \max_{k \leq n} [\tau + (E|\xi_{k,n}|; |\xi_{k,n}| > \tau)] \leq \tau_n + M_1(\tau_n) \rightarrow 0. \quad (6)$$

Из (5) следует, в частности, свойство пренебрежимой малости слагаемых $\xi_{k,n}$.

Будем говорить, что $\xi_{k,n}$ *пренебрежимо малы* или, что то же, *обладают свойством (S)*, если при любом $\varepsilon > 0$

$$\max_{k \leq n} P(|\xi_{k,n}| > \varepsilon) \rightarrow 0. \quad (S)$$

Свойство (S) можно было бы назвать *равномерной сходимостью $\xi_{k,n}$ по вероятности к нулю*. Свойство (S) сразу следует из (5) и неравенства Чебышева. Оно следует также из более сильных соотношений, вытекающих из (M_1) :

$$\begin{aligned}
P(\max_{k \leq n} |\xi_{k,n}| > \varepsilon) &= P(\bigcup \{|\xi_{k,n}| > \varepsilon\}) \leq \\
&\leq \sum_{k \leq n} P(|\xi_{k,n}| > \varepsilon) \leq \varepsilon^{-1} \sum_{k \leq n} E(|\xi_{k,n}|; |\xi_{k,n}| > \varepsilon) \rightarrow 0.
\end{aligned} \quad (S_1)$$

Обратимся теперь к доказательству закона больших чисел. Мы приведем две версии доказательства. Первая будет иллюстрировать классический метод характеристических функций. Вторая версия основана на элементарных неравенствах и приводит к более сильному утверждению о сходимости в среднем¹⁾.

Рассмотрим сначала первую версию.

¹⁾ Вторая версия предложена А. И. Саханенко.

Доказательство теоремы 3²⁾. Обозначим

$$\varphi_{k,n}(t) = E e^{it\xi_{k,n}}, \quad \Delta_k(t) = \varphi_{k,n}(t) - 1.$$

Надо доказать, что при каждом t и при $n \rightarrow \infty$

$$\varphi_{\zeta_n}(t) = E e^{it\zeta_n} = \prod_{k=1}^n \varphi_{k,n}(t) \rightarrow 1.$$

В силу леммы 7.3

$$\begin{aligned} |\varphi_{\zeta_n}(t) - 1| &= \left| \prod_{k=1}^k \varphi_{k,n}(t) - \prod_{k=1}^n 1 \right| \leq \sum_{k=1}^n |\Delta_k(t)| = \\ &= \sum_{k=1}^n |E(e^{it\xi_{k,n}} - 1)| = \sum_{k=1}^n |E(e^{it\xi_{k,n}} - 1 - it\xi_{k,n})|. \end{aligned}$$

В силу леммы 7.2 ($g_1(x) = \min(|x|, x^2)$)

$$|e^{itx} - 1 - itx| \leq \min\left(2|x|, \frac{t^2 x^2}{2}\right) \leq 2g_1(tx) \leq 2h(t)g_1(x),$$

где $h(t) = \max(|t|, |t|^2)$. Поэтому

$$|\varphi_{\zeta_n}(t) - 1| \leq 2h(t) \sum_{k=1}^n E g_1(\xi_{k,n}) = 2h(t) D_1 \rightarrow 0.$$

Последнее неравенство показывает, что $|\varphi_{\zeta_n}(t) - 1|$ может быть оценено непосредственно в терминах D_1 . Оказывается, что в терминах D_1 можно оценить и $E|\zeta_n|$. Мы приведем теперь более сильную версию закона больших чисел.

Теорема 3А. При выполнении условий (1), (D_1) справедливо $E|\zeta_n| \rightarrow 0$ (или, что то же, $\zeta_n \xrightarrow{(1)} 0$).

Утверждение теоремы 3А означает, очевидно, равномерную интегрируемость ζ_n ; из него следует теорема 3, так как

$$P(|\zeta_n| > \varepsilon) \leq \frac{E|\zeta_n|}{\varepsilon} \rightarrow 0 \quad \text{при } n \rightarrow \infty.$$

Доказательство теоремы 3А. Положим

$$\xi'_{k,n} = \begin{cases} \xi_{k,n}, & \text{если } |\xi_{k,n}| \leq 1, \\ 0 & \text{в противном случае,} \end{cases}$$

$\xi''_{k,n} = \xi_{k,n} - \xi'_{k,n}$. Тогда $\xi_{k,n} = \xi'_{k,n} + \xi''_{k,n}$, $\zeta_n = \zeta'_n + \zeta''_n$ при очевидном соглашении относительно обозначений ζ'_n , ζ''_n . В силу неравенства Коши—Буняковского

$$E|\zeta_n| \leq E|\zeta'_n - E\zeta'_n| + E|\zeta''_n - E\zeta''_n| \leq \sqrt{E(\zeta'_n - E\zeta'_n)^2} + E|\zeta''_n| + |E\zeta''_n| \leq$$

²⁾ Существует еще одно, «прямое» доказательство теоремы 3, использующее не х.ф., а так называемые срезки случайных величин и оценки их дисперсий. Однако здесь и в связи с дальнейшим нам удобнее использовать аппарат х.ф.

$$\begin{aligned} &\leq \sqrt{\sum D\xi'_{k,n}} + 2 \sum E|\xi''_{k,n}| \leq \sqrt{\sum E(\xi'_{k,n})^2} + 2 \sum E|\xi''_{k,n}| = \\ &= \left[\sum E(\xi_{k,n}^2; |\xi_{k,n}| \leq 1) \right]^{1/2} + 2 \sum E(|\xi_{k,n}|; |\xi_{k,n}| > 1) \leq \sqrt{D_1} + 2D_1 \rightarrow 0, \end{aligned}$$

если $D_1 \rightarrow 0$. Теорема доказана. ■

Замечание 2. Из доказательства теоремы 3А видно, что все ее рассуждения сохраняются, если независимость $\xi_{k,n}$ заменить более слабым условием некоррелированности $\xi'_{k,n}$. Они сохраняются и в том случае, когда $\xi'_{k,n}$ лишь слабо коррелированы, настолько, что

$$E(\zeta'_n - E\zeta'_n)^2 \leq c \sum D(\xi'_{k,n}), \quad c < \infty.$$

Если $\{\xi_k\}$ есть заданная фиксированная (не зависящая от n) последовательность независимых случайных величин, $S_n = \sum_{k=1}^n \xi_k$, $E\xi_k = a_k$, то объектом изучения в законе больших чисел будут последовательности

$$\xi_{k,n} = \frac{\xi_k - a_k}{b(n)}, \quad \zeta_n = \sum \xi_{k,n} = \frac{S_n - \sum_{k=1}^n a_k}{b(n)}, \quad (7)$$

где $\xi_{k,n}$ удовлетворяют (1), $b(n)$ — некоторая неограниченно возрастающая последовательность. В качестве $b(n)$ в ряде случаев естественно брать $b(n) = \sum_{k=1}^n E|\xi_k|$, если эта сумма неограниченно возрастает. Положим, не ограничивая общности, $a_k = 0$. Из теоремы 3А вытекает

Следствие 1. Если выполнено условие

$$D_1 = \frac{1}{b(n)} \sum E \min \left(|\xi_k|, \frac{\xi_k^2}{b(n)} \right) \rightarrow 0$$

или условия

$$M_1(\tau) = \frac{1}{b(n)} \sum E(|\xi_k|; |\xi_k| > \tau b(n)) \rightarrow 0, \quad b(n) = \sum_{k=1}^n E|\xi_k| \rightarrow \infty,$$

при любом $\tau > 0$, то $\zeta_n \xrightarrow{(1)} 0$.

Следствие 2. При выполнении условий теоремы 1 справедливо $S_n/n \xrightarrow{(1)} a$.

Доказательство. Если ξ_k одинаково распределены, $E|\xi_k| < \infty$, то, очевидно, выполнено условие $M_1(\tau) \rightarrow 0$ при $b(n) = n$, так как

$$E(|\xi_k|; |\xi_k| > \tau n) \rightarrow 0 \quad \text{при } n \rightarrow \infty.$$

Следствие 2 доказано. ■

Вернемся к произвольным ξ_k . Для выполнения (D_1) достаточно выполнения следующего условия: при некотором s , $2 \geq s > 1$,

$$\sum_{k=1}^n E|\xi_{k,n}|^s \rightarrow 0. \quad (L_s)$$

Это утверждение очевидно, так как $g_1(x) \leq |x|^s$ при $2 \geq s > 1$. Условия (L_s) можно было бы назвать условием типа Ляпунова (ср. с условием Ляпунова (L_s) в § 4).

Для доказательства теоремы 3А мы использовали так называемые «срезки» $\xi'_{k,n}$ случайных величин $\xi_{k,n}$. В заключение этого раздела мы приведем еще одно утверждение закона больших чисел, в котором условия выражены в терминах срезок.

Обозначим через $\xi^{(N)}$ срезку случайной величины ξ на уровне N :

$$\xi^{(N)} = \max[-N, \min(N, \xi)].$$

Теорема 3В. *Предположим, что последовательность $\{\xi_k\}$ случайных величин в (7) удовлетворяет следующему условию: для любого заданного $\varepsilon > 0$ существуют N_k такие, что*

$$\frac{1}{b(n)} \sum_{k=1}^n E |\xi_k - \xi_k^{(N_k)}| < \varepsilon, \quad \frac{1}{b(n)} \sum_{k=1}^n N_k < N < \infty.$$

Тогда последовательность ζ_n сходится к нулю в среднем $\zeta_n \xrightarrow{(1)} 0$.

Доказательство. Очевидно, что $a_k^{(N_k)} = E \xi_k^{(N_k)} \rightarrow a_k$ при $N_k \rightarrow \infty$, $|a_k^{(N_k)}| \leq N_k$. Далее имеем

$$\begin{aligned} E|\zeta_n| &= \frac{1}{b(n)} E \left| \sum (\xi_k - a_k) \right| \leq \frac{1}{b(n)} \sum E |\xi_k - \xi_k^{(N_k)}| + \\ &+ E \left| \sum \frac{(\xi_k^{(N_k)} - a_k^{(N_k)})}{b(n)} \right| + \frac{1}{b(n)} \sum |a_k^{(N_k)} - a_k|. \end{aligned}$$

Здесь среднее слагаемое в правой части сходится к нулю, так как сумма под знаком E удовлетворяет условиям теоремы 3 и ограничена. Первое и последнее слагаемые не превосходят ε . Так как левая часть от ε не зависит, то $E|\zeta_n| \rightarrow 0$ при $n \rightarrow \infty$. ■

Следствие 3. *Если $b(n) = n$ и при достаточно большом N и всех $k \leq n$*

$$E |\xi_k - \xi_k^{(N)}| < \varepsilon,$$

то справедливо утверждение теоремы 3В.

Следствие 3 вытекает из теоремы 3В, так как условия следствия очевидным образом влекут выполнение условий теоремы 3В.

Ясно, что для одинаково распределенных ξ_k условие следствия 3 всегда выполнено, и мы вновь получаем следствие 2.

Если $E|\xi_k|^r < \infty$, то аналогичным образом можно установить также, что

$$\frac{S_n}{n} \xrightarrow{(r)} a.$$

Замечание 3. Условие (D_1) (или (M_1)) не является необходимым для сходимости $\zeta_n \xrightarrow{p} 0$ даже при выполнении (2), (5), на что указывает следующий пример: $\xi_{k,n}$

принимает значения $-n, 0, n$, соответственно с вероятностями $1/n^2, 1-2/n^2, 1/n^2$. Здесь $\zeta_n \xrightarrow{p} 0$, так как $P(\zeta_n \neq 0) \leq P(\bigcup \{\xi_{k,n} \neq 0\}) \leq 2/n \rightarrow 0$, $E|\xi_{k,n}| = 2/n \rightarrow 0$,

$M_1 = \sum E|\xi_{k,n}| = 2 < \infty$. В то же время $\sum E(|\xi_{k,n}|; |\xi_{k,n}| \geq 1) = 2 \not\rightarrow 0$ и условия (D_1) , (M_1) не выполнены.

Однако, если потребовать, чтобы

$$\begin{aligned} \xi_{k,n} &\geq -\varepsilon_{k,n}, \quad \varepsilon_{k,n} \geq 0, \\ \max_{k \leq n} \varepsilon_{k,n} &\rightarrow 0, \quad \sum_{k=1}^n \varepsilon_{k,n} \leq c < \infty, \end{aligned} \quad (8)$$

то условие (D_1) станет необходимым.

Прежде чем доказывать это утверждение, приведем некоторые вспомогательные соотношения, которые будут полезны нам в дальнейшем.

Лемма 1.

$$\sum |\Delta_k(t)| \leq |t| M_1.$$

При выполнении условия (S) при $n \rightarrow \infty$ и при каждом t

$$\max_{k \leq n} |\Delta_k(t)| \rightarrow 0.$$

Если случайная величина ξ , $E\xi = 0$, ограничена слева: $\xi > -c$, $c > 0$, то $E|\xi| \leq 2c$.

Доказательство. В силу леммы 7.2

$$\begin{aligned} |\Delta_k(t)| &\leq E|e^{it\xi_{k,n}} - 1| \leq |t| E|\xi_{k,n}|, \\ \sum |\Delta_k(t)| &\leq |t| M_1. \end{aligned}$$

Далее,

$$|\Delta_k(t)| \leq E(|e^{it\xi_{k,n}} - 1|; |\xi_{k,n}| \leq \varepsilon) + E(|e^{it\xi_{k,n}} - 1|; |\xi_{k,n}| > \varepsilon) \leq |t|\varepsilon + 2P(|\xi_{k,n}| > \varepsilon).$$

Если воспользоваться здесь условием (S) и произвольностью ε , то получим второе утверждение леммы.

Положим $\xi^+ = \max(0; \xi) \geq 0$, $\xi^- = -(\xi - \xi^+) \geq 0$. Тогда $E\xi = E\xi^+ - E\xi^- = 0$, $E|\xi| = E\xi^+ + E\xi^- = 2E\xi^+ \leq 2c$. ■

Из последнего утверждения леммы следует, что условие (8) влечет за собой выполнение (2), (5).

Лемма 2. Пусть выполнены условия (S) и (2). Тогда для сходимости $\varphi_{\zeta_n}(t) \rightarrow \varphi(t)$ необходимо и достаточно, чтобы

$$\sum_{k=1}^n \Delta_k(t) \rightarrow \ln \varphi(t).$$

Доказательство. Заметим, что

$$\operatorname{Re} \Delta_k(t) = \operatorname{Re}(\varphi_{k,n}(t) - 1) \leq 0, \quad |e^{\Delta_k(t)}| \leq 1,$$

и, следовательно, по лемме 7.3

$$\begin{aligned} |\varphi_{\zeta_n}(t) - e^{\sum \Delta_k(t)}| &= \left| \prod_{k=1}^n \varphi_{k,n}(t) - \prod_{k=1}^n e^{\Delta_k(t)} \right| \leq \\ &\leq \sum_{k=1}^n |\varphi_{k,n}(t) - e^{\Delta_k(t)}| = \sum_{k=1}^n |e^{\Delta_k(t)} - 1 - \Delta_k(t)| \leq \\ &\leq \frac{1}{2} \sum_{k=1}^n \Delta_k^2(t) \leq \frac{1}{2} \max_k |\Delta_k(t)| \sum_{k=1}^n |\Delta_k(t)|. \end{aligned}$$

По лемме 1 в силу условий (S) и (2) оцениваемое выражение сходится к 0 при $n \rightarrow \infty$. Поэтому, если $\varphi_{\zeta_n}(t) \rightarrow \varphi(t)$, то $e^{\sum \Delta_k(t)} \rightarrow \varphi(t)$, и наоборот. ■

Следующее утверждение дополняет теорему 3.

Теорема 4. Пусть справедливы соотношения (1), (8). Тогда условие (D_1) (или условие (M_1)) необходимо для выполнения закона больших чисел.

Доказательство. Если выполнен закон больших чисел, то $\varphi_{\zeta_n}(t) \rightarrow 1$ и, следовательно, по лемме 2 (напомним, что (8) влечет за собой (2), (5), (S))

$$\sum_{k=1}^n \Delta_k(t) = \sum_{k=1}^n \mathbb{E}(e^{it\xi_{k,n}} - 1 - it\xi_{k,n}) \rightarrow 0.$$

Кроме того, по лемме 7.2

$$\begin{aligned} \sum_{k=1}^n \mathbb{E}(|e^{it\xi_{k,n}} - 1 - it\xi_{k,n}|; |\xi_{k,n}| \leq \varepsilon_{k,n}) &\leq \\ &\leq \frac{1}{2} \sum_{k=1}^n \mathbb{E}(|\xi_{k,n}|^2; |\xi_{k,n}| \leq \varepsilon_{k,n}) \leq \\ &\leq \sum_{k=1}^n \varepsilon_{k,n}^2 \leq \max_k \varepsilon_{k,n} \sum_{k=1}^n \varepsilon_{k,n} \rightarrow 0. \end{aligned}$$

Поэтому, если имеет место закон больших чисел, то в силу (8)

$$\sum_{k=1}^n \mathbb{E}(e^{it\xi_{k,n}} - 1 - it\xi_{k,n}; \xi_{k,n} > \varepsilon_{k,n}) \rightarrow 0.$$

Рассмотрим функцию $\alpha(x) = (e^{ix} - 1)/ix$. Нетрудно видеть, что неравенство $|\alpha(x)| \leq 1$, доказанное в лемме 7.2, является при $x > \varepsilon > 0$ строгим, и, значит, существует $\delta(\tau) > 0$ при $\tau > 0$ такое, что $\operatorname{Re}(1 - \alpha(x)) \geq \delta(\tau)$ при $x > \tau$. Это равносильно тому, что $\operatorname{Im}(1 + ix - e^{ix}) \geq \delta(\tau)x$ и, следовательно,

$$x \leq \frac{1}{\delta(\tau)} \operatorname{Im}(1 + ix - e^{ix}) \quad \text{при } x > \tau.$$

Отсюда находим

$$\begin{aligned} M_1(\tau) = \sum_{k=1}^n \mathbb{E}(|\xi_{k,n}|; |\xi_{k,n}| > \tau) &= \sum_{k=1}^n \mathbb{E}(\xi_{k,n}; \xi_{k,n} > \tau) \leq \\ &\leq \frac{1}{\delta(\tau)} \operatorname{Im} \sum_{k=1}^n \mathbb{E}(1 + i\xi_{k,n} - e^{i\xi_{k,n}}; \xi_{k,n} > \varepsilon_{k,n}) \rightarrow 0. \end{aligned}$$

Таким образом, выполнено условие (M_1) , которое вместе с соотношением (2), вытекающим из (8), влечет за собой (D_1) . ■

По-видимому, наряду с теоремой 4 имеет место и более сильное утверждение о том, что условие (D_1) необходимо для сходимости $\zeta_n \xrightarrow{(1)} 0$ в среднем (условия (8) влекут эту сходимость).

§ 4. Центральная предельная теорема для сумм произвольных независимых случайных величин

Как и в § 3, будем рассматривать серии случайных величин

$$\xi_{1,n}, \dots, \xi_{n,n}$$

и их суммы

$$\zeta_n = \sum_{k=1}^n \xi_{k,n}. \quad (9)$$

Мы будем предполагать, что $\xi_{k,n}$ имеют конечные вторые моменты $\sigma_{k,n}^2 = D\xi_{k,n} < \infty$, и считать, не ограничивая общности, что

$$E\xi_{k,n} = 0, \quad \sum_{k=1}^n \sigma_{k,n}^2 = D\zeta_n = 1. \quad (10)$$

Введем в рассмотрение условие: *при некотором $s > 2$*

$$D_2 \equiv \sum_{k=1}^n E \min(\xi_{k,n}^2, |\xi_{k,n}|^s) \rightarrow 0, \quad (D_2)$$

которое будет играть важную роль в дальнейшем. При этом многие рассуждения, связанные с условием (D_2) и с условиями (M_2) , (L_s) , вводимыми ниже, будут вполне аналогичны рассуждениям, связанным с условиями (D_1) , (M_1) , (L_s) в § 3.

При выполнении (10) необходимым и достаточным для выполнения (D_2) является условие Линдеберга:

$$M_2(\tau) \equiv \sum_{k=1}^n E(|\xi_{k,n}|^2; |\xi_{k,n}| > \tau) \rightarrow 0 \quad (M_2)$$

при любом $\tau > 0$ и $n \rightarrow \infty$.

Действительно, положим

$$g_2(x) = \min(x^2, |x|^s), \quad s > 2.$$

Тогда (ср. с (3), (4); $\tau \leq 1$)

$$\begin{aligned} D_2 = \sum_{k=1}^n E g_2(\xi_{k,n}) &\leq \sum_{k=1}^n E(\xi_{k,n}^2; |\xi_{k,n}| > \tau) + \sum_{k=1}^n E(|\xi_{k,n}|^s; |\xi_{k,n}| \leq \tau) \leq \\ &\leq M_2(\tau) + \tau^{s-2} M_2(0) = M_2(\tau) + \tau^{s-2}. \end{aligned}$$

Так как τ произвольно, то $D_2 \rightarrow 0$ при $n \rightarrow \infty$, $M_2(\tau) \rightarrow 0$.

Обратно, при $\tau \leq 1$ (условие (10) здесь не требуется)

$$M_2(\tau) \leq \sum_{k=1}^n E(\xi_{k,n}^2; |\xi_{k,n}| > 1) + \frac{1}{\tau^{s-2}} \sum_{k=1}^n (|\xi_{k,n}|^s; \tau < |\xi_{k,n}| \leq 1) \leq \frac{1}{\tau^{s-2}} D_2 \rightarrow 0$$

при любом $\tau > 0$ и $n \rightarrow \infty$.

Итак, при выполнении (10) условие (D_2) инвариантно по $s > 2$.

Условию (D_2) можно придать более общую форму:

$$\sum_{k=1}^n E \xi_{k,n}^2 h(\xi_{k,n}) \rightarrow 0,$$

где $h(x)$ — любая функция, обладающая свойствами $h(x) > 0$ при $x > 0$, $h(x) \uparrow$, $h(x) \rightarrow 0$ при $x \rightarrow 0$, $h(x) \rightarrow c < \infty$ при $x \rightarrow \infty$. При этом все основные свойства условия (D_2) сохраняются. Условие Линденберга выясняет существование условия (D_2) с несколько иной точки зрения. В нем $h(x) = I_{(\tau, \infty)}$, $\tau \in (0, 1)$. Аналогичное замечание можно сделать по поводу условий (D_1) , (M_1) в § 3.

Аналогично тому, как это сделано в § 3 относительно условия (M_1) , легко проверить, что условие (M_2) , влечет за собой сходимость (см. (6))

$$\max_{k \leq n} D \xi_{k,n} \rightarrow 0 \quad (11)$$

и пренебрежимую малость $\xi_{k,n}$ (свойство (S)). Очевидно также неравенство

$$M_1(\tau) \leq \frac{1}{\tau} M_2(\tau).$$

Если $\{\xi_k\}$ есть заданная фиксированная (не зависящая от n) последовательность независимых случайных величин, $S_n = \sum_{k=1}^n \xi_k$, $E \xi_k = a_k$, $D \xi_k = \sigma_k^2$, то для нее будем изучать поведение нормированных сумм

$$\zeta_n = \frac{S_n - \sum_{k=1}^n a_k}{B_n}, \quad B_n^2 = \sum_{k=1}^n \sigma_k^2,$$

которые, очевидно, также имеют вид (9), где $\xi_{k,n} = (\xi_k - a_k)/B_n$.

Условия (D_2) , (M_2) для ξ_k будут иметь вид

$$\begin{aligned} D_2 &= \frac{1}{B_n^2} \sum_{k=1}^n E \min \left((\xi_k - a_k)^2, \frac{|\xi_k - a_k|^s}{B_n^{s-2}} \right) \rightarrow 0, \\ M_2(\tau) &= \frac{1}{B_n^2} \sum_{k=1}^n E ((\xi_k - a_k)^2; |\xi_k - a_k| > \tau B_n) \rightarrow 0. \end{aligned} \quad (12)$$

Теорема 5 (центральная предельная теорема). Если последовательности независимых случайных величин $\{\xi_{k,n}\}_{k=1}^n$, $n = 1, 2, \dots$, удовлетворяют условиям (10), (D_2) , то $P(\zeta_n < x) \rightarrow \Phi(x)$ равномерно по x .

Здесь сохраняют свою силу замечания к теореме 2 о равномерной интегрируемости ζ_n^2 .

Доказательство. Достаточно убедиться, что

$$\varphi_{\zeta_n}(t) = \prod_{k=1}^n \varphi_{k,n}(t) \rightarrow e^{-t^2/2}.$$

В силу леммы 7.3

$$\begin{aligned} \left| \varphi_{\zeta_n}(t) - e^{-t^2/2} \right| &= \left| \prod_{k=1}^n \varphi_{k,n}(t) - \prod_{k=1}^n e^{-t^2 \sigma_{k,n}^2/2} \right| \leq \sum_{k=1}^n \left| \varphi_{k,n}(t) - e^{-t^2 \sigma_{k,n}^2/2} \right| \leq \\ &\leq \sum_{k=1}^n \left| \varphi_{k,n}(t) - 1 + \frac{1}{2} t^2 \sigma_{k,n}^2 \right| + \sum_{k=1}^n \left| e^{-t^2 \sigma_{k,n}^2/2} - 1 + \frac{1}{2} t^2 \sigma_{k,n}^2 \right|. \end{aligned} \quad (13)$$

Так как по лемме 7.2

$$\left| e^{ix} - 1 - ix + \frac{x^2}{2} \right| \leq \min \left(x^2, \frac{|x^3|}{6} \right) \leq g_2(x)$$

(при $s \leq 3$), то первая сумма в (13) не превосходит

$$\begin{aligned} \sum_{k=1}^n \left| \mathbb{E} \left(e^{it\xi_{k,n}} - 1 - it\xi_{k,n} + \frac{t^2 \xi_{k,n}^2}{2} \right) \right| &\leq \\ &\leq \sum_{k=1}^n \mathbb{E} g_2(|t\xi_{k,n}|) \leq h(t) \sum_{k=1}^n \mathbb{E} g_2(|\xi_{k,n}|) \leq h(t) D_2 \rightarrow 0, \end{aligned}$$

где $h(t) = \max(t^2, |t|^3)$. Вторая сумма в (13) (снова по лемме 7.2) не превосходит (см. (10), (11))

$$\frac{t^4}{8} \sum_{k=1}^n \sigma_{k,n}^4 \leq \frac{t^4}{8} \max_k \sigma_{k,n}^2 \rightarrow 0 \quad \text{при } n \rightarrow \infty.$$

Иногда утверждение теоремы 5 бывает удобным в следующей форме.

Следствие 1. Пусть $\mathbb{E} \xi_{k,n} = 0$, $D \zeta_n \rightarrow \sigma^2 > 0$ и выполнено условие (D_2) или (M_2) , тогда $\zeta_n \Leftrightarrow \Phi_{0,\sigma^2}$.

Замечание 3. Утверждение теоремы 2 об асимптотической нормальности $\zeta_n = (S_n - a_n)/\sigma\sqrt{n}$, когда ξ_k одинаково распределены, можно было бы получить в качестве следствия теоремы 5. Нужно лишь проверить выполнение условия Линдеберга. Так как в этом случае $B_n^2 = n\sigma^2$, то (см. (12))

$$M_2(\tau) = \frac{1}{\sigma^2} \mathbb{E}((\xi_k - a)^2; |\xi_k - a| > \tau \sigma \sqrt{n}) \rightarrow 0 \quad \text{при } n \rightarrow \infty$$

для любого $\tau > 0$.

Замечание 4. Достаточным условием для выполнения (D_2) , (M_2) является более ограничительное условие *Ляпунова*. Проверять его в ряде случаев легче. Предположим, что выполнено (10). Величину

$$L_s = \sum_{k=1}^n \mathbb{E} |\xi_{k,n}|^s$$

при некотором $s > 2$ называют *отношением Ляпунова порядка s* . Условие

$$L_s \rightarrow 0 \quad (L_s)$$

называют *условием Ляпунова*. Величина L_s получила название *отношения* потому, что для $\xi_{k,n} = (\xi_k - a_k)/B_n$ ($a_k = E\xi_k$, $B^2 = \sum_{k=1}^n D\xi_k$), ξ_k от n не зависят) L_s имеет вид

$$L_s = \frac{\sum_{k=1}^n E|\xi_k - a_k|^s}{B_n^s}.$$

Если ξ_k одинаково распределены, $a_k = a$, $D\xi_k = \sigma^2$, $E|\xi_k - a|^s = \mu < \infty$, то

$$L_s = \frac{\mu}{\sigma^s n^{(s-2)/2}} \rightarrow 0.$$

Достаточность условия Ляпунова вытекает из очевидных неравенств $g_2(x) \leq |x|^s$ при любом s , $D_2 \leq L_s$.

Замечание 5. Условия (D_2) или (M_2) не являются необходимыми для сходимости распределения ζ_n к нормальному. Достаточно рассмотреть следующий пример: $\xi_{1,n} \Rightarrow \Phi_{0,1}$, $\xi_{2,n} = \dots = \xi_{n,n} = 0$. Очевидно, что условия (10) выполнены, $P(\zeta_n < x) = \Phi(x)$, а слагаемые $\xi_{k,n}$ не являются пренебрежимо малыми и, следовательно, не удовлетворяют условиям (D_2) , (M_2) .

Однако, если наряду со сходимостью $\zeta_n \Rightarrow \Phi_{0,1}$ потребовать, чтобы $\xi_{k,n}$ были пренебрежимо малы, то условия (D_2) , (M_2) становятся необходимыми.

Теорема 6. Пусть последовательности независимых величин $\{\xi_{k,n}\}_{k=1}^n$ удовлетворяют условиям (10), (S). Тогда условие (D_2) (или (M_2)) необходимо и достаточно для сходимости $\zeta_n \in \Phi_{0,1}$.

Заметим предварительно, что утверждения лемм 1, 2 при замене условий (2), (S) на (10), (S) сохраняются с точностью до несущественных изменений.

Лемма 1А. Пусть выполнены условия (10), (S). Тогда

$$\max_{k \leq n} |\Delta_k(t)| \rightarrow 0, \quad \sum |\Delta_k(t)| \leq \frac{t^2}{2},$$

и полностью сохранится утверждение леммы 2.

Доказательство. Все рассуждения в доказательствах лемм 1, 2 полностью сохраняются за исключением одного места, связанного с оценкой $\sum |\Delta_k(t)|$. В новых условиях в силу леммы 7.2 имеем

$$|\Delta_k(t)| = |\varphi_{k,n}(t) - 1 - itE\xi_{k,n}| \leq E|e^{it\xi_{k,n}} - 1 - it\xi_{k,n}| \leq \frac{t^2}{2} E\xi_{k,n}^2,$$

так что

$$\sum |\Delta_k(t)| \leq \frac{t^2}{2}.$$

Другие изменения в доказательствах лемм 1, 2 отсутствуют.

Доказательство теоремы 6. Достаточность уже доказана. Чтобы доказать необходимость, воспользуемся леммой 1А. Если $\varphi_{\zeta_n}(t) \rightarrow e^{-t^2/2}$, то в силу этой леммы $(\Delta_k(t) = \varphi_{k,n}(t) - 1)$

$$\sum_{k=1}^n \Delta_k(t) \rightarrow \ln \varphi(t) = \frac{-t^2}{2}.$$

При $t = 1$ выписанное соотношение можно записать в форме

$$R_n \equiv \sum_{k=1}^n \mathbb{E} \left(e^{i\xi_{k,n}} - 1 - i\xi_{k,n} + \frac{\xi_{k,n}^2}{2} \right) \rightarrow 0. \quad (14)$$

Обозначим $\alpha(x) = (e^{ix} - 1 - ix)/x^2$. Нетрудно видеть, что неравенство $|\alpha(x)| \leq 1/2$, доказанное в лемме 7.2, является строгим при $x \neq 0$ и

$$\sup_{|x| \geq \tau} |\alpha(x)| < \frac{1}{2} - \delta(\tau),$$

где $\delta(\tau) > 0$ при $\tau > 0$. Это означает, что при $|x| \geq \tau > 0$

$$\operatorname{Re} \left[\alpha(x) + \frac{1}{2} \right] \geq \delta(\tau) > 0,$$

$$x^2 \leq \frac{1}{\delta(\tau)} \operatorname{Re} \left(e^{ix} - 1 - ix + \frac{x^2}{2} \right),$$

$$\mathbb{E}(\xi_{k,n}^2; |\xi_{k,n}| > \tau) \leq \frac{1}{\delta(\tau)} \operatorname{Re} \mathbb{E} \left(e^{i\xi_{k,n}} - 1 - i\xi_{k,n} + \frac{\xi_{k,n}^2}{2} \right),$$

$$M_2(\tau) \leq \frac{1}{\delta(\tau)} |R_n| \rightarrow 0$$

в силу (14) при $n \rightarrow \infty$ для любого $\tau > 0$. ■

Следствие 2. Пусть выполнено (10) и

$$\max_{k \leq n} D\xi_{k,n} \rightarrow 0. \quad (15)$$

Тогда для сходимости $\zeta_n \Leftrightarrow \Phi_{0,1}$ необходима и достаточна сходимость

$$\eta_n \equiv \sum_{k=1}^n \xi_{k,n}^2 \Leftrightarrow I_1$$

(или сходимость $\eta_n \xrightarrow{p} 1$).

Доказательство. Пусть $\eta_n \Leftrightarrow I_1$. Случайные величины $\xi'_{k,n} = \xi_{k,n}^2 - \sigma_{k,n}^2$ в силу (15) удовлетворяют условию (8) и удовлетворяют закону больших чисел: $\sum_{k=1}^n \xi'_{k,n} = \eta_n - 1 \xrightarrow{p} 0$. Следовательно, по теореме 4 $\xi'_{k,n}$ удовлетворяют условию (M₁):

$$\sum_{k=1}^n \mathbb{E}(|\xi_{k,n}^2 - \sigma_{k,n}^2|; |\xi_{k,n}^2 - \sigma_{k,n}^2| > \tau) \rightarrow 0 \quad (16)$$

при любом $\tau > 0$. Но по (15) это условие эквивалентно, очевидно, условию (M₂) для $\xi_{k,n}$. Отсюда следует $\zeta_n \Leftrightarrow \Phi_{0,1}$.

Наоборот, если $\zeta_n \Leftrightarrow \Phi_{0,1}$, то выполнено (M₂) для $\xi_{k,n}$ и, значит, условие (16). Так как, кроме того,

$$\sum_{k=1}^n \mathbb{E}|\xi'_{k,n}| \leq 2 \sum_{k=1}^n D\xi_{k,n} = 2,$$

то для $\xi'_{k,n}$ выполнено (2), и по теореме 3

$$\sum_{k=1}^n \xi'_{k,n} = \eta_n - 1 \xrightarrow{p} 0.$$

Пример 1. Пусть ξ_k , $k = 1, 2, \dots$, — независимые случайные величины с распределениями

$$\xi_k = \begin{cases} k^\alpha & \text{с вероятностью } 1/2, \\ -k^\alpha & \text{с вероятностью } 1/2. \end{cases}$$

Здесь ($s = 3$)

$$\sigma_k^2 = k^{2\alpha}, \quad \mu_k \equiv E|\xi_k|^3 = k^{3\alpha},$$

$$B_n^2 = \sum_{k=1}^n k^{2\alpha}, \quad C_n^3 \equiv \sum_{k=1}^n \mu_k = \sum_{k=1}^n k^{3\alpha}.$$

Имеем при $2\alpha + 1 > 0$

$$B_n^2 = \sum_{k=1}^n k^{2\alpha} = n^{2\alpha+1} \left(\frac{1}{n} + \left(\frac{n-1}{n} \right)^{2\alpha} \frac{1}{n} + \dots + \left(\frac{1}{n} \right)^{2\alpha} \frac{1}{n} \right) \sim$$

$$\sim n^{2\alpha+1} \int_0^1 x^{2\alpha} dx = \frac{n^{2\alpha+1}}{2\alpha+1}.$$

При $\alpha = -1/2$

$$B_n^2 = \sum_{k=1}^n k^{-1} \sim \ln n.$$

Если $\alpha < -1/2$, то B_n ограничены при $n \rightarrow \infty$ и, следовательно, не будет выполнено условие пренебрежимой малости слагаемых.

Аналогичным образом находим при $\alpha > -1/3$

$$C_n^3 = \sum_{k=1}^n k^{3\alpha} = n^{3\alpha+1} \left(\frac{1}{n} + \left(\frac{n-1}{n} \right)^{3\alpha} \frac{1}{n} + \dots + \left(\frac{1}{n} \right)^{3\alpha} \frac{1}{n} \right) \sim$$

$$\sim n^{3\alpha+1} \int_0^1 x^{3\alpha} dx = \frac{n^{3\alpha+1}}{3\alpha+1}.$$

При $\alpha = -1/3$ имеем $C_n^3 \sim \ln n$; если $\alpha < -1/3$, то C_n ограничены. Так как $n^{\alpha+1/3} = o(n^{\alpha+1/2})$, то при $\alpha \geq -1/2$ $C_n = o(B_n)$, и условие Ляпунова $L_3 = C_n^3/B_n^3 \rightarrow 0$ будет выполнено. Следовательно, суммы $S_n \sqrt{2\alpha+1}/n^{\alpha+1/2}$ при $\alpha > -1/2$ будут асимптотически нормальны с параметрами $(0, 1)$. Если же $\xi_k = \pm 1/\sqrt{k}$ с вероятностями $1/2$, то асимптотически нормальной будет величина $S_n/\sqrt{\ln n}$. В частности, при больших n величина $|S_n|$ не будет превосходить $2\sqrt{\ln n}$ с вероятностью, равной примерно $2\Phi(2) - 1 \approx 0,955$. Для сравнения можно заметить, что $\sum_{k=1}^n \frac{1}{\sqrt{k}}$ растет как $\sqrt{n}/2$, а знакопередающиеся суммы $\sum_{k=1}^n \frac{(-1)^k}{\sqrt{k}}$ ограничены постоянным числом.

Очень наглядной и хорошо известной иллюстрацией действия доказанных теорем является рассеивание артиллерийских снарядов при стрельбе по цели. Дело в том, что на траекторию полета снаряда действует большое количество независимых факторов, влияние каждого из которых само по себе невелико. Это отклонения в количестве заряда, отклонения в весе и размерах снаряда, отклонения по влажности и температуре воздуха, направлении и силе ветра на разных высотах и т. д. В результате отклонение снаряда от точки прицеливания удивительно точно описывается нормальным законом.

Такого же рода наблюдения можно сделать относительно ошибок в измерениях, когда на их точность влияет много «мелких» факторов. (Существует даже «теория ошибок», главным пунктом которой является центральная предельная теорема.)

В целом центральная предельная теорема находит чрезвычайно много применений в самых разных областях. Это происходит благодаря ее *универсальности*, ее *устойчивости* относительно незначительных отклонений от условий теоремы и ее сравнительно *высокой точности* даже при умеренных значениях n . Первые два отмеченных качества означают, что:

1) теорема применима к величинам $\xi_{k,n}$ с *любыми распределениями*, лишь бы существовали дисперсии $\xi_{k,n}$ и они были «пренебрежимо малы»;

2) появление «умеренной» зависимости³⁾ между $\xi_{k,n}$ не меняет нормальности предельного распределения.

Чтобы проиллюстрировать точность нормального приближения, приведем такой пример. Пусть $F_n(x) = P(S_n/\sqrt{n} < x)$ есть функция распределения нормированных сумм S_n , образованных независимыми величинами ξ_k , равномерно распределенными на $[-\sqrt{3}, \sqrt{3}]$, так что $D\xi_k = 1$. Тогда оказывается, что уже при $n = 5$ (!) максимальная величина разности $|F_n(x) - \Phi(x)|$ на всей оси значений x не превосходит 0,006 (максимум достигается около точек $x = \pm 0,7$).

И все, же несмотря на отмеченные обстоятельства, при использовании центральной предельной теоремы следует проявлять осторожность. Например, нельзя требовать точности от нормального приближения при оценке вероятностей редких событий, скажем, при изучении вероятностей больших уклонений (этот вопрос уже обсуждался в §5.3). Ведь теорема обеспечивает лишь малость при больших n разности

$$|\Phi(x) - P(\zeta_n < x)|. \quad (17)$$

Допустим, мы хотим с помощью нормального приближения указать такое x_0 , чтобы гарантировать появление события $\{\zeta_n > x_0\}$ в среднем не чаще, чем 1 раз на 1000 случаев (такая задача может возникнуть, скажем, у экспериментатора, когда он хочет быть уверенным, что в отдельном эксперименте событие такого вида не произойдет). Если даже разность (17) не превосходит 0,02 (что может считаться хорошим приближением), то, используя нормальное приближение, мы рискуем ошибиться в десятки раз, поскольку может оказаться, что $1 - \Phi(x_0) = 10^{-3}$, а $P(\zeta_n > x_0) \approx 0,02$, и событие $\{\zeta_n > x_0\}$ будет появляться значительно чаще — в среднем 1 раз на 50 случаев.

В связи с этим в §8 рассматривается задача о вероятностях больших уклонений, где отыскивается функция $P(n, x)$ такая, что $P(\zeta_n > x)/P(n, x) \rightarrow 1$ при $n \rightarrow \infty$, $x \rightarrow \infty$. При этом функция $P(n, x)$ оказывается, вообще говоря, отличной от $1 - \Phi(x)$. Надо отметить, однако, что использование приближения $P(n, x)$ требует

³⁾ Существует целый ряд условий, характеризующих допустимую зависимость между $\xi_{k,n}$. Знакомство с ними выходит за рамки настоящей книги и требует специальной литературы (см., например, [11]).

выполнения более жестких условий на $\xi_{k,n}$. В § 6 приводится также *локальная предельная теорема*, которая устанавливает сближение *плотностей* ζ_n с плотностью нормального закона и позволяет оценивать вероятности редких событий другого сорта, скажем, вида $\{a < \zeta_n < b\}$, где числа a и b близки.

§ 5. Другой подход к доказательству предельных теорем. Оценки погрешности

Изложенный в § 1–4 подход к доказательству основных предельных теорем для распределения сумм случайных величин основан на х. ф. Однако это далеко не единственный метод доказательства такого рода утверждений. В настоящее время известно несколько весьма простых доказательств как законов больших чисел, так и центральной предельной теоремы, не использующих аппарата х. ф. (Это не умаляет значимости последнего, представляющего собой мощный, хорошо разработанный и весьма общий аппарат исследований.) Эти доказательства позволяют к тому же получать в некоторых случаях более общие результаты. В качестве иллюстрации приведем ниже доказательство центральной предельной теоремы, обобщающей в известном смысле теорему 6 и дающей оценку скорости сходимости (правда, не наилучшую).

Наряду со случайными величинами $\xi_{k,n}$ в схеме серий и при выполнении (10) рассмотрим независимые между собой и от последовательности $\{\xi_{k,n}\}_{k=1}^n$ случайные

величины $\eta_{k,n} \in \Phi_{0,\sigma_{k,n}^2}$, $\sigma_{k,n}^2 = D\xi_{k,n}$, так что $\eta_n = \sum_{k=1}^n \eta_{k,n} \in \Phi_{0,1}$.

Обозначим $\mu_{k,n} = E|\xi_{k,n}|^3$, $\nu_{k,n} = E|\eta_{k,n}|^3 = c_3\sigma_{k,n}^3 \leq c_3\mu_{k,n}$ ⁴⁾, $\mu_{k,n}^0 = \int |x|^3 |d(F_{k,n}(x) - \Phi_{k,n}(x))| \leq \mu_{k,n} + \nu_{k,n}$, $L_3 = \sum_{k=1}^n \mu_{k,n}$, $N_3 = \sum_{k=1}^n \nu_{k,n}$, $L_3^0 = \sum_{k=1}^n \mu_{k,n}^0 \leq L_3 + N_3 \leq (1 + c_3)L_3$. Здесь $F_{k,n}$, $\Phi_{k,n}$ — функции распределения $\xi_{k,n}$ и $\eta_{k,n}$ соответственно. Числа L_3 и N_3 суть отношения Ляпунова третьего порядка для последовательностей $\{\xi_{k,n}\}$ и $\{\eta_{k,n}\}$. Числа $\mu_{k,n}^0$ называются *псевдомоментами* третьего порядка, а величина L_3^0 — отношением Ляпунова для псевдомоментов. Очевидно, что $N_3 \leq c_3 L_3 \rightarrow 0$, если выполнено условие Ляпунова. Как уже отмечалось, для $\xi_{k,n} = (\xi_k - a_k)/B_n$, где $a_k = E\xi_k$, $B_n^2 = \sum_{k=1}^n D\xi_k$, ξ_k от n не зависят,

$L_3 = \frac{1}{B_n^3} \sum_{k=1}^n \mu_k$, $\mu_k = E|\xi_k - a_k|^3$. Если к тому же ξ_k одинаково распределены, то $L_3 = \mu_1/(\sigma^3\sqrt{n})$.

Нашей первой задачей будет оценка близости $Ef(\zeta_n)$ и $Ef(\zeta_n)$ для достаточно гладких f . Эта задача может представлять и самостоятельный интерес. Предположим, что f принадлежит классу C_3 всех ограниченных функций с равномерно непрерывной и ограниченной третьей производной: $\sup_x |f^{(3)}(x)| \leq f_3$.

⁴⁾ Если $\eta \in \Phi_{0,1}$, то

$$c_3 = E|\eta|^3 = \frac{2}{\sqrt{2\pi}} \int_0^\infty x^3 e^{-x^2/2} dx = \frac{4}{\sqrt{2\pi}} \int_0^\infty te^{-t} dt = \frac{4}{\sqrt{2\pi}}.$$

Теорема 7. Если $f \in C_3$, то

$$|Ef(\zeta_n) - Ef(\eta_n)| \leq \frac{f_3 L_3^0}{6} \leq \frac{f_3}{6} (L_3 + N_3). \quad (18)$$

Доказательство. Обозначим

$$\begin{aligned} X_l &= \xi_{1,n} + \dots + \xi_{l-1,n} + \eta_{l,n} + \dots + \eta_{n,n}, \\ Z_l &= \xi_{1,n} + \dots + \xi_{l-1,n} + \eta_{l+1,n} + \dots + \eta_{n,n}, \\ X_1 &= \eta_n, \quad X_{n+1} = \zeta_n. \end{aligned}$$

Тогда

$$X_{n+1} = Z_1 + \xi_{l,n}, \quad X_l = Z_l + \eta_{l,n}, \quad (19)$$

$$f(\zeta_n) - f(\eta_n) = \sum_{l=1}^n [f(X_{l+1}) - f(X_l)]. \quad (20)$$

Воспользуемся теперь следующей леммой.

Лемма 3. Пусть $f \in C_3$, Z , ξ и η суть независимые случайные величины,

$$Ef = E\eta = a, \quad E\xi^2 = E\eta^2 = \sigma^2, \quad \mu^0 = \int |x^3| |d(F_\xi(x) - F_\eta(x))| < \infty.$$

Тогда

$$|Ef(Z + \xi) - Ef(Z + \eta)| \leq \frac{f_3 \mu^0}{6}. \quad (21)$$

Применяя эту лемму к (19), получим

$$|E[f(X_{l+1}) - f(X_l)]| \leq \frac{f_3 \mu_{l,n}^0}{6},$$

что после суммирования дает (18). ■

Итак, для доказательства теоремы 7 осталось провести

Доказательство леммы 3. Обозначим $g(x) = Ef(Z + x)$. Очевидно, что g , будучи усреднением f , обладает всеми свойствами гладкости f и, в частности, $|g'''(x)| \leq f_3$. В силу независимости Z , ξ и η

$$Ef(Z + \xi) - Ef(Z + \eta) = \int g(x) d(F_\xi(x) - F_\eta(x)). \quad (22)$$

Под интегралом воспользуемся разложением

$$g(x) = g(0) + xg'(0) + \frac{x^2}{2}g''(0) + \frac{x^3}{6}g'''(\theta_x), \quad \theta_x \in [0, x].$$

В силу совпадения моментов ξ и η до второго порядка для правой части (22) получим

$$\left| \frac{1}{6} \int x^3 g'''(\theta_x) d(F_\xi(x) - F_\eta(x)) \right| \leq \frac{f_3 \mu^0}{6}. \quad \blacksquare$$

Замечание 6. Совершенно аналогично можно получить представление

$$|E(\zeta_n) - Ef(\eta_n)| \leq \frac{g'''(0)}{6} \sum_{k=1}^n E(\xi_{k,n}^3 - \eta_{k,n}^3) + \frac{f_4 L_4^0}{24} \quad (23)$$

при очевидных соглашениях относительно обозначений f_4 и L_4^0 . Эта оценка может улучшать (18), если разности $E(\xi_{k,n}^3 - \eta_{k,n}^3)$ малы. Если, например, $\xi_{k,n} = (\xi_k - a)/(\sigma\sqrt{n})$, ξ_k одинаково распределены и третьи моменты $\xi_{k,n}$ и $\eta_{k,n}$ совпадают, то справа в (23) будет стоять величина порядка $1/n$.

Теорема 7 обобщает теорему 5 при $s = 3$. Обобщение состоит в том, что теперь для сходимости $\zeta_n \Rightarrow \Phi_{0,1}$ не требуется предельной пренебрегаемости $\xi_{k,n}$. Если, например, $\xi_{1,n} \in \Phi_{0,1/2}$ (в этом случае $\mu_{1,n}^0 = 0$) и $L_3^0 \rightarrow 0$, то $Ef(\zeta_n) \rightarrow Ef(\eta)$, $\eta \in \Phi_{0,1}$ для любой f из класса C_3 . Так как C_3 определяет распределение (см. главу 6), то остается воспользоваться следствием 6.7.

Высказанное утверждение можно усилить.

Теорема 7А.

$$|P(\zeta_n < x) - \Phi(x)| < c(L_3^0)^{1/4}, \quad (24)$$

где c — абсолютная постоянная.

Доказательство. Возьмем какую-нибудь функцию $h \in C_3$, $0 \leq h \leq 1$, такую, что $h(x) = 1$, $x \leq 0$; $h(x) = 0$, $x \geq 1$, и обозначим $h_3 = \sup_x |h'''(x)|$. Тогда для функции $f(x) = h((x-t)/\varepsilon)$ будем иметь $f_3 = \sup_x |f'''(x)| \leq h_3/\varepsilon^3$,

$$\begin{aligned} P(\zeta_n < t) &\leq Ef(\zeta_n) \leq Ef(\eta) + \frac{f_3 L_3^0}{6} \leq \\ &\leq P(\eta < t + \varepsilon) + \frac{h_3 L_3^0}{6\varepsilon^3} \leq P(\eta < t) + \frac{\varepsilon}{\sqrt{2\pi}} + \frac{h_3 L_3^0}{6\varepsilon^3}. \end{aligned}$$

Последнее неравенство имеет место, так как максимум производной нормальной функции распределения $\Phi(t) = P(\eta < t)$ равен $1/\sqrt{2\pi}$. Устанавливая точно так же обратное неравенство и полагая $\varepsilon = (L_3^0)^{1/4}$, получим (24). ■

Оценка в теореме 7А, конечно, не является наилучшей. И все же неравенство (24) показывает, что мы будем иметь хорошее нормальное приближение для $P(\zeta_n < x)$ и в области больших отклонений (т. е. при $|x| \rightarrow \infty$), по крайней мере для x , при которых при $n \rightarrow \infty$

$$(1 - \Phi(|x|))(L_3^0)^{-1/4} \rightarrow \infty. \quad (25)$$

Действительно, в этом случае, например, при $x = |x| > 0$

$$\left| \frac{P(\zeta_n \geq x)}{1 - \Phi(x)} - 1 \right| \leq \frac{c(L_3^0)^{1/4}}{1 - \Phi(x)} \rightarrow 0.$$

Так как по правилу Лопиталя

$$1 - \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-t^2/2} dt \sim \frac{1}{\sqrt{2\pi} x} e^{-x^2/2} \quad \text{при } x \rightarrow \infty,$$

то (25) справедливо для $|x| < c_1 \sqrt{-\ln L_3^0}$ при подходящей постоянной c_1 .

В главе 19 мы получим обобщение теорем 7, 7А.

Вопрос об уточнениях и оценках погрешности в центральной предельной теореме и в других предельных теоремах является одним из важных в теории вероятностей, поскольку от ответа на него зависит, насколько точными и эффективными будут применения этих теорем к задачам практики. Надо, прежде всего, найти правильный порядок убывания $\Delta_n = \sup_x |P(\zeta_n < x) - \Phi(x)|$ по n (или, скажем, по L_3

в случае разнораспределенных слагаемых). Существует по крайней мере два подхода к отысканию правильных оценок Δ_n . Один из них, *метод характеристических функций*, исходит из неулучшаемой оценки близости характеристических функций

$$\left| \ln \varphi_{\zeta_n}(t) + \frac{t^2}{2} \right| < cL_3,$$

которую читатель может получить самостоятельно, пользуясь леммой 7.2 и усложнив немного рассуждения в доказательстве теоремы 5. Основные технические трудности здесь состоят в том, чтобы с помощью формулы обращения получить такой же порядок малости для Δ_n .

Второй подход, так называемый *метод композиций*, был проиллюстрирован в этом параграфе в теореме 7 (в какой-то мере существо его заложено в соотношении (20)). Именно этим методом в Приложении 5 для случая одинаково распределенных слагаемых будет доказан следующий общий результат (Крамер, Берри, Эссен):

Теорема. Если ξ_k не зависит от n , $\xi_{k,n} = (\xi_k - a_k)/B_n$, то

$$\sup_x |P(\zeta_n - x) - \Phi(x)| \leq cL_3,$$

где c — абсолютная постоянная.

В случае одинаково распределенных ξ_k правая часть неравенства превращается в $c\mu_1/\sigma^3\sqrt{n}$. Установлено, что $(2\pi)^{-1/2} < c < 0,8^5$.

Здесь надо иметь в виду, что приведенные теоремы и оценка постоянной c универсальны и выполняются, следовательно, в самых неблагоприятных с точки зрения приближений условиях. Поэтому скорость сходимости в реальных задачах, как правило, оказывается лучше.

§ 6. Локальная предельная теорема

В предыдущих параграфах мы установили, что распределение нормированной суммы $S_n = \sum_{j=1}^n \xi_j$ одинаково распределенных и независимых величин ξ_j с $E\xi_j = 0$,

$D\xi_j = \sigma^2 < \infty$ сходится к нормальному закону. Допустим теперь, что распределение $\zeta_n = S_n/(\sigma\sqrt{n})$ имеет плотность. Будет ли она равномерно сближаться с плотностью нормального приближения? Положительный ответ на этот вопрос позволил бы оценивать вероятности попадания ζ_n в произвольные измеримые множества⁶⁾,

⁵⁾ См. [28].

⁶⁾ Слабая сходимость распределений $P_n \Rightarrow P$ означает сходимость вероятностей $P_n(B) \rightarrow P(B)$ лишь для множеств B , граница которых ∂B (это есть разность между замыканием B и дополнением к замыканию $\mathbb{R} - B$) обладает свойством $P(\partial B) = 0$. Существенность этого условия хорошо иллюстрируется следующим примером. Пусть ξ_i целочисленны, $\sigma = 1$, а B есть множество всех

и в том числе вероятности событий вида $\{S_n \in \Delta\}$, где Δ — фиксированный интервал (из центральной предельной теоремы следует лишь, что $P(S_n \in \Delta) = P(\zeta_n \in \Delta/(\sigma\sqrt{n})) \rightarrow 0$ при $n \rightarrow \infty$).

Теорема 8. Пусть $\{\xi_k\}_{k=1}^\infty$ — последовательность независимых одинаково распределенных случайных величин,

$$E\xi_k = 0, D\xi_k = \sigma^2, 0 < \sigma^2 < \infty, \varphi(t) = Ee^{it\xi_1}.$$

Предположим, что при некотором $m \geq 1$ распределение суммы S_m имеет плотность, и, кроме того, х.ф. $\varphi^m(t)$ интегрируема. Тогда плотности распределения $f_{\zeta_n}(x)$ случайных величин $\zeta_n = S_n/(\sigma\sqrt{n})$ равномерно по x сходятся к плотности нормального закона $f_{\zeta_n}(x) \rightarrow 1/\sqrt{2\pi}e^{-x^2/2}$.

Например, плотность равномерного на $[-1, 1]$ распределения удовлетворяет этой теореме, так как для нее $|\varphi^2(t)| \leq 1/t^2$ (см. пример 7.7) и, следовательно, $|\varphi^2(t)|$ интегрируема. Следовательно, по теореме 6 плотность распределения S_n в точке $u\sigma\sqrt{n}$ ($\sigma^2 = 1/3$) будет вести себя как $1/(\sigma\sqrt{2\pi n})e^{-u^2/2}$ (ср. с замечанием к примеру 3.4).

Интегрируемость $\varphi^2(t)$ имеет место также во всех случаях, когда плотность ξ_k есть функция ограниченной вариации. В этом легко убедиться, если проинтегрировать х.ф. по частям.

Доказательство. В силу формулы обращения при $n \geq m$

$$f_{\zeta_n}(x) = \frac{1}{2\pi} \int e^{-itx} \varphi^n \left(\frac{t}{\sigma\sqrt{n}} \right) dt. \quad (26)$$

Здесь, как мы знаем, при каждом t

$$\varphi^n \left(\frac{t}{\sigma\sqrt{n}} \right) \rightarrow e^{-t^2/2}. \quad (27)$$

Поэтому, чтобы доказать требуемое равенство

$$\lim_{n \rightarrow \infty} f_{\zeta_n}(x) = \frac{1}{2\pi} \int e^{-itx} e^{-t^2/2} dt = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad (28)$$

достаточно убедиться, что $\varphi^n(1/(\sigma\sqrt{n}))$ при всех n мажорируется интегрируемой функцией.

Так как дисперсия конечна, то $\varphi''(t)$ непрерывна, $\varphi''(0) = -\sigma^2$. Пусть $\delta > 0$ таково, что при $|t| < \delta$

$$|\varphi''(t) - \varphi''(0)| \leq \varepsilon \sigma^2$$

для заданного $\varepsilon > 0$. Тогда при $|t| < \delta$

$$\varphi(t) = 1 - \frac{t^2 \sigma^2}{2} - \frac{t^2}{2} (\varphi''(t) - \varphi''(0)) = 1 - \frac{t^2 \sigma^2}{2} (1 + \theta_1),$$

рациональных чисел. Тогда

$$P_n(B) = P(\zeta_n \in B) = P \left(\frac{S_n}{\sqrt{n}} \in B \right) = 1,$$

если $n = k^2$ (k — целое), и $P_n(B) = 0$ в противном случае: так что $\lim_{n \rightarrow \infty} P_n(B)$ не существует. В этом

примере для предельного нормального распределения P , очевидно, выполняется $P_n \Rightarrow P$, $P(\partial B) = 1$, так как ∂B совпадает здесь со всей вещественной прямой \mathbb{R} .

Одновременно можно заметить, что для любого распределения P , имеющего плотность (в том числе для нормального), в силу счетности B будем иметь $P(B) = 0$.

где $\bar{t} \in [0, t]$, $|\theta_1| < \varepsilon$. Используя неравенство $|\ln(1 - \beta) - \beta| \leq \beta^2$, получаем, что

$$\ln \varphi(t) = -\frac{t^2 \sigma^2}{2} (1 + \theta_2), \quad |\theta_2| < \varepsilon. \quad (29)$$

Отсюда находим, что при $\left| \frac{t}{\sigma\sqrt{n}} \right| < \delta$

$$\left| \varphi^n \left(\frac{t}{\sigma\sqrt{n}} \right) \right| = e^{n \operatorname{Re} \ln \varphi} \leq e^{-t^2/2(1-\varepsilon)}.$$

В дополнительной области $\left| \frac{t}{\sigma\sqrt{n}} \right| \geq \delta$ действует оценка

$$\sup_{\left| \frac{t}{\sigma\sqrt{n}} \right| \geq \delta} \left| \varphi \left(\frac{t}{\sigma\sqrt{n}} \right) \right| < 1 - \gamma < 1,$$

которая справедлива в силу того, что S_m имеет плотность. Поэтому

$$\begin{aligned} \int_{\left| \frac{t}{\sigma\sqrt{n}} \right| < \delta} \left| \varphi^n \left(\frac{t}{\sigma\sqrt{n}} \right) \right| dt &\leq (1 - \gamma)^{n-m} \int_{\left| \frac{t}{\sigma\sqrt{n}} \right| > \delta} \left| \varphi^m \left(\frac{t}{\sigma\sqrt{n}} \right) \right| dt \leq \\ &\leq (1 - \gamma)^{n-m} \sigma\sqrt{n} \int |\varphi^m(u)| du \leq (1 - \gamma)^n \sqrt{n} \frac{c}{(1 - \gamma)^m} \rightarrow 0 \end{aligned}$$

при $n \rightarrow \infty$ в силу интегрируемости $|\varphi^m(u)|$. Соотношение (28) доказано.

Пользуясь тем, что $\varphi^n(t/\sqrt{n})$ мажорируется интегрируемой функцией, легко установить, что $f_{\zeta_n}(x)$ равномерно по n непрерывны. Это делается с помощью представления (26) точно так же, как доказывается непрерывность х.ф. Равностепенная непрерывность f_{ζ_n} , непрерывность функции $e^{-x^2/2}$ и соотношение (28) влекут за собой равномерную сходимость $f_{\zeta_n}(x)$ к плотности нормального закона. ■

Следствие 3. Из доказанной теоремы следует, в частности, что при $n \rightarrow \infty$ и при $|z|/\sqrt{n} < c$

$$P \left(\frac{z}{\sigma\sqrt{n}} < \zeta_n < \frac{z+a}{\sigma\sqrt{n}} \right) \sim \frac{1}{\sqrt{2\pi}} \frac{a}{\sigma\sqrt{n}} e^{-z^2/(2\sigma^2 n)}, \quad (30)$$

где c, a — фиксированные числа.

Мы предлагаем читателю сравнить это соотношение с локальной предельной теоремой для схемы Бернулли (см., в частности, формулу (5.7)).

Соотношение (30) можно получить и при более широких условиях.

Теорема 9. Соотношение (30) останется справедливым, если в теореме 8 условие интегрируемости $\varphi^m(t)$ заменить на интегрируемость $(\varphi^m(t))/(1 + |t|)$.

Доказательство. По формуле обращения

$$F_{\zeta_n}(y) - F_{\zeta_n}(x) = \frac{1}{2\pi} \int \frac{e^{-itz} - e^{-ity}}{it} \varphi^n \left(\frac{t}{\sigma\sqrt{n}} \right) dt.$$

При $x = z/(\sigma\sqrt{n})$, $y = x + a/(\sigma\sqrt{n})$ в правой части получим

$$\frac{1}{2\pi\sigma\sqrt{n}} \int e^{-itz} \frac{1 - e^{-ita/(\sigma\sqrt{n})}}{it/(\sigma\sqrt{n})} \varphi^n \left(\frac{t}{\sigma\sqrt{n}} \right) dt,$$

где $\frac{1-e^{-itx/(\sigma\sqrt{n})}}{it/(\sigma\sqrt{n})} \rightarrow a$, $\varphi^n(t/(\sigma\sqrt{n})) \rightarrow e^{-t^2/2}$. Здесь можно опять использовать теорему о мажорируемой сходимости (по тем же причинам, что и прежде, и при очевидных изменениях, связанных с появлением под интегралом множителя $1/t$). Получим в результате, что

$$F_{\zeta_n}\left(x + \frac{a}{\sigma\sqrt{n}}\right) - F_{\zeta_n}(x) \sim \frac{a}{\sigma\sqrt{n}} \frac{1}{2\pi} \int e^{-itx - \frac{t^2}{2}} dt = \frac{a}{\sigma\sqrt{n}} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

В дальнейшем нам понадобится несколько более общая постановка задачи. Мы предположим, что распределения случайных величин ξ_j зависят от n (схема серий). Следовательно, будут зависеть от n и х. ф. $\varphi(t) = \varphi_n(t) = E e^{it\xi_j}$.

Теорема 10. Пусть

- 1) $\varphi_n^m(t)$ при некотором $m \geq 1$ интегрируемы равномерно по n ;
- 2) $\sup_n |\varphi_n(t)| < 1 - \delta < 1$ при всех t , $|t| > \varepsilon > 0$, $\delta = \delta(\varepsilon)$;
- 3) $\varphi_n''(t)$ равномерно по n непрерывны в точке $t = 0$. Именно, для заданного $\varepsilon > 0$ существует $\delta > 0$ такое, что при $|t| < \delta$

$$\sup_n |\varphi_n''(t) - \varphi_n''(0)| < \varepsilon, \quad 0 < c_1 < |\varphi_n''(0)| < c_2 < \infty.$$

Тогда справедливо утверждение теоремы 8.

Если в условии 1) $\varphi_n^m(t)$ заменить на $\varphi_n^m(t)/(1+|t|)$, то справедливо утверждение теоремы 9.

Приведенные условия таковы, что они позволяют получить доказательство теоремы 10, совершенно ничего не меняя в рассуждениях теорем 8, 9. Соотношение (27), для выполнения которого в условиях схемы серий требовалось бы условие Линдеберга, следует из (29).

Теорема 10 показывает, что переход к схеме серий в локальной теореме даже в случае одинаково распределенных величин сопровождается введением целого ряда новых условий.

§ 7. Закон больших чисел и центральная предельная теорема в многомерном случае

В этом параграфе мы будем предполагать, что $\xi_{1,n}, \dots, \xi_{n,n}$ есть случайные векторы в схеме серий, $E\xi_{k,n} = 0$, $\zeta_n = \sum_{k=1}^n \xi_{k,n}$.

Закон больших чисел $\zeta_n \xrightarrow{p} 0$ получается немедленно из теоремы 3, если предположить, что условиям теоремы 3 удовлетворяют координаты $\xi_{k,n}$. Поэтому ничто не мешает считать, что теорема 3 сформулирована и доказана для векторов.

Чуть сложнее обстоит дело с центральной предельной теоремой. Здесь мы будем предполагать, что $E|\xi_{k,n}|^2 < \infty$, где $|x|^2 = (x, x)$ есть квадрат нормы x . Обозначим

$$\sigma_{k,n}^2 = E\xi_{k,n}^T \xi_{k,n}, \quad \sigma_n^2 = \sum_{k=1}^n \sigma_{k,n}^2$$

(индекс T означает транспонирование, так что $\xi_{k,n}^T$ есть вектор-столбец).

Введем в рассмотрение условие

$$\sum_{k=1}^n \mathbb{E} \min(|\xi_{k,n}|^2, |\xi_{k,n}|^s) \rightarrow 0, \quad s > 2, \quad (D_2)$$

и условие Линдеберга

$$\sum_{k=1}^n \mathbb{E}(|\xi_{k,n}|^2; |\xi_{k,n}| > \tau) \rightarrow 0 \quad (M_2)$$

при $n \rightarrow \infty$ и любом $\tau > 0$. Как и в одномерном случае, легко убедиться, что условия (D_2) и (M_2) при $\text{tr} \sigma_n^2 = \sum_{j=1}^d (\sigma_n^2)_{jj} < c < \infty$ равносильны.

Теорема 11. Если $\sigma_n^2 \rightarrow \sigma^2$, где σ^2 — положительно определенная матрица, и выполнено условие (D_2) (или (M_2)), то

$$\zeta_n \Rightarrow \Phi_{0, \sigma^2}.$$

Следствие 4 («обычная» центральная предельная теорема). Если ξ_1, ξ_2, \dots — последовательность независимых одинаково распределенных векторов, $\mathbb{E} \xi_k = 0$, $\sigma^2 = \mathbb{E} \xi_k^T \xi_k$, $S_n = \sum_{k=1}^n \xi_k$, то при $n \rightarrow \infty$

$$\frac{S_n}{\sqrt{n}} \Rightarrow \Phi_{0, \sigma^2}.$$

Это утверждение является следствием теоремы 11, так как случайные величины $\xi_{k,n} = \xi_k / \sqrt{n}$ удовлетворяют ее условиям.

Доказательство теоремы 11. Рассмотрим характеристические функции

$$\varphi_{k,n}(t) = \mathbb{E} e^{i(t, \xi_{k,n})}, \quad \varphi_n(t) = \mathbb{E} e^{i(t, \zeta_n)} = \prod_{k=1}^n \varphi_{k,n}(t).$$

Для доказательства надо убедиться, что при любом t и при $n \rightarrow \infty$

$$\varphi_n(t) \rightarrow \exp \left\{ -\frac{1}{2} t \sigma^2 t^T \right\}.$$

Воспользуемся теоремой 5. Функции $\varphi_{k,n}(t)$ и $\varphi_n(t)$ можно рассматривать как характеристические функции

$$\varphi_{k,n}^\theta(v) = \mathbb{E} \exp(i v \xi_{k,n}^\theta), \quad \varphi_n^\theta(v) = \mathbb{E} \exp(i v \zeta_n^\theta)$$

случайных величин $\xi_{k,n}^\theta = (\xi_{k,n}, \theta)$, $\zeta_n^\theta = (\zeta_n, \theta)$, где $\theta = t/|t|$, $v = |t|$. Покажем, что скалярные случайные величины $\xi_{k,n}^\theta$ удовлетворяют условиям теоремы 5 (или следствия 1) для одномерного случая. Очевидно,

$$\begin{aligned} \mathbb{E} \xi_{k,n}^\theta &= 0, \\ \sum_{k=1}^n \mathbb{E} (\xi_{k,n}^\theta)^2 &= \sum_{k=1}^n \mathbb{E} (\xi_{k,n}, \theta)^2 = \theta \sigma_n^2 \theta^T \rightarrow \theta \sigma^2 \theta^T > 0. \end{aligned}$$

Выполнение условия (D₂) следует из очевидных неравенств

$$(\xi_{k,n}, \theta)^2 = |\xi_{k,n}^\theta|^2 \leq |\xi_{k,n}|^2, \\ \sum_{k=1}^n E g_2(\xi_{k,n}^\theta) \leq \sum_{k=1}^n E g_2(|\xi_{k,n}|),$$

где $g_2(x) = \min(x^2, |x|^s)$, $s > 2$. Таким образом, для любых v и θ (т.е. для любых t) в силу следствия 1 теоремы 5

$$\varphi_n(t) = E \exp\{i v \zeta_n^\theta\} \rightarrow \exp\left\{-\frac{1}{2} v^2 \theta \sigma^2 \theta^T\right\} = \exp\left\{-\frac{1}{2} t \sigma^2 t^T\right\}.$$

В теореме 11 остается вне поля зрения случай, когда элементы матрицы σ_n^2 неограниченно растут или ведут себя так, что ранг предельной матрицы σ^2 становится меньше размерности векторов $\xi_{k,n}$. Такое может происходить, когда дисперсии разных координат $\xi_{k,n}$ имеют разные порядки убывания (или роста). В этом случае вместо ζ_n следует рассматривать преобразованные суммы $\zeta'_n = \zeta_n \sigma_n^{-1}$. По существу теорема 11 является следствием следующего более общего утверждения, которое, в свою очередь, вытекает из теоремы 11.

Теорема 11А. Если случайные величины $\xi'_{k,n} = \xi_{k,n} \sigma_n^{-1}$ удовлетворяют условию (D₂) (или (M₂)), то $\zeta'_n \Leftrightarrow \Phi_{0,E}$, где E — единичная матрица.

§ 8*. Вероятности больших уклонений

Пусть ξ_1, ξ_2, \dots — последовательность независимых одинаково распределенных величин, $E\xi_k = 0$, $D\xi_k = 1$. Предположим, что надо оценить вероятность $P(S_n \geq x)$, когда x зависит от n и $x/\sqrt{n} \rightarrow \infty$ при $n \rightarrow \infty$. Центральная предельная теорема в этом случае позволяет лишь заключить, что $P(S_n \geq x) \rightarrow 0$. Задача состоит в отыскании функции $P(n, x)$ такой, что

$$P(S_n \geq x) = P(n, x)(1 + o(1)).$$

Отметим еще раз, что в отличие от нормального приближения здесь мы требуем малость *относительной*, а не *абсолютной* погрешности.

Такого рода задачи нередко возникают, например, в математической статистике при оценке потерь в статистических критериях и в теории надежности.

В этом параграфе будет изложен метод Крамера исследования вероятностей больших уклонений, который в известном смысле сводит задачу к центральной предельной теореме.

В дальнейшем будем предполагать, что выполнено следующее основное

Условие (А) (условие Крамера). Существует $\lambda > 0$ такое, что

$$\int e^{\lambda y} dF(y) = E e^{\lambda \xi} < \infty.$$

Смысл этого условия понять нетрудно — оно эквивалентно экспоненциальному убыванию $P(\xi > y) = 1 - F(y)$. Действительно, интеграл в (А) сходится или расходится вместе с интегралом

$$\int_0^\infty e^{\lambda y} dF(y) = - \int_0^\infty e^{\lambda y} d(1 - F(y)) = 1 - F(0) + \lambda \int_0^\infty e^{\lambda y} (1 - F(y)) dy.$$

Поэтому, если $1 - F(y) \leq ce^{-\mu y}$, то (A) выполнено при $\lambda < \mu$. Наоборот, из (A) следует по неравенству Чебышева, что

$$P(\xi > y) \leq e^{-\lambda y} E e^{\lambda \xi}.$$

Обозначим $\lambda_+ = \sup\{\lambda: E e^{\lambda \xi} < \infty\}$. Если, например, ξ ограничена сверху, то $\lambda_+ = \infty$. То же самое верно для нормально распределенной ξ . Если же $P(\xi > y) = ce^{-\mu y}$, $\mu > 0$, то, очевидно, $\lambda_+ = \mu$.

Если выполнено условие (A), то х. ф. $\varphi(t) = E e^{it\xi}$ — аналитическая в плоскости комплексного переменного в полосе $0 > \text{Im } t > -\lambda_+$. Это следует из возможности дифференцирования в плоскости комплексного переменного функции

$$\varphi(t) = \int e^{ity} dF(y),$$

поскольку $\int |ye^{ity}| dF(y)$ при указанных значениях t равномерно сходится. Если условие (A) выполняется и при некоторых отрицательных λ , то $\varphi(t)$ аналитическая в полосе, содержащей прямую $\text{Im } t = 0$.

Рассмотрим теперь подробнее функцию $\psi(\lambda) = E e^{\lambda \xi} = \varphi(-i\lambda)$. Очевидно, что (имеются в виду правые производные)

$$\psi'(0) = E\xi, \quad \psi''(0) = D\xi, \quad \psi''(\lambda) = \int y^2 e^{\lambda y} dF(y) > 0.$$

Нетрудно видеть, что $\psi''(\lambda)/\psi(\lambda)$ имеет смысл второго момента случайной величины η с функцией распределения

$$G(y) = \frac{1}{\psi(\lambda)} \int_{-\infty}^y e^{\lambda t} dF(t). \quad (31)$$

Дисперсия этой величины равна

$$\frac{\psi''}{\psi} - \left(\frac{\psi'}{\psi} \right)^2 = (\ln \psi)''.$$

Так как дисперсия обязана быть положительной, то отсюда следует, что функции ψ и $\ln \psi$ являются выпуклыми на $[0, \lambda_+]$. Функции

$$\psi_\alpha(\lambda) = e^{-\alpha \lambda} \psi(\lambda) = E e^{\lambda(\xi - \alpha)}, \quad \ln \psi_\alpha(\lambda) = -\alpha \lambda + \ln \psi(\lambda),$$

очевидно, также являются выпуклыми, при этом

$$\psi'_\alpha(0) = E(\xi - \alpha) = -\alpha.$$

Поэтому, если $\alpha \geq 0$ и $\psi'_\alpha(\lambda_+) > 0$ (мы полагаем $\psi'_\alpha(\lambda_+) = \lim_{\lambda \uparrow \lambda_+} \psi'_\alpha(\lambda)$), то на полуинтервале $[0, \lambda_+)$ существует единственная точка $\lambda(\alpha)$, в которой $\psi'_\alpha(\lambda) = 0$ и функции ψ_α и $\ln \psi_\alpha$ достигают своего минимума.

Так как $\psi'_\alpha(\lambda) = \psi_\alpha(\lambda)(-\alpha + \psi'(\lambda)/\psi(\lambda))$, то условия существования точки $\lambda(\alpha)$ можно записать в виде

$$0 \leq \alpha \leq \alpha_+ \equiv \frac{\psi'(\lambda_+)}{\psi(\lambda_+)}.$$

Ясно, что при достаточно малых $\alpha > 0$ эти условия всегда выполнены. Они выполнены при любом $\alpha \geq 0$, если $\psi'(\lambda_+) = \infty$.

Обозначим

$$\Lambda(\alpha) = \sup_{\lambda} (\alpha\lambda - \ln \psi(\lambda)) = \alpha\lambda(\alpha) - \ln \psi(\lambda(\alpha)). \quad (32)$$

Функцию $\Lambda(\alpha)$ назовем *функцией уклонений* (смысл такого названия станет ясным позже). Точка $\lambda(\alpha)$, где достигается значение $\Lambda(\alpha)$, удовлетворяет тождеству

$$\frac{\psi'(\lambda(\alpha))}{\psi(\lambda(\alpha))} = \alpha. \quad (33)$$

Так как $\ln \psi$ — выпуклая функция, то производная $(\psi'(\lambda)/\psi(\lambda))' = (\ln \psi)'' > 0$ и, значит, функция $\lambda(\alpha)$, являясь обратной к ψ'/ψ , будет аналитической в $(0, \lambda_+)$. Заметим, что $\lambda(0) = 0$, $\Lambda(0) = 0$ и что в силу (32), (33)

$$\Lambda'(\alpha) = \lambda(\alpha) + \alpha\lambda'(\alpha) - \frac{\psi'(\lambda(\alpha))}{\psi(\lambda(\alpha))} \lambda'(\alpha) = \lambda(\alpha),$$

получим $\Lambda(\alpha) = \int_0^\alpha \lambda(u) du$. Так как $\lambda'(0) = (D\xi)^{-1} = 1$, то отсюда следует, в частности,

что $\Lambda(0) = \Lambda'(0) = 0$, $\Lambda''(0) = 1$. Случайную величину $\xi^{(\alpha)}$ с распределением G в (31) нам будет удобно называть *преобразованием Крамера* над случайной величиной ξ . Характеристическая функция этой случайной величины равна

$$g(u) = \int e^{iuy} \frac{1}{\psi(\lambda)} e^{\lambda y} dF(y) = \frac{\varphi(u - i\lambda)}{\varphi(-i\lambda)}. \quad (34)$$

Функция распределения для преобразования Крамера над S_n в силу (31) будет иметь вид

$$\psi^{-n}(\lambda) \int_{-\infty}^y e^{\lambda t} dF_{S_n}(t), \quad (35)$$

а соответствующая х. ф. будет равна в силу (34)

$$\frac{\varphi_{S_n}(u - i\lambda)}{\psi^n(\lambda)} = \frac{\varphi^n(u - i\lambda)}{\psi^n(\lambda)}. \quad (36)$$

Покажем, что (35) есть не что иное, как функция распределения F_{Z_n} суммы $Z_n = \sum_{k=1}^n \xi_k^{(\alpha)}$, где $\xi_k^{(\alpha)} = \xi^{(\alpha)}$ и независимы. Другими словами, *преобразование Крамера над суммой есть сумма преобразований Крамера над слагаемыми*. Это следует из того, что х. ф. Z_n равна $\left[\frac{\varphi(u - i\lambda)}{\psi(\lambda)} \right]^n$ и совпадает с (36). Таким образом,

$$F_{Z_n}(y) = \psi^{-n}(\lambda) \int_{-\infty}^y e^{\lambda t} dF_{S_n}(t), \quad F_{S_n}(y) = \psi^n(\lambda) \int_{-\infty}^y e^{-\lambda t} dF_{Z_n}(t). \quad (37)$$

Если S_n имеет плотность $f_{S_n}(y)$, то Z_n также имеет плотность

$$f_{Z_n}(y) = \psi^{-n}(\lambda) e^{\lambda y} f_{S_n}(y). \quad (38)$$

Пусть $0 < \alpha < \alpha_+$. Выберем в формулах (31), (34) $\lambda = \lambda(\alpha)$. Тогда случайная величина $\xi^{(\alpha)}$ с характеристической функцией $g(u)$ будет обладать свойствами

$$\begin{aligned} E(\xi^{(\alpha)} - \alpha) &= \frac{\psi'(\lambda(\alpha))}{\psi(\lambda(\alpha))} - \alpha = 0, \\ D(\xi^{(\alpha)} - \alpha) &= \frac{\psi''(\lambda(\alpha))}{\psi(\lambda(\alpha))} - \alpha^2 \equiv \sigma_\alpha^2 > 0. \end{aligned} \quad (39)$$

Можно сформулировать теперь следующую теорему. Обозначим

$$H_n = Z_n - \alpha n = \sum_{k=1}^n (\xi_k^{(\alpha)} - \alpha).$$

Теорема 12. Пусть выполнено условие (A) и

$$\limsup_{n \rightarrow \infty} \frac{x}{n} < \alpha_+ = \frac{\psi'(\lambda_+)}{\psi(\lambda_+)}$$

(напомним, что $\psi(\lambda) = E e^{\lambda \xi}$, $\lambda_+ = \sup\{\lambda: \psi(\lambda) < \infty\}$). Тогда

$$F_{S_n}(x+a) - F_{S_n}(x) = \exp\{-n\Lambda(\alpha)\} \int_0^a e^{-\lambda(\alpha)t} dF_{H_n}(t), \quad (40)$$

где $\alpha = x/n$.

Если, кроме того, ξ имеет плотность, то

$$f_{S_n}(x+a) = \exp\{-n\Lambda(\alpha) - a\lambda(\alpha)\} f_{H_n}(a). \quad (41)$$

Эта теорема позволяет сводить задачи о больших отклонениях для S_n к задаче о «нормальных» отклонениях для H_n .

Доказательство следует немедленно из формул (37), (38), если положить в них $y = x + a = \alpha n + a$, $\lambda = \lambda(\alpha)$ и заметить, что

$$F_{Z_n}(x+a) = F_{Z_n - \alpha n}(a) = F_{H_n}(a). \quad \blacksquare$$

Формулы (40), (41) позволяют вычислять $F_{S_n}(x+a)$ и $f_{S_n}(x+a)$, поскольку распределение суммы $H_n = \sum_{k=1}^n (\xi_k^{(\alpha)} - \alpha)$ в области «нормальных отклонений» известно. Мы знаем, например, что в широких предположениях в формуле (41) $f_{H_n}(a) \sim 1/(\sigma_\alpha \sqrt{n}) \cdot 1/\sqrt{2\pi}$. Для большей точности воспользуемся локальной теоремой 10, где вместо $\varphi_n(t)$ нужно рассмотреть х.ф. $\xi_1^{(\alpha)} - \alpha$, равную

$$\gamma(t) = e^{-it\alpha} \frac{\varphi(t - i\lambda(\alpha))}{\varphi(-i\lambda(\alpha))}$$

(ясно, что мы находимся в условиях схемы серий, так как распределение $\xi_1^{(\alpha)} - \alpha$ зависит от n через $\alpha = x/n$).

Введем в рассмотрение

Условие (В). $F(x)$ имеет плотность, а х. ф. $\varphi(t)$ такова, что $\varphi^m(u - i\lambda)$ при каком-нибудь целом $m \geq 1$ интегрируема по u при всех $\lambda \in [0, \lambda_+)$ ⁷⁾.

Нетрудно видеть, что при выполнении этого условия выполняются условия 1–3 теоремы 10, при этом равномерно по α . Равномерность по α можно проверять непосредственно. Допустим, например, относительно условия 2 противное: что существует $\varepsilon > 0$ и последовательности α_k, t_k такие, что $|t_k| > \varepsilon$ и

$$|\gamma(t_k)| = \left| \frac{\varphi(t_k - i\lambda(\alpha_k))}{\varphi(-i\lambda(\alpha_k))} \right| \rightarrow 1 \quad (42)$$

при $k \rightarrow \infty$. Так как $\alpha_k \in [0, a]$, $a < \alpha_+$, то можно, пользуясь компактностью отрезка, не ограничивая общности, считать, что $\alpha_k \rightarrow \alpha_0 \in [0, a]$, $a < \infty$. Далее, производная $\varphi'(t)$ равномерно ограничена в области $0 \geq \operatorname{Im} t \geq -\lambda(a)$, поскольку $|\varphi'(t)| \leq \psi'(\lambda(a))$. Поэтому вместе с (42) при $k \rightarrow \infty$ будет выполняться

$$|t_k| > \varepsilon, \quad \varphi(-i\lambda(\alpha_k)) \rightarrow \varphi(-i\lambda(\alpha_0)) > 0, \quad \left| \frac{\varphi(t_k - i\lambda(\alpha_0))}{\varphi(-i\lambda(\alpha_0))} \right| \rightarrow 1.$$

Это противоречит существованию плотности у распределения с характеристической функцией $\frac{\varphi(t - i\lambda(\alpha_0))}{\varphi(-i\lambda(\alpha_0))}$. ■

Таким образом, условия теоремы 10 выполнены. Это означает, что

$$f_{H_n}(a) = \frac{1}{\sigma_\alpha \sqrt{n}} f_{\zeta_n^*} \left(\frac{a}{\sigma_\alpha \sqrt{n}} \right), \quad f_{\zeta_n^*}(v) \rightarrow \frac{1}{\sqrt{2\pi}} e^{-v^2/2} \quad (43)$$

равномерно по v , где $\zeta_n^* = H_n/(\sigma_\alpha \sqrt{n})$. Сопоставляя это с теоремой 12, получаем следующее утверждение.

Теорема 13. Пусть выполнены условия теоремы 12 и условие (В). Тогда

$$f_{S_n}(x) \sim \frac{1}{\sigma_\alpha \sqrt{2\pi n}} \exp\{-n\Lambda(\alpha)\},$$

где по-прежнему $\alpha = x/n$, σ_α определяются формулой (39).

При любом $a > 0$ таком, что $a = o(\sqrt{n})$,

$$P(x \leq S_n \leq x + a) \sim \frac{1}{\sigma_\alpha \sqrt{2\pi n}} \exp\{-n\Lambda(\alpha)\} \frac{1 - e^{-a\lambda(\alpha)}}{\lambda(\alpha)}. \quad (44)$$

Если $x/\sqrt{n} \rightarrow \infty$, то

$$P(S_n \geq x) \sim \frac{1}{\sigma_\alpha \lambda(\alpha) \sqrt{2\pi n}} \exp\{-n\Lambda(\alpha)\}. \quad (45)$$

Доказательство. Соотношение (44) получается интегрированием (41), (43) по a .

Чтобы получить (45), надо выбрать $a = o(\sqrt{n})$ так, чтобы $a\lambda(\alpha) \rightarrow \infty$ (или, что то же, $ax/n \rightarrow \infty$). Это, очевидно, всегда возможно, и это делает правую часть в (44) асимптотически эквивалентной (45). ■

В соотношениях (44), (45) следует иметь в виду, что $\sigma_\alpha \rightarrow 1$, $\lambda(\alpha) \sim \alpha$ при $\alpha \rightarrow 0$.

⁷⁾ Для этого достаточно, чтобы интегрируема была $\varphi^m(t)$.

Замечание 7. Для получения соотношений (44), (45) нет необходимости пользоваться локальной предельной теоремой 10, для которой введено условие (B). Достаточно воспользоваться формулой (40) и теоремой 9, для которой, как мы знаем, в условии (B) интегрируемость $\varphi^m(t)$ можно ослабить до интегрируемости $\varphi^m(t)/(1+|t|)$.

Соотношение (40) при $a = \infty$ допускает следующую интерпретацию, связанную с тем, что интеграл в правой части представляет собой в этом случае свертку двух распределений. Пусть $\gamma \in \Gamma_{\lambda(\alpha)}$ (т. е. $P(\gamma > x) = e^{-\lambda(\alpha)x}$, $x \geq 0$) и не зависит от H_n . Тогда (40) можно записать в виде

$$\begin{aligned} P(S_n > x) &= \exp\{-n\Lambda(\alpha)\} [P(H_n > 0) - P(H_n > \gamma)] = \\ &= \exp\{-n\Lambda(\alpha)\} \lambda(\alpha) \int_0^\infty e^{-\lambda(\alpha)z} P(0 < H_n < z) dz. \end{aligned}$$

По центральной предельной теореме отсюда следует лишь, что при $\alpha > \alpha_0 > 0$

$$P(S_n > x) = o(e^{-n\Lambda(\alpha)}).$$

Замечание 8. Соотношение (45) показывает, насколько точным является неравенство типа Чебышева

$$P(\xi \geq x) \leq e^{-\lambda x} E e^{\lambda \xi},$$

полученное в §4.7. Применяя его к случайной величине S_n и выбирая $\lambda = \lambda(\alpha)$, мы получим $E e^{\lambda S_n} = \psi(\lambda)^n$,

$$P(S_n \geq x) \leq \exp n\{-\alpha\lambda(\alpha) + \ln \psi(\lambda(\alpha))\} = \exp\{-n\Lambda(\alpha)\}.$$

Замечание 9. Теоремы 12, 13 показывают, что функция $\Lambda(\alpha)$ играет определяющую роль при описании вероятностей больших отклонений. Соотношения (45) позволяют записать для функции $\Lambda(\alpha)$ при каждом α , $0 \leq \alpha < \alpha_+$, равенство

$$\Lambda(\alpha) = -\lim_{n \rightarrow \infty} \frac{1}{n} \ln P\left(\frac{S_n}{n} \geq \alpha\right). \quad (46)$$

Легко видеть также, что

$$\Lambda(\alpha) = -\lim_{h \rightarrow 0} \lim_{n \rightarrow \infty} \frac{1}{n} \ln P\left(\frac{S_n}{n} \in (\alpha - h, \alpha + h)\right). \quad (47)$$

Это последнее соотношение оказывается верным уже без каких бы то ни было ограничений на распределение ξ_k и α .

Равенства (46), (47) объясняют в какой-то мере термин «функция отклонений» для $\Lambda(\alpha)$, которую следует считать определенной при всех α с помощью первого соотношения в (32). Можно отметить также, что если условие (A) не выполнено, то $\Lambda(\alpha) = 0$ при $\alpha > 0$.

Замечание 10. Возвращаясь к теореме 9, видим, что если $x \geq \alpha_0 n$, $\alpha_0 > 0$, то убывание $P(S_n \geq x)$ оказывается экспоненциальным. Если функция $\psi(\lambda)$ аналитична в точке $\lambda = 0$, то аналитической будет и $\Lambda(\alpha)$ в точке $\alpha = 0$, и можно воспользоваться разложением $\Lambda(\alpha) = \sum \Lambda_k \alpha^k$, где Λ_k полностью определяются k моментами ξ . В этом случае при $x/n = \alpha \rightarrow 0$ получим

$$\exp\{-n\Lambda(\alpha)\} \sim \exp\left\{-n \sum_{k=2}^s \Lambda_k \alpha^k\right\},$$

если $n\alpha^{s+1} = x^{s+1}/n^s \rightarrow 0$ при $n \rightarrow \infty$. В частности, если $x = o(n^{2/3})$, то

$$P(S_n \geq x) \sim \left(2\pi \frac{x^2}{n}\right)^{-1/2} \exp\left\{-\frac{x^2}{2n}\right\} \sim 1 - \Phi\left(\frac{x}{\sqrt{n}}\right). \quad (48)$$

Можно отметить, что последний результат согласуется с замечаниями относительно применимости теоремы Муавра—Лапласа в области больших уклонений. Если же окажется, что $\Lambda_3 \neq 0$, то пользоваться нормальным приближением, когда $\Lambda_3 x^3/n^2$ не является малым числом, нельзя.

Функция уклонений $\Lambda(\alpha)$ для некоторых распределений может быть вычислена в явном виде. Например, для нормального закона с параметрами $(0, 1)$

$$\lambda_+ = \alpha_+ = \infty, \quad \lambda(\alpha) = \alpha, \quad \Lambda(\alpha) = \alpha^2/2.$$

Для центрированного показательного распределения с параметром μ

$$P(\xi > x) = e^{-\mu(x+1/\mu)} \text{ при } x \geq \frac{-1}{\mu}, \quad P(\xi > x) = 1 \text{ при } x < \frac{-1}{\mu},$$

имеем

$$\lambda_+ = \mu, \quad \alpha_+ = \infty, \quad \psi(\lambda) = e^{-\lambda/\mu} \frac{\mu}{\mu - \lambda}, \quad \lambda(\alpha) = \mu - \frac{\mu}{\alpha\mu + 1} = \frac{\alpha\mu^2}{\alpha\mu + 1}.$$

Замечание 11. Для схемы Бернулли функция уклонений по существу нами была вычислена и использовалась в § 5.2. В связи с этим можно отметить, что формулы (45) сохраняются и для решетчатых ξ_k (ξ_k имеют вид kh , $h > 0$, k целое), однако при этом x должно быть кратным шагу решетки.

Из обобщений теоремы 13, которые доступны читателю без преодоления каких-либо новых трудностей, можно указать следующее. Вспомним, для чего требовалось условие (В) — для возможности использовать формулу обращения (интегрируемость φ^m) и для оценки $|\gamma(u)| < 1 - \delta$ при $|u| > \varepsilon$ (см. условия теоремы 10). Но всего этого можно добиться и несколько иным путем, введя в S_n «сглаживающее» слагаемое ξ^* с финитной плотностью (гладкой и сосредоточенной на $[0, h]$) и предположив, что выполнено

Условие (С) (это условие иногда также называют условием Крамера).

$$\lim_{|u| \rightarrow \infty} \sup |\varphi(u - i\lambda)| < \varphi(-i\lambda), \quad \lambda \in [0, \lambda_+).$$

Введение ξ^* будет обеспечивать существование плотности $S_n + \xi^*$ и возможность пользоваться формулой обращения, поскольку х.ф. ξ^* интегрируема (см. доказательство теоремы непрерывности). Из условия (С) будет следовать, что $|\gamma(u)| < 1 - \delta$ при $|u| > \varepsilon$ ($\text{Im } u = 0$).

Для выполнения условия (С) достаточно⁸⁾, чтобы $F(x)$ содержала абсолютно непрерывную компоненту (ведь х.ф., соответствующая этой компоненте, сходится к нулю при $|u| \rightarrow \infty$, $\text{Im } u = 0$). Условие (С) заведомо не будет выполнено, если ξ дискретна — тогда $|\varphi(t)|$ будет принимать при $t \rightarrow \infty$ значения, сколь угодно близкие (или равные) единице.

⁸⁾ Условие (С) будет выполнено, если $\lim_{|u| \rightarrow \infty} \sup |\varphi(u)| < 1$ ($\text{Im } u = 0$).

Таким образом, если случайная величина ξ^* не зависит от $\{\xi_k\}$ и имеет финитную плотность, сосредоточенную на $[0, h]$, то ее х.ф. $\varphi^*(t)$ интегрируема, аналитична во всей плоскости, и плотность $S_n + \xi^*$ равна

$$f_{S_n + \xi^*}(x + a) = \frac{1}{2\pi} \int e^{-itx} \varphi^n(t) \varphi^*(t) dt.$$

Повторяя с очевидными изменениями проведенные выкладки, получим

$$f_{S_n + \xi^*}(x + a) = \exp\{-n\Lambda(\alpha) - a\lambda(\alpha)\} f_{H_n + \xi^{**}}(a) \varphi^*(-i\lambda(\alpha)),$$

где ξ^{**} также имеет финитную плотность и х.ф., равную $\frac{\varphi^*(u - i\lambda(\alpha))}{\varphi^*(-i\lambda(\alpha))}$. Теперь надо вновь воспользоваться тем, что при $a = o(\sqrt{n})$ равномерно по a

$$f_{H_n + \xi^{**}}(a) = \frac{1}{\sigma_\alpha \sqrt{2\pi n}} (1 + o(1)).$$

Мы получим тогда так же, как в (44), что при $a > 0$, $a = o(\sqrt{n})$

$$P(x < S_n + \xi^* < x + a) \sim \frac{e^{-n\Lambda(\alpha)}}{\sigma_\alpha \lambda(\alpha) \sqrt{2\pi n}} (1 - e^{-a\lambda(\alpha)}) \varphi^*(-i\lambda(\alpha)). \quad (49)$$

Отсюда уже нетрудно получить (44) и (45), пользуясь неравенствами

$$P(x + h < S_n + \xi^* < x + a) \leq P(x \leq S_n < x + a) \leq P(x \leq S_n + \xi^* < x + a + h)$$

и тем, что величина h может быть выбрана произвольно малой. Надо считать при этом, что $a \geq a_0 = \text{const} > 0$. Величина $\varphi^*(-i\lambda(\alpha))$ в (49) выбором h может быть сделана сколь угодно близкой единице.

Мы доказали таким образом, что верна

Теорема 14. Если выполнены условия (A) и (C), то при дополнительном условии $a \geq a_0 > 0$ справедливы соотношения (44), (45).

В заключение этого параграфа отметим еще одно свойство функции уклонений. Будем обозначать через $\Lambda_\xi(\alpha)$, $\lambda_\xi(\alpha)$, $\psi_\xi(\lambda)$ функции Λ , λ и ψ , определенные выше и соответствующие случайной величине ξ . Тогда если ξ и η независимы, то

$$\begin{aligned} \Lambda_{\xi+\eta}(\alpha) &= \sup_{\lambda} (\alpha\lambda - \ln \psi_\xi(\lambda) - \ln \psi_\eta(\lambda)) = \\ &= \sup_{\lambda} (\beta\lambda - \ln \psi_\xi(\lambda) + (\alpha - \beta)\lambda - \ln \psi_\eta(\lambda)) \leq \Lambda_\xi(\beta) + \Lambda_\eta(\alpha - \beta). \end{aligned} \quad (50)$$

С другой стороны, так как $\lambda_\xi(\beta)$ есть монотонно возрастающая функция, то можно подобрать β так, чтобы $\lambda_\xi(\beta) = \lambda_\eta(\alpha - \beta)$. При таком β в (50) будет знак равенства, и, следовательно,

$$\Lambda_{\xi+\eta}(\alpha) = \inf_{\beta} (\Lambda_\xi(\beta) + \Lambda_\eta(\alpha - \beta)).$$

Из соображений симметрии для одинаково распределенных ξ_k отсюда находим, что

$$\Lambda_{\xi_1 + \xi_2}(2\alpha) = 2\Lambda_{\xi_1}(\alpha), \quad \Lambda_{S_n}(n\alpha) = n\Lambda_\xi(\alpha).$$

Замечая, что в формуле (45) порядок стремления к бесконечности у последовательностей $n\lambda^2(\alpha)$ и $n\Lambda(\alpha)$ одинаков, мы могли бы эту формулу написать теперь в виде

$$P(S_n \geq x) \sim c(\alpha) (\Lambda_{S_n}(x))^{-1/2} \exp\{-\Lambda_{S_n}(x)\},$$

где $c(\alpha)$ заключено в фиксированных положительных пределах. В таком виде теорема о больших отклонениях становится весьма общей и сохраняющей свою силу при отклонениях от условий теоремы 14, когда допускается зависимость между слагаемыми и их разнораспределенность.

§ 9*. Сходимость к другим устойчивым законам

Как это видно из предыдущих разделов, нормальный закон занимает особое место среди других распределений — он выступает в роли предельного для нормированных сумм произвольно распределенных слагаемых. Возникает естественный вопрос: существуют ли другие предельные законы для сумм независимых величин?

Из доказательства теоремы 2 для одинаково распределенных случайных величин видно, что характер предельного распределения S_n определяется поведением х. ф. в окрестности нуля. Если существуют $E\xi = 0$ и $E\xi^2 = \sigma^2$, то

$$\varphi\left(\frac{t}{\sqrt{n}}\right) = 1 + \frac{\varphi''(0)t^2}{2n} + o\left(\frac{1}{n}\right).$$

А это определяет асимптотическое поведение х. ф. S_n/\sqrt{n} , равной $\varphi^n(t/\sqrt{n})$, которое приводит к нормальному закону. Следовательно, в поисках других предельных законов для сумм $S_n = \xi_1 + \dots + \xi_n$ надо отказаться от конечности дисперсии $E\xi^2$, или, что то же, от существования $\varphi''(0)$. Мы не будем здесь рассматривать все распределения, обладающие этими свойствами, а ограничимся рассмотрением лишь достаточно «правильных» случаев, когда при $\alpha < 2$ выполнено

Условие (D).

$$\begin{aligned} 1 - F(x) &= ax^{-\alpha} + o(x^{-\alpha}), \\ F(-x) &= bx^{-\alpha} + o(x^{-\alpha}) \end{aligned}$$

при $x \rightarrow \infty$.

Здесь предполагается, что по крайней мере одно из чисел a и b положительно. Нетрудно видеть, что для таких распределений $\int x^2 dF(x) = \infty$. Величины $E\xi$, если $\alpha > 1$, не ограничивая общности, будем считать равными нулю.

Ответ на поставленный вопрос дает следующая

Теорема 15. При $0 < \alpha < 2$ существуют последовательности A_n такие, что при $n \rightarrow \infty$

$$\frac{S_n}{|\alpha(a+b)n|^{1/\alpha}} - A_n \Leftrightarrow F_{\alpha,\beta}, \quad (51)$$

где $F_{\alpha,\beta}$ есть закон распределения, зависящий лишь от параметров α и $\beta = (a-b)/(a+b)$.

Логарифм х. ф. закона $F_{\alpha,\beta}$ равен

$$\begin{aligned} \ln \varphi(t) &= \int \left(e^{itz} - 1 - \frac{itz}{1+x^2} \right) |x|^{-\alpha-1} D(x) dx, \\ D(x) &= \begin{cases} \frac{a}{a+b} & \text{при } x > 0, \\ \frac{b}{a+b} & \text{при } x \leq 0. \end{cases} \end{aligned} \quad (52)$$

Роль параметров α и β поясняется в замечании 13, помещенном в конце параграфа.

Помимо $F_{\alpha,\beta}$, $0 < \alpha < 2$, $-1 \leq \beta \leq 1$, предельными, как мы видели, являются нормальные законы Φ_{α,σ^2} и вырожденные I_α . Множество всех этих законов обозначим через \mathcal{L}_0 . Далее, нетрудно видеть, что если F — какой-нибудь предельный закон распределения, то закон, отличающийся от F изменением масштаба и сдвигом (т.е. закон $F((x - a_1)/b_1)$), также является предельным.

Оказывается, что класс законов распределения \mathcal{L} , полученный таким расширением \mathcal{L}_0 , исчерпывает все предельные законы для сумм одинаково распределенных независимых величин.

Возможна и другая характеристика класса предельных законов \mathcal{L} .

Определение 1. Мы назовем закон распределения $F(x)$ *устойчивым*, если для любых $a_1, a_2, b_1 > 0, b_2 > 0$ найдутся a и $b > 0$ такие, что

$$F\left(\frac{x - a_1}{b_1}\right) * F\left(\frac{x - a_2}{b_2}\right) = F\left(\frac{x - a}{b}\right).$$

В терминах х. ф. свойство устойчивости можно сформулировать так: для любых $b_1 > 0, b_2 > 0$ найдутся a и $b > 0$ такие, что

$$\varphi(tb_1)\varphi(tb_2) = e^{ita}\varphi(tb).$$

Класс всех устойчивых законов обозначим через \mathcal{L}^* . Замечательный факт состоит в том, что класс всех предельных законов \mathcal{L} и класс всех устойчивых законов \mathcal{L}^* совпадают.

Принадлежность $\mathcal{L} \subset \mathcal{L}^*$ следует из того, что законы Φ_{α,σ^2} , I_α и $F_{\alpha,\beta}$ являются устойчивыми (устойчивость $F_{\alpha,\beta}$ проще доказывать, пользуясь замечаниями 12, 13).

Любопытно отметить, что закон Пуассона (напомним, что его х. ф. $\varphi(t)$ равна $\exp\{\mu(e^{it} - 1)\}$) и гамма-распределение ($\varphi(t) = (1 - it/\alpha)^{-\lambda}$, см. § 7.7) не являются устойчивыми в смысле определения 1. Однако они обладают другим свойством, называемым *безграничной делимостью*. Это свойство состоит в том, что функция $\varphi^{1/n}(t)$ при любом n снова должна быть х. ф. (т.е. величина ξ с х. ф. φ представима как сумма независимых $\xi_{k,n}$, $k = 1, \dots, n$, с х. ф. $\varphi^{1/n}$). Класс безгранично делимых законов содержит в себе \mathcal{L}^* и играет ту же роль, что и класс \mathcal{L} : он совпадает с классом всех предельных законов для сумм независимых одинаково распределенных величин, но в *схеме серий*⁹⁾.

Доказательство теоремы 15. Общая схема доказательства здесь примерно та же, что и в центральной предельной теореме. Обозначим $B_n = [\alpha n(a + b)]^{1/\alpha}$, $\zeta_n = S_n/B_n - A_n$, где A_n выберем позже. Характеристическая функция ζ_n равна

$$\varphi_{\zeta_n}(t) = e^{-itA_n} \varphi_\xi^n\left(\frac{t}{B_n}\right).$$

Так как $\varphi_\xi(t) \rightarrow 1$ при $t \rightarrow 0$, то при каждом фиксированном t и $n \rightarrow \infty$

$$\left| \ln \varphi_\xi\left(\frac{t}{B_n}\right) - \left(\varphi_\xi\left(\frac{t}{B_n}\right) - 1\right) \right| < \left| \varphi_\xi\left(\frac{t}{B_n}\right) - 1 \right|^2,$$

$$\ln \varphi_{\zeta_n}(t) = -itA_n + n \left(\varphi_\xi\left(\frac{t}{B_n}\right) - 1 \right) + R_n.$$
(53)

⁹⁾ Подробнее об этом см. [9, 3].

Последнее слагаемое R_n по модулю не превосходит $n \left| \varphi_\xi \left(\frac{t}{B_n} \right) - 1 \right|^2$ и сходится к нулю при $n \rightarrow \infty$. Действительно, из неравенств

$$|e^{i\tau} - 1| \leq |\tau|, \quad |e^{i\tau} - 1| \leq 2$$

вытекает, что при $p < 1$

$$|e^{i\tau} - 1| < 2|\tau|^p.$$

Так как $m = E|\xi|^p < \infty$ при $p < \alpha$, то

$$|\varphi_\xi(\tau) - 1| \leq E|e^{i\tau\xi} - 1| \leq 2E|\tau\xi|^p,$$

$$n \left| \varphi_\xi \left(\frac{t}{B_n} \right) - 1 \right|^2 \leq n \left| \frac{t}{B_n} \right|^{2p} 4m^2 \leq \text{const} \cdot n^{1-2p/\alpha} \rightarrow 0$$

при $2p > \alpha$.

Рассмотрим теперь второе слагаемое в правой части (53). Его можно записать в виде

$$\begin{aligned} n \int (e^{itx} - 1) dF(xB_n) &= \\ &= \int \left(e^{itx} - 1 - \frac{itx}{1+x^2} \right) \frac{1+x^2}{x^2} dG_n(x) + itn \int \frac{x}{1+x^2} dF(xB_n), \end{aligned} \quad (54)$$

$$\text{где }^{10)} G_n(v) = \int_{-\infty}^v \frac{nx^2}{1+x^2} dF(xB_n).$$

Введем в рассмотрение функцию

$$G(v) = \int_{-\infty}^v \frac{|x|^{1-\alpha} D(x)}{1+x^2} dx,$$

где функция $D(x)$ определена в (52). Интегрируя по частям, находим при $v > 0$

$$\left(d \left(\frac{x^2}{1+x^2} \right) = \frac{2x dx}{(1+x^2)^2} \right)$$

$$\begin{aligned} G_n(v) - G_n(0) &= n \int_0^v \frac{x^2}{1+x^2} dF(xB_n) = \\ &= -n \frac{x^2}{1+x^2} (1 - F(xB_n)) \Big|_0^v + 2n \int_0^v (1 - F(xB_n)) \frac{x dx}{(1+x^2)^2}. \end{aligned} \quad (55)$$

В силу (D) получаем

$$\begin{aligned} \lim_{n \rightarrow \infty} (G_n(v) - G_n(0)) &= \\ &= -\frac{v^2}{1+v^2} \lim_{n \rightarrow \infty} na(vB_n)^{-\alpha} + 2 \int_0^v \lim_{n \rightarrow \infty} na(xB_n)^{-\alpha} \frac{x}{(1+x^2)^2} dx = \\ &= -\frac{v^{2-\alpha}}{1+v^2} \cdot \frac{a}{\alpha(a+b)} + 2 \frac{a}{\alpha(a+b)} \int_0^v \frac{x^{1-\alpha}}{(1+x^2)^2} dx. \end{aligned}$$

¹⁰⁾ Если $\alpha > 1$ и, следовательно, существует $E\xi = 0$, то удобнее вместо правой части в (54) записать $\int (e^{itx} - 1 - itx) \frac{1+x^2}{x^2} dG_n(x)$. В общем случае мы добавляем и вычитаем $\int \frac{x}{1+x^2} dF(xB_n)$, поскольку этот интеграл всегда существует.

Законность предельного перехода под знаком интеграла (включая случай $v = \infty$) здесь проверяется без труда, и мы предоставляем это читателю¹¹⁾.

Производя «обратное» интегрирование по частям, получим

$$\lim_{n \rightarrow \infty} (G_n(v) - G_n(0)) = \frac{a}{a+b} \int_0^v \frac{x^2}{1+x^2} x^{-\alpha-1} dx = G(v) - G(0).$$

Совершенно аналогично устанавливаем, что при $v < 0$

$$G_n(0) - G_n(v) = n \int_v^0 \frac{x^2}{1+x^2} dF(xB_n) \rightarrow \frac{b}{a+b} \int_v^0 \frac{x^2}{1+x^2} |x|^{-\alpha-1} dx = G(0) - G(v).$$

Таким образом, последовательность $G_n(v)/G_n(\infty)$ образует последовательность функций распределения

$$G_n(\infty) \rightarrow G(\infty), \quad \frac{G_n(v)}{G_n(\infty)} \Rightarrow \frac{G(v)}{G(\infty)}. \quad (56)$$

Вернемся теперь к (54). Первый интеграл в правой части (54) равен

$$\int f(x) dG_n(x) = G_n(\infty) \int f(x) d\frac{G_n(x)}{G_n(\infty)},$$

где $f(x) = (e^{itx} - 1 - itx/(1+x^2))(1+x^2)/(x^2)$ есть при каждом t непрерывная и ограниченная функция от x . Ограниченность и непрерывность ее при $x \neq 0$ очевидна. В окрестности точки $x = 0$, пользуясь разложением e^{itx} в ряд, получаем, что $f(x)$ ведет себя как

$$\frac{1}{x^2} \left[\left(itx - \frac{t^2 x^2}{2} + O(x^3) \right) (1+x^2) - itx \right] = -\frac{t^2}{2} + O(x).$$

На основании (56) и свойств слабой сходимости заключаем, что

$$\int f(x) dG_n(a) \rightarrow \int f(x) dG(x) = \int \left(e^{itx} - 1 - \frac{itx}{1+x^2} \right) |x|^{-\alpha-1} D(x) dx.$$

Если теперь положить $A_n = n \int \frac{x}{1+x^2} dF(xB_n)$, то получим, что существует

$$\lim_{n \rightarrow \infty} \ln \varphi_{\zeta_n}(t) = \int \left(e^{itx} - 1 - \frac{itx}{1+x^2} \right) |x|^{-\alpha-1} D(x) dx \equiv \psi(t). \quad (57)$$

Функция ψ , как нетрудно видеть, обладает свойством $\psi(t) \rightarrow 0$ при $t \rightarrow 0$ и, следовательно, непрерывна в точке $t = 0$ (ср. с доказательством непрерывности х. ф.). По теореме непрерывности (теорема 7.3) это означает, что $\varphi = e^\psi$ является х. ф. Сходимость (51) доказана.

¹¹⁾ Если в правой части (55) отбросить $\int_0^{v_n}$, где v_n таково, что $B_n v_n \rightarrow \infty$, $v_n \sqrt{n} \rightarrow 0$, то $n \int_0^{v_n} \rightarrow 0$,

а подынтегральная функция в $\int_{v_n}^{\infty}$ мажорируется интегрируемой функцией.

Тот факт, что $F_{\alpha,\beta}$ зависит только от α и β , проверяется непосредственно. Действительно, если обозначить

$$\int_0^{\infty} \left(e^{itz} - 1 - \frac{itz}{1+x^2} \right) |x|^{-\alpha-1} dx = u + iv \quad (u, v \text{ вещественны}),$$

то $\int_{-\infty}^0 \left(e^{itz} - 1 - \frac{itz}{1+x^2} \right) |x|^{-\alpha-1} dx = u - iv$, и мы получим (см. (52))

$$\ln \varphi(t) = \frac{a}{a+b}(u+iv) + \frac{b}{a+b}(u-iv) = u + i\beta v.$$

Числа u и v в этих формулах можно найти в явном виде.

Замечание 12. Формулировку теоремы можно несколько упростить, рассматривая отдельно случаи $\alpha < 1$, $\alpha = 1$, $\alpha > 1$. В первом случае нетрудно показать, что последний интеграл в (54)

$$nit \int \frac{x}{1+x^2} dF(xB_n) = it \int \frac{1}{x} dG_n(x) \rightarrow it \int \frac{1}{x} dG(x)$$

(надо воспользоваться интегрированием по частям, подобно (55)). Одновременно можно заметить, что при $\alpha < 1$ существует часть интеграла (57), равная

$$\int \frac{itz}{1+x^2} |x|^{-\alpha-1} D(x) dx = it \int \frac{1}{x} dG(x).$$

Поэтому при $\alpha < 1$ можно положить $A_n = 0$, а утверждение теоремы сформулировать так:

$$\frac{S_n}{B_n} \Rightarrow F_{\alpha,\beta}^*, \quad (58)$$

где $F_{\alpha,\beta}^*$ есть устойчивый закон, логарифм х. ф. которого равен

$$\ln \varphi^*(t) = \int (e^{itz} - 1) |x|^{-\alpha-1} D(x) dx = |t|^\alpha \int (e^{iy} - 1) |y|^{-\alpha-1} D(y \operatorname{sign} t) dy. \quad (59)$$

Последнее равенство здесь есть результат замены переменных $tx = y$.

Если $\alpha > 1$, то $E\xi = 0$ и, как уже отмечалось в сноске 10) на с. 189, мы можем положить $A_n = 0$. Логарифм х. ф. распределения $F_{\alpha,\beta}^*$ в (58) будет при этом характеризоваться равенством

$$\ln \varphi^*(t) = \int (e^{itz} - 1 - itx) |x|^{-\alpha-1} D(x) dx. \quad (60)$$

Это выражение интегрированием по частям можно привести к виду, подобному (59).

В случае $\alpha = 1$ последовательность A_n устранить не удастся, но можно найти асимптотическое поведение A_n и показать, что

$$A_n = \beta \ln n + o(\ln n).$$

Будем считать для простоты, что $a + b = 1$. Тогда при $\alpha = 1$, $B_n = n$

$$A_n = n \int \frac{x}{1+x^2} dF(xn).$$

Рассмотрим сначала \int_0^∞ . Ясно, что при достаточно малом $\varepsilon > 0$ значение $n \int_0^\varepsilon \frac{x}{1+x^2} dF(xn)$ будет близко к $n \int_0^\varepsilon x dF(xn)$. Можно считать, кроме того, что при некотором $c > 0$

$$1 - F(x) < \min(1, cx^{-1}), \quad x > 0.$$

Поэтому, интегрируя по частям, получаем

$$n \int_0^\varepsilon x dF(xn) = -n\varepsilon(1 - F(\varepsilon n)) + n \int_0^\varepsilon (1 - F(xn)) dx.$$

Здесь $n\varepsilon(1 - F(\varepsilon n)) \rightarrow a$, $n \int_0^\varepsilon = n \int_0^{c/n} + n \int_{c/n}^{(\ln n)/n} + n \int_{(\ln n)/n}^\varepsilon$. Находим

$$\begin{aligned} n \int_0^{c/n} &\leq c, \quad n \int_{c/n}^{(\ln n)/n} < n \int_{c/n}^{(\ln n)/n} c(nx)^{-1} dx = c \ln x \Big|_{c/n}^{(\ln n)/n} = \\ &= c[\ln \ln n - \ln n - \ln c + \ln n] = c[\ln \ln n - \ln c]. \end{aligned}$$

В последнем интеграле аргумент $xn > \ln n \rightarrow \infty$ при $n \rightarrow \infty$, так что

$$n \int_{(\ln n)/n}^\varepsilon \sim n \int_{(\ln n)/n}^\varepsilon a(xn)^{-1} dx = a \ln n \Big|_{(\ln n)/n}^\varepsilon = a[\ln \varepsilon - \ln \ln n + \ln n].$$

Таким образом,

$$n \int_0^\varepsilon \frac{x}{1+x^2} dF(xn) \sim a \ln n \quad (= o(\ln n) \text{ при } a = 0).$$

Остальная часть интеграла конечна:

$$\begin{aligned} n \int_\varepsilon^\infty \frac{x}{1+x^2} dF(xn) &= \frac{n\varepsilon}{1+\varepsilon^2}(1 - F(\varepsilon n)) + n \int_\varepsilon^\infty (1 - F(xn)) d\left(\frac{x}{1+x^2}\right) < \\ &< \frac{c}{1+\varepsilon^2} + \int_\varepsilon^\infty \frac{c}{x} \left| d\left(\frac{x}{1+x^2}\right) \right| < \infty. \end{aligned}$$

Производя такие же вычисления с $\int_{-\infty}^0$, получим, что

$$A_n = (a - b) \ln n + o(\ln n).$$

Замечание 13. Устойчивость законов $F_{\alpha,\beta}^*$, описанных в замечании 12, проверяется весьма просто, поскольку в силу (59), (60) $\ln \varphi^*(tb) = b^\alpha \ln \varphi^*(t)$. Так как $F_{\alpha,\beta}$ и $F_{\alpha,\beta}^*$ отличаются лишь сдвигом аргумента, то отсюда следует и устойчивость $F_{\alpha,\beta}$.

Чтобы избежать в дальнейшем несущественных оговорок и пояснений, мы не будем различать $\mathbf{F}_{\alpha,\beta}$ и $\mathbf{F}_{\alpha,\beta}^*$ и будем понимать под $\mathbf{F}_{\alpha,\beta}$ распределения с более простыми по форме х. ф. (59), (60).

По поводу роли параметров α и β отметим следующее. Параметр α характеризует скорость сходимости к нулю $\mathbf{F}_{\alpha,\beta}(-x)$ и $1 - \mathbf{F}_{\alpha,\beta}(x)$ при $x \rightarrow \infty$. Оказывается, что $\mathbf{F}_{\alpha,\beta}$ сохраняет свойство (D) с точностью до постоянного множителя. Параметр β , принимающий значения из отрезка $[-1, 1]$, является мерой асимметрии распределения $\mathbf{F}_{\alpha,\beta}$. Если, например, $\beta = 1$ ($a > 0$, $b = 0$), то при $\alpha < 1$ распределение $\mathbf{F}_{\alpha,1}(x)$ есть распределение, целиком сосредоточенное на правой полуоси. Это видно из того, что в этом случае $\mathbf{F}_{\alpha,1}(x)$ можно рассматривать как предельное распределение для нормированных сумм случайных величин $\xi_k \geq 0$ ($F(x) = 0$ при $x < 0$). Так как все допредельные распределения при этом будут сосредоточены на правой полуоси, то там же будет сосредоточено и предельное. Аналогично при $\beta = -1$ ($a = 0$, $b > 0$), $\alpha < 1$ распределение $\mathbf{F}_{\alpha,-1}(x)$ целиком сосредоточено на левой полуоси. При $\beta = 0$ х. ф. $\mathbf{F}_{\alpha,0}$ становится вещественной, а само распределение $\mathbf{F}_{\alpha,0}$ — симметричным (см. свойства х. ф.).

Как мы отмечали, х. ф. для $\mathbf{F}_{\alpha,\beta}$ могут быть найдены в явном виде. Что же касается самих функций распределения $\mathbf{F}_{\alpha,\beta}$, то в явной форме известны лишь немногие законы. Одним из них является закон Коши $\mathbf{F}_{1,0}$ с плотностью $2/(\pi^2 + 4x^2)$. Преобразование сжатия в $\pi/2$ раз приводит эту плотность к виду $1/\pi(1 + x^2)$, соответствующему $\mathbf{K}_{0,1}$.

В заключение рассмотрим один пример. В § 11.8 мы увидим, что в безобидной игре, рассмотренной нами в примере 4.8, время разорения $\eta(z)$ игрока с начальным капиталом z удовлетворяет соотношению $P(\eta(z) \geq n) \sim z\sqrt{2/\pi n}$ при $n \rightarrow \infty$. В частности, при $z = 1$

$$P(\eta(1) \geq n) \sim \sqrt{2/\pi n}. \quad (61)$$

Нетрудно видеть (см. об этом также в главе 11), что $\eta(z)$ будет распределено так же, как сумма $\eta_1 + \eta_2 + \dots + \eta_z$, где η_j независимы и распределены, как $\eta(1)$. Таким образом, при изучении распределения $\eta(z)$, когда z велико, можно в силу (61) воспользоваться теоремой 15 (при $a = \sqrt{2/\pi}$, $b = 0$, $\alpha = 1/2$), по которой (см. также замечание 12)

$$\lim_{z \rightarrow \infty} P\left(\frac{2\pi\eta(z)}{z^2} < x\right) = \mathbf{F}_{1/2,1}^*(x) \quad (62)$$

есть устойчивый закон с показателями $\alpha = 1/2$, $\beta = 1$.

Это утверждение позволяет получить предельное распределение для числа пересечений траекторией S_1, \dots, S_n произвольной полосы $[u, v]$ в случае $\xi \in \mathbf{B}_p$, $p = 1/2$ ($P(\xi_k = 1) = p = P(\xi_k = -1)$). Действительно, пусть для простоты $u = 0$. Первым моментом положительного пересечения полосы $[0, v]$ назовем марковский момент $\eta_+ = \min\{k : S_k = v\}$. Первым моментом отрицательного пересечения назовем момент $\eta_+ + \eta_-$, где $\eta_- = \min\{k : S_{\eta_++k} = 0\}$. Момент $\eta_1 = \eta_+ + \eta_-$ будет также моментом «двойного пересечения» $[0, v]$. Величины η_{\pm} распределены как $\eta(v)$ и независимы, поэтому η_1 распределено как $\eta(2v)$. Величина $H_k = \eta(2v) + \dots + \eta_k(2v)$, где $\eta_i(2v)$ распределены как $\eta(2v)$ и независимы, будет моментом k -го двойного пересечения. Поэтому

$$\nu(n) = \max\{k : H_k \leq n\} = \min\{k : H_k > n\} - 1$$

есть число двойных пересечений полосы $[0, v]$ за время n . Мы можем доказать теперь следующее утверждение:

$$\lim_{n \rightarrow \infty} P\left(\frac{\nu(n)}{\sqrt{n}} \geq x\right) = F_{1/2,1}^*\left(\frac{\pi}{2v^2x^2}\right). \quad (63)$$

Для доказательства воспользуемся следующим соотношением (которое в его более общем виде будет играть важную роль в следующей главе):

$$\{\nu(n) \geq k\} = \{H_k \leq n\},$$

где H_k распределено как $\eta(2vk)$. Если $\frac{n}{k^2} \rightarrow s^2$ при $n \rightarrow \infty$, то в силу (62)

$$P(H_k \leq n) = P\left(\frac{2\pi H_k}{(2vk)^2} \leq \frac{2\pi n}{(2vk)^2}\right) \rightarrow F_{1/2,1}^*\left(\frac{\pi s^2}{2v^2}\right).$$

и, следовательно,

$$P\left(\frac{\nu(n)}{\sqrt{n}} \geq x\right) = P(\nu(n) \geq x\sqrt{n}) = P(H_{[x\sqrt{n}]} \leq n) \rightarrow F_{1/2,1}^*\left(\frac{\pi}{2v^2x^2}\right).$$

(Здесь для $k = [x\sqrt{n}]$ выполняется $n/k^2 \rightarrow s^2 = 1/x^2$.) Утверждение (63) доказано. ■

Ясно, что утверждение (63) сохранится и для числа пересечений $[u, v]$, $u \neq 0$; надо лишь в правой части (63) v заменить на $v - u$. Ясно также, что (63) позволяет найти и предельное распределение для числа «простых» (не двойных) пересечений $[u, v]$, поскольку последнее равно $2\nu(n)$ или $2\nu(n) + 1$.

Г л а в а 9

Элементы теории восстановления

§ 1. Процессы восстановления, функции восстановления

Перейдем теперь к рассмотрению простейших случайных процессов, которые не требуют для своего описания каких-либо специальных конструкций и целиком определяются последовательностями обычных случайных величин.

Пусть $\{\xi_j\}_{j=1}^{\infty}$ есть такая последовательность, причем ξ_j независимы.

Определение 1. *Процессом восстановления* будем называть совокупность случайных величин $\eta(t)$, зависящих от параметра t и определенных на $(\Omega, \mathfrak{F}, P)$ равенством

$$\eta(t) = \min(k: S_k > t), \quad t \geq 0, \quad (1)$$

где $S_k = \sum_{j=1}^k \xi_j$, $S_0 = 0$.

Величины $\eta(t)$ определены еще не полностью. Мы не знаем, чему равны $\eta(t)$ на тех исходах ω , на которых уровень t никогда превзойден не будет. В этом случае естественно положить

$$\eta(t) = \infty, \quad \text{если все } S_k \leq t. \quad (2)$$

Очевидно, что $\eta(t)$ является моментом остановки.

Обычно случайные величины ξ_2, ξ_3, \dots предполагаются одинаково распределенными и имеющими конечное математическое ожидание. Распределение случайной величины ξ_1 может быть произвольным.

Пусть сначала все ξ_j положительны. В этом случае определение (1) позволяет рассматривать $\eta(t)$ как случайную функцию, которую можно описать следующим образом. Если на вещественной оси нанести точки $S_0 = 0, S_1, S_2, \dots$, то $\eta(0) = 1$ на полуинтервале $[0, S_1)$, $\eta(t) = 2$ на полуинтервале $[S_1, S_2)$ и т. д.

Часто процессом восстановления называют и саму последовательность $\{S_k\}_{k=0}^{\infty}$. Иногда последовательность $\{S_k\}$ мы будем называть также *случайным блужданием*.

Сам термин «процесс восстановления» связан с тем, что функцию $\eta(t)$ и последовательность $\{S_k\}$ часто используют для описания работы различных физических устройств, содержащих сменяемые элементы. Если, например, ξ_j — время бесперебойной работы прибора, после которого он выходит из строя и требует замены или ремонта («восстановления»), то время S_k будет означать момент k -го «восстановления» прибора. Если момент $S_0 = 0$ также считать моментом «восстановления», то $\eta(t)$ будет равно числу «восстановлений», которые произошли до момента времени t .

Процессы восстановления нередко используются также и в собственно вероятностных исследованиях, а также при изучении других процессов, где присутствуют моменты «регенерации», после которых развитие процесса начинается как бы сначала. С примерами такого использования процессов восстановления мы столкнемся ниже.

Вернемся к случаю произвольных (разнозначных) ξ_j

Определение 2. Функция

$$H(t) = E\eta(t), \quad t \geq 0,$$

называется *функцией восстановления последовательности* $\{S_k\}_{k=0}^{\infty}$.

В существующей литературе чаще используется другое определение.

Определение 2А. Функцией восстановления последовательности $\{S_k\}_{k=0}^{\infty}$ называется функция

$$U(t) = \sum_{j=0}^{\infty} P(S_j \leq t).$$

Если $\xi_j \geq 0$, то эти определения эквивалентны. Действительно, рассмотрим случайную величину

$$\nu(t) = \sum_{j=0}^{\infty} I(S_j \leq t),$$

где $I(A)$ — индикатор события A . Это есть время, проведенное траекторией $\{S_j\}_{j=0}^{\infty}$ (включая момент времени 0) на отрезке $[0, t]$, что для положительных ξ_j совпадает с числом восстановлений на $[0, t]$, так что $\nu(t) = \eta(t)$.

Если ξ_j разнозначны, то, очевидно, $\nu(t) \geq \eta(t)$, и с положительной вероятностью $\nu(t) > \eta(t)$ (траектория $\{S_j\}$ после прохождения уровня t может вернуться в область $(-\infty, t]$). Следовательно, в этом случае $U(t) > H(t)$. Таким образом, для разнозначных ξ_j мы приходим к двум версиям понятия функции восстановления, данным в определениях 2 и 2А. Мы будем называть их соответственно первой и второй версиями. В этой главе мы будем рассматривать главным образом первую версию (определение 2). О второй версии см. Приложение 6.

Отметим также, что определенные нами функции $H(t)$ и $U(t)$ являются непрерывными справа. В существующей литературе часто рассматривают непрерывные слева версии функций восстановления, определенные, соответственно, как

$$H(t-0) = E \min\{k: S_k \geq t\} \quad \text{и} \quad U(t-0) = \sum_{j=0}^{\infty} P(S_j < t).$$

Если все ξ_j одинаково распределены, $F^{*k}(t)$ есть k -значная свертка функции распределения $F(t) = P(\xi_j < t)$, то вторая непрерывная слева версия функции восстановления может быть представлена также в виде $\sum_{k=0}^{\infty} F^{*k}(t)$, где F^{*0} соответствует распределению, сосредоточенному в 0.

С точки зрения рассмотрений, которые будут проводиться ниже, нет никакой разницы, какой вариант непрерывности выбрать. В силу ряда причин в этой главе нам будет удобнее иметь дело с функциями восстановления, непрерывными справа. Все сказанное ниже в равной степени будет справедливо и для функций, непрерывных слева.

Теорема 1 (интегральная теорема восстановления). Пусть $E\xi_1 > -\infty$, ξ_j , $j \geq 2$, одинаково распределены, $E\xi_j = a > 0$. Тогда существует

$$\lim_{t \rightarrow \infty} \frac{H(t)}{t} = \frac{1}{a}.$$

Это соотношение остается справедливым при $a = 0$ и $a = \infty$, то есть левая часть его обращается в ∞ при $a = 0$ и в 0 при $a = \infty$.

Доказательство. Пусть сначала все ξ_j одинаково распределены. Нам понадобится

Определение 3. Случайную величину

$$\chi(t) = S_{\eta(t)} - t > 0$$

мы будем называть *эксцессом* (или «перескоком») уровня t для случайного блуждания $\{S_j\}$.

Так как $\eta(t)$ есть момент остановки, то для одинаково распределенных ξ_j с $E\xi_j = a > 0$ в силу тождества Вальда

$$ES_{\eta(t)} = t + E\chi(t) = aE\eta(t)$$

или, что то же

$$H(t) = E\eta(t) = \frac{t + E\chi(t)}{a}.$$

Отсюда сразу следует оценка снизу

$$H(t) \geq \frac{t}{a}. \quad (3)$$

С другой стороны, для ограниченных сверху случайных величин $\xi_j \leq N$ очевидно выполняется $\chi(t) \leq N$ и, стало быть,

$$H(t) \leq \frac{t + N}{a}.$$

Введем теперь для произвольных ξ_j , $a = E\xi_j > 0$ «срезки» $\xi_j^{(N)} = \min(\xi_j, N)$ и выберем N так, что $a_N = E\xi_j^{(N)} > 0$. Тогда, если функцию восстановления, соответствующую $\xi_j^{(N)}$, обозначить $H_{(N)}(t)$, то мы получим

$$H(t) \leq H_{(N)}(t) \leq \frac{t + N}{a_N}, \quad \limsup_{t \rightarrow \infty} \frac{H(t)}{t} \leq \frac{1}{a_N}.$$

Так как N произвольно, то

$$\limsup_{t \rightarrow \infty} \frac{H(t)}{t} \leq \frac{1}{a}.$$

Вместе с оценкой (3) снизу это доказывает утверждение для одинаково распределенных ξ_j .

Обозначим $H_1(t)$ функцию восстановления в том случае, когда распределения ξ_1 и ξ_j , $j \geq 2$, различны. Тогда

$$H_1(t) = 1 + \int_{-\infty}^t P(\xi_1 \in du) H(t-u).$$

Из приведенных выше рассуждений следует, что существуют постоянные c_1, c_2 такие, что при всех $t \geq 0$

$$H(t) \leq c_1 + c_2 t.$$

Имеем

$$\frac{1}{t} \int_{-\infty}^t P(\xi_1 \in du) H(t-u) = \int_{|u| \leq N} P(\xi_1 \in du) \frac{H(t-u)}{t} + I,$$

где в случае $E\xi_1 > -\infty$ интеграл

$$I \leq \left(\int_{-\infty}^{-N} + \int_N^t \right) P(\xi_1 \in du) \frac{c_1 + c_2(t-u)}{t}$$

выбором N может быть сделан произвольно малым. Отсюда следует, что

$$\frac{H_1(t)}{t} \rightarrow \frac{1}{a}.$$

Последнее утверждение теоремы в случае $a = \infty$ вытекает из оценок сверху, приведенных в доказательстве. Чтобы получить утверждение в случае $a = 0$, надо ввести в рассмотрение функции восстановления $H_{(\varepsilon)}(t)$, соответствующие случайным величинам $\xi_j + \varepsilon$, и заметить, что

$$\dot{H}(t) > H_{(\varepsilon)}(t) > \frac{t}{\varepsilon}.$$

Опишем теперь, как такое же, как теорема 1, утверждение можно доказать для функции $U(t)$ (второй версии функции восстановления) с помощью аналитических подходов.

Пусть $\varphi(v)$ — характеристическая функция ξ_j . Тогда преобразование Фурье—Стилтьеса $\int e^{ivt} dU(t)$ над функцией $U(t)$ будет равно

$$\int e^{ivt} dU(t) = \sum_{k=0}^{\infty} \int e^{ivt} P(S_k \in dt) = \sum_{k=0}^{\infty} \varphi^k(v) = \frac{1}{1 - \varphi(v)}.$$

Положим $\psi(\lambda) = \varphi(i\lambda) = \int e^{-\lambda y} dF(y)$ и предположим для простоты, что выполнено условие $\psi(\lambda) < \infty$ при некотором $\lambda > 0$ (это есть условие Крамера (ср. с условием (A) в § 8.8), означающее экспоненциальное убывание $P(\xi_j < t) = F(t)$ при $t \rightarrow -\infty$). Очевидно, что $\psi(\lambda) = 1 - a\lambda + o(\lambda)$ при $\lambda \downarrow 0$. Отсюда следует, что

$$A(\lambda) = \int e^{-\lambda t} dU(t) = \frac{1}{1 - \psi(\lambda)} \sim \frac{1}{a\lambda}$$

при $\lambda \downarrow 0$, так что $A(\lambda)$ ведет себя асимптотически при $\lambda \downarrow 0$ так же, как

$$\int_0^{\infty} e^{-\lambda y} d\left(\frac{y}{a}\right) = \frac{1}{a\lambda}.$$

Остается воспользоваться так называемыми тауберовыми теоремами, которые для монотонных функций U позволяют находить асимптотическое поведение $U(t)$ при

$t \rightarrow \infty$ по асимптотическому поведению при $\lambda \downarrow 0$ преобразования Лапласа—Стилтьеса $A(\lambda)$. В силу этих теорем¹⁾ $U(t) \sim t/a$ при $t \rightarrow \infty$.

Мы не формулируем это утверждение в виде доказанной теоремы по двум причинам:

(1) мы не предполагаем, что читатель знаком с тауберовыми теоремами,

(2) некоторое более сильное утверждение будет доказано в Приложении 6. Однако исторически многие исследования асимптотических свойств функции $U(t)$ следовали именно этому аналитическому пути.

Значительно более трудной задачей является исследование асимптотического при $t \rightarrow \infty$ поведения приращений $H(t+u) - H(t)$ (или $U(t+u) - U(t)$). С ней тесно связана более общая задача, часто возникающая в приложениях, об асимптотическом поведении интегралов (см., например, гл. 12)

$$\int_0^t g(t-y) dH(y)$$

при $t \rightarrow \infty$ для функций $g(v)$ таких, что $\int_0^\infty g(v) dv < \infty$.

Теоремы, описывающие это асимптотическое поведение, будут называться *основными теоремами восстановления*. Этим теоремам будут посвящены ближайшие разделы и Приложение 6.

Отметим, что в ряде частных случаев названные проблемы решаются очень просто, т. к. функция восстановления $H(t)$ может быть найдена в явном виде. Как это следует из тождества Вальда, использованного выше, для этого достаточно найти в явном виде $E\chi(t)$. Если, например, ξ_j целочисленны, $P(\xi_j = 1) > 0$, $P(\xi_j \geq 2) = 0$, то $\chi(t) = 1$, и тождество Вальда дает $H(t) = (t+1)/a$. Аналогичные равенства будут справедливы, если $P(\xi_j > t) = ce^{-\alpha t}$ при $t > 0$, $\alpha > 0$ (если ξ_j целочисленны, то t принимает лишь целочисленные значения). В этом случае распределение $\chi(t)$ будет экспоненциальным и не будет зависеть от t (подробнее об этом см. ниже и в гл. 14).

§ 2. Основная теорема восстановления в решетчатом случае

Мы будем различать два типа распределений ξ_j : решетчатые (или арифметические) и нерешетчатые.

(L). Мы будем говорить, что распределение ξ_j *решетчатое*, если возможные значения ξ_j кратны некоторому числу, которое, не ограничивая общности, мы будем считать равным 1. При этом мы будем считать, что 1 есть наибольшее число, обладающее названным свойством, то есть наибольший общий делитель (н.о.д.) возможных значений ξ_j равен 1. Такие распределения называют также *арифметическими*.

¹⁾ При невыполнении условия Крамера ($\psi(\lambda) < \infty$ при некотором $\lambda > 0$) требуются более тонкие рассмотрения. Отметим, что $U(0) = \infty$ при $\int_{-\infty}^0 t^2 dF(t) = \infty$.

Если условие (L) не выполнено, то распределение ξ_j мы будем называть *нерешетчатым*. Ясно, например, что случайная величина, принимающая с положительными вероятностями значения 1 и $\sqrt{2}$, не может быть решетчатой.

В этом параграфе мы будем рассматривать решетчатый случай. Здесь функции $P(\xi_j < t)$ и $H(t)$ полностью определяются их значениями в целых точках $t = k$, $k = 0, 1, 2, \dots$.

Рассмотрим сначала случай, когда ξ_j положительны. В этом случае разности

$$h(k) = H(k) - H(k-1) = \sum_{j=0}^{\infty} P(S_j = k), \quad k \geq 1,$$

равны математическим ожиданиям числа попаданий последовательности $\{S_j\}$ в точку k . Положим

$$q_k = P(\xi_1 = k), \quad p_k = P(\xi_j = k) \quad \text{при } j \geq 2.$$

Определение 2. Процесс восстановления $\eta(t)$ будем называть *однородным* и обозначать $\eta_0(t)$, если

$$q_k = \frac{1}{a} \sum_{j=k}^{\infty} p_j, \quad k = 1, 2, \dots, \quad \sum_{k=1}^{\infty} q_k = 1. \quad (4)$$

Если обозначить $p(z)$ производящую функцию

$$p(z) = Ez^{\xi_1} = \sum_{k=1}^{\infty} p_k z^k,$$

то производящая функция $q(z) = Ez^{\xi_1} = \sum_{k=1}^{\infty} q_k z^k$ будет равна

$$q(z) = \frac{1}{a} \sum_{k=1}^{\infty} z^k \sum_{j=k}^{\infty} p_j = \frac{z}{a} \sum_{j=1}^{\infty} p_j \sum_{k=0}^{j-1} z^k = \frac{z}{a} \sum_{j=1}^{\infty} p_j \frac{1-z^j}{1-z} = \frac{z(1-p(z))}{a(1-z)}.$$

Как мы увидим позже, термин «однородный» применительно к процессу $\eta_0(t)$ вполне оправдан. Одним из свидетельств тому может служить

Теорема 1А.

$$H_0(k) \equiv E\eta_0(k) = 1 + \frac{k}{a}.$$

Доказательство. Рассмотрим производящую функцию $\psi(z)$ последовательности $h_0(k) = H_0(k) - H_0(k-1)$:

$$\begin{aligned} \psi(z) &= \sum_{k=1}^{\infty} z^k h_0(k) = \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} z^k P(S_j = k) = \\ &= \sum_{j=1}^{\infty} Ez^{S_j} = q(z) \sum_{j=0}^{\infty} p^j(z) = \frac{q(z)}{1-p(z)} = \frac{z}{a(1-z)}. \end{aligned}$$

Отсюда следует, что $h_0(k) = 1/a$. ■

Процесс $\eta_0(t)$ иногда называют также *стационарным*. Как мы увидим позже, его правильнее было бы называть *процессом со стационарными приращениями* (см. § 21.1).

Правильное поведение функции $H_0(k)$ асимптотически при $k \rightarrow \infty$ сохраняется и при произвольном ξ_1 .

Обозначим через d наибольший общий делитель (н.о.д.) возможных значений ξ_2 :

$$d = \text{н.о.д. } \{k: p_k > 0\}$$

и пусть $g(k)$, $k = 0, 1, \dots$ — произвольная последовательность такая, что $\sum_{k=0}^{\infty} |g(k)| < \infty$.

Теорема 2 (основная теорема восстановления). Если $d = 1$, ξ_i — произвольная целочисленная собственная случайная величина, то

$$h(k) = H(k) - H(k-1) \rightarrow \frac{1}{a}, \quad \sum_{l=1}^k h(l)g(k-l) \rightarrow \frac{1}{a} \sum_{l=0}^{\infty} g(l) \quad \text{при } k \rightarrow \infty.$$

Для доказательства нам потребуются два вспомогательных утверждения.

Лемма 1. Если $\nu \geq 1$ — марковская величина, то σ -алгебра, порожденная случайными величинами $\nu, \xi_1, \dots, \xi_\nu$, и σ -алгебра $\sigma(\xi_{\nu+1}, \dots)$, порожденная последовательностью $\{\xi_{\nu+1}, \xi_{\nu+2}, \dots\}$, независимы. Последовательность $\{\xi_{\nu+1}, \xi_{\nu+2}, \dots\}$ распределена так же, как $\{\xi_2, \xi_3, \dots\}$.

Таким образом, несмотря на случайные номера, элементы последовательности $\xi_{\nu+j}$ распределены как ξ_j .

Доказательство. Для заданных борелевских множеств $B_1, B_2, \dots, C_1, \dots, C_k$ обозначим

$$A = \{\nu \in N, \xi_1 \in B_1, \dots, \xi_\nu \in B_\nu\}, \quad D_\nu = \{\xi_{\nu+1} \in C_1, \dots, \xi_{\nu+k} \in C_k\},$$

где N есть заданное множество целых чисел. В силу теоремы 3.4 достаточно доказать что $P(D_\nu|A) = P(D_1)$ или, что то же, $P(D_\nu|A) = P(D_1)P(A)$.

По формуле полной вероятности имеем

$$P(D_\nu|A) = \sum_{j \in N} P(D_\nu|A\{\nu = j\}) = \sum_{j \in N} P(D_j|A\{\nu = j\}).$$

Но событие $A\{\nu = j\}$ принадлежит $\sigma(\xi_1, \dots, \xi_j)$, в то время как $D_j \in \sigma(\xi_{j+1}, \dots, \xi_{j+k})$. Поэтому события D_j и $A\{\nu = j\}$ независимы,

$$P(D_j|A\{\nu = j\}) = P(D_j)P(A\{\nu = j\}); \quad P(D_j) = P(D_1), \quad j \geq 1.$$

Отсюда следует, очевидно, что $P(D_\nu|A) = P(D_1)P(A)$. ■

Лемма 2. Пусть ζ_1, ζ_2, \dots — независимые, целочисленные, одинаково и симметрично распределенные случайные величины с конечным математическим ожиданием ($E\zeta_j = 0$) и н.о.д. возможных значений, равным 1. Положим $Z_n = \sum_{j=1}^n \zeta_j$. Тогда при любом целом k

$$\nu_k = \min\{n: Z_n = k\}$$

есть собственная случайная величина: $P(\nu_k < \infty) = 1$.

Доказательство этой леммы содержится в § 12.3 (см. следствие 12.2).

Доказательство теоремы 2. Рассмотрим две независимые между собой последовательности случайных величин (считаем их заданными на одном вероятностном пространстве): последовательность ξ_1, ξ_2, \dots , где ξ_1 имеет произвольное распределение, и последовательность ξ'_1, ξ'_2, \dots , где $P(\xi'_1 = k) = q_k$, $P(\xi'_j = k) = P(\xi_j = k) = p_k$ при $j \geq 2$ (так что $\xi'_j = \xi_j$ при $j \geq 2$; процесс $\eta'(t)$,

построенный по суммам $S'_k = \sum_{j=1}^k \xi'_j$, является однородным).

Обозначим $\nu = \min\{n \geq 1: S_n = S'_n\}$. Это есть, очевидно, марковский момент относительно последовательности $(\xi_1, \xi'_1), (\xi_2, \xi'_2), \dots$. Покажем, что $P(\nu < \infty) = 1$. Положим

$$Z_n = \sum_{j=2}^n (\xi_j - \xi'_j) \quad \text{при } n \geq 2, \quad Z_1 = 0, \quad Z_0 = \xi_1 - \xi'_1.$$

Тогда

$$\nu = \min\{n \geq 1: Z_n = -Z_0\}.$$

По лемме 2 ($\zeta_j = \xi_j - \xi'_j$ имеют при $j \geq 2$ симметричное распределение) для каждого целого k величина $\nu_k = \min\{n \geq 1: Z_n = k\}$ является собственной. Так как Z_n при $n \geq 1$ и Z_0 независимы, то

$$P(\nu < \infty) = \sum_k P(Z_0 = -k)P(\nu_k < \infty) = 1.$$

Произведем теперь «склеивание» последовательностей $\{S_n\}$ и $\{S'_n\}$. Так как $S_\nu = S'_\nu$ и ν есть марковский момент, то можно по лемме 1 заменить $\xi_{\nu+1}, \xi_{\nu+2}, \dots$ на $\xi'_{\nu+1}, \xi'_{\nu+2}, \dots$ (и тем самым $S_{\nu+1}, S_{\nu+2}, \dots$ на $S'_{\nu+1}, S'_{\nu+2}, \dots$), не изменяя при этом распределения последовательности $\{S_n\}$.

Следовательно, на множестве $\{S_\nu < k\}$ выполняется $\eta(t) = \eta'(k)$ при $t \geq k - 1$ и, значит,

$$\begin{aligned} h(k) &= E(\eta(k) - \eta(k-1)) = \\ &= E[\eta'(k) - \eta'(k-1); S_\nu < k] + E[\eta(k) - \eta(k-1); S_\nu \geq k] = \\ &= \frac{1}{a} - E[\eta'(k) - \eta'(k-1); S_\nu \geq k] + E[\eta(k) - \eta(k-1); S_\nu \geq k]. \end{aligned}$$

Так как $|\eta(k) - \eta(k-1)| \leq 1$, то

$$\left| h(k) - \frac{1}{a} \right| \leq P(S_\nu \geq k) \rightarrow 0$$

при $k \rightarrow \infty$. Первое утверждение теоремы доказано.

Так как $h(k) \leq 1$, то

$$\left| \sum_{l=1}^{k-N} h(l)g(k-l) \right| \leq \sum_{l=N+1}^{k-1} |g(l)| \leq \sum_{l=N+1}^{\infty} |g(l)| \rightarrow 0 \quad \text{при } N \rightarrow \infty.$$

Кроме того

$$\sum_{l=k-N}^k h(l)g(k-l) \rightarrow \frac{1}{a} \sum_{l=0}^N g(l) \quad \text{при } k \rightarrow \infty.$$

Поэтому второе утверждение теоремы является простым следствием первого. ■

Замечание 1. Склеивание $\{S_n\}$ и $\{S'_n\}$ в доказательстве теоремы 2 можно произвести и раньше, в момент

$$\gamma = \min\{n \geq 1: S_n \in S'\},$$

где S' есть множество точек $S' = \{S'_1, S'_2, \dots\}$.

Теорема 2А. Для произвольных (разнозначных) ξ_j утверждение теоремы 2 сохраняется.

Доказательство. Мы редуцируем задачу к случаю $\xi_1 > 0$. Пусть сначала все ξ_j одинаково распределены. Рассмотрим случайную величину $\chi_1 = \chi(0)$, которую мы будем называть первой положительной суммой. В гл. 11 будет показано (см. следствие 11.2), что $E\chi_1 < \infty$, если $a = E\xi_j < \infty$. В соответствии с леммой 1 последовательность $\xi_{\eta(0)+1}, \xi_{\eta(0)+2}, \dots$ будет иметь то же распределение, что и ξ_1, ξ_2, \dots . Поэтому «вторая положительная сумма» χ_2 или, что то же, первая положительная сумма величин $\xi_{\eta(0)+1}, \xi_{\eta(0)+2}, \dots$, будет распределена так же, как χ_1 , и не будет зависеть от нее. То же будет справедливо и для последующих «перескоков» уже достигнутых уровней $\chi_1, \chi_1 + \chi_2, \dots$. Рассмотрим теперь случайное блуждание $\{H_k = \sum_{i=1}^k \chi_i\}_{k=1}^{\infty}$ и обозначим для него

$$\eta^*(t) = \min\{k: H_k > t\}, \quad \chi^*(t) = H_{\eta^*(t)} - t, \quad H^*(t) = E\eta^*(t).$$

Так как $\chi_k > 0$, то применима теорема 2 и, стало быть, в силу тождества Вальда

$$H^*(k) - H^*(k-1) = \frac{1 + E\chi^*(k) - E\chi^*(k-1)}{E\chi_1} \rightarrow \frac{1}{E\chi_1},$$

$$E\chi^*(k) - E\chi^*(k-1) \rightarrow 0.$$

Заметим теперь, что распределения случайных величин $\chi(t)$ и $\chi^*(t)$ совпадают. Поэтому

$$H(k) - H(k-1) = \frac{1 + E\chi(k) - E\chi(k-1)}{a} = \frac{1 + E\chi^*(k) - E\chi^*(k-1)}{a} \rightarrow \frac{1}{a}.$$

Пусть теперь распределения ξ_1 и ξ_j , $j \geq 2$, различны. Тогда функция восстановления $H_1(t)$ для такого блуждания будет равна

$$H_1(k) = 1 + \sum_{i=-\infty}^k P(\xi_1 = i)H(k-i),$$

$$h_1(k) = H_1(k) - H_1(k-1) = \sum_{i=-\infty}^k P(\xi_1 = i)h(k-i).$$

Если допустить, что $h(k) < c < \infty$ при всех k , то отсюда будет следовать $h_1(k) \rightarrow 1/a$, что доказывает теорему.

Требуемое неравенство $h(k) < c$ вытекает из следующего общего предложения, справедливого для произвольных (не обязательно решетчатых) ξ_j .

Лемма 3. Если все ξ_j одинаково распределены, то при всех t, u

$$H(t+u) - H(t) \leq H(u) \leq c_1 + c_2 u.$$

Доказательство. Разность $\eta(t+u) - \eta(t)$ есть число скачков траектории $\{\tilde{S}_k\}$, стартовавшей в точке $t + \chi(t) > t$ до первого прохождения уровня $t + u$, при этом последовательность $\{\tilde{S}_k\}$ распределена так же, как последовательность $\{S_k\}$, и независима от нее (см. лемму 1). Другими словами, $\eta(t+u) - \eta(t)$ имеет то же распределение, что и $\tilde{\eta}(t - \chi(t)) \leq \tilde{\eta}(t)$, где $\tilde{\eta}$ соответствует $\{\tilde{S}_k\}$, если $\chi(t) \leq u$, и $\eta(t+u) - \eta(t) = 0$, если $\chi(t) > u$. Поэтому $H(t+u) - H(t) \leq H(u)$. Неравенство для $H(u)$ вытекает из теоремы 1. ■

Теорема 2А доказана. ■

§ 3. Экссесс и дефект случайного блуждания. Предельное распределение в решетчатом случае

Наряду с эксцессом $\chi(t) = S_{\eta(t)} - t$ введем в рассмотрение еще одну случайную величину, тесно связанную с $\chi(t)$.

Определение 3 А. Случайная величина

$$\gamma(t) = t - S_{\eta(t)-1} \geq 0$$

называется величиной *дефекта* уровня t в блуждании $\{S_n\}^2$.

Величина $\chi(t)$ имеет смысл времени, которое еще проработает прибор, работающий в момент t ; $\gamma(t)$ есть время, которое этот прибор уже работает.

Не следует думать, что сумма $\chi(t) + \gamma(t)$ распределена так же, как ξ_j — ведь эта сумма равна значению ξ со случайным индексом $\eta(t)$. В частности, как мы увидим ниже, может оказаться, что $E\chi(t) > E\xi_j$ при больших t . С этим обстоятельством связан кажущийся парадокс в следующем примере. Пассажира, пришедший на автобусную остановку, через которую проходят автобусы с интервалами $\xi_1 > 0$, $\xi_2 > 0, \dots$, ($E\xi_j = a$), будет ждать прихода очередного автобуса время χ , среднее значение которого $E\chi$ может оказаться больше, чем a .

Существенным фактом теории восстановления является утверждение о том, что в широких предположениях существует совместное предельное распределение $\chi(t)$ и $\gamma(t)$ при $t \rightarrow \infty$, так что при больших t распределение $\chi(t)$ перестает зависеть от t и становится *стационарным*. Обозначим это предельное распределение $\chi(t)$ через G :

$$G(x) = \lim_{t \rightarrow \infty} P(\chi(t) < x). \quad (5)$$

Если в качестве распределения ξ_1 взять распределение G , то по самому его построению распределение величины $\chi(t)$ для построенного процесса от t зависеть не будет. В самом деле, в этом случае $\{S_j\}$ можно представлять себе как положительные моменты восстановления для процесса, который построен по последовательности $\{\xi_j\}$ и начало которого отнесено в точку $-N$, где N велико. Так как в силу (5) можно считать, что распределения $\chi(N)$ и $\chi(N+t)$ совпадают, то распределение величины $\chi(t)$ (которую можно отождествить с $\chi(N+t)$) от t не зависит и совпадает с распределением ξ_1 . Формальное доказательство этого факта мы опускаем, поскольку дальше он использоваться не будет. Однако читатель может провести его, используя явный вид $G(x)$ в (5), полученный ниже.

²⁾ В литературе $\chi(t)$ называют часто первым «перескоком» через уровень t , а $\gamma(t)$ «недоскоком». Эти термины весьма наглядны, но их можно упрекнуть в недостаточной «литературности».

Распределение G в решетчатом случае и есть распределение (4), с помощью которого строился однородный процесс восстановления $\eta_0(t)$. Этот факт будет доказан чуть позже.

Из того, что распределение $\chi(t)$ для процесса $\eta_0(t)$ не зависит от t и совпадает с распределением ξ_1 , будет следовать, что распределение $\eta_0(t+u) - \eta_0(t)$ совпадает с распределением $\eta_0(u)$ и, следовательно, от t также не зависит. Это свойство и определяет стационарность приращений процесса восстановления; мы назвали его однородностью процесса. Оно означает, что распределение числа восстановлений в интервале времени длины u не зависит от того, с какого момента мы начали отсчет, и, следовательно, зависит только от u .

Теоремы о предельном распределении $\chi(t)$ и $\gamma(t)$ представляют интерес не только с точки зрения приложений. Они понадобятся нам и для целого ряда других задач. Мы рассмотрим опять сначала случай, когда величины $\xi_j > 0$ целочисленны. В этом случае «время» t можно считать дискретным, и мы будем, как и прежде, обозначать его буквами n и k .

Теорема 3. Пусть н. о. д. возможных значений ξ_2 равен 1, существует $E\xi_2 = a$, ξ_1 есть произвольная целочисленная случайная величина. Тогда существует

$$\lim_{k \rightarrow \infty} P(\gamma(k) = i, \chi(k) = j) = \frac{p_{i+j}}{a}. \quad (6)$$

Из теоремы 3 следует, что

$$\begin{aligned} \lim_{k \rightarrow \infty} P(\chi(k) = i) &= \frac{1}{a} \sum_{j=i}^{\infty} p_j, \\ \lim_{k \rightarrow \infty} P(\gamma(k) = i) &= \frac{1}{a} \sum_{j=i+1}^{\infty} p_j. \end{aligned} \quad (7)$$

Доказательство теоремы 3. В силу теоремы восстановления имеем при $k > i$

$$\begin{aligned} P(\gamma(k) = i, \chi(k) = j) &= \sum_{l=1}^{\infty} P(S_l = k - i, \xi_{l+1} = i + j) = \\ &= \sum_{l=1}^{\infty} P(S_l = k - i) P(\xi_2 = i + j) = h(k - i) p_{i+j} \rightarrow \frac{p_{i+j}}{a}. \quad \blacksquare \end{aligned}$$

Если $E\xi_2^2 = m_2 < \infty$, то теорема 3 позволяет уточнить теорему 2.

Следствие 3. Если $m_2 < \infty$, то случайные величины $\chi(k)$ равномерно интегрируемы,

$$E\chi(k) \rightarrow \frac{1}{a} \sum_{i=0}^{\infty} i \sum_{j=i}^{\infty} p_i = \frac{m_2 + a}{2a}. \quad (8)$$

Доказательство. Равномерная интегрируемость следует из неравенств $h(k) \leq 1$,

$$P(\chi(k) = j) = \sum_{i=0}^k h(k-i) p_{i+j} \leq \sum_{i=j}^{\infty} p_i.$$

Отсюда следует (8) (см. § 6.1). ■

Теорема 2А. Если все ξ_j одинаково распределены, $E\xi_j^2 = m_2 < \infty$, то

$$H(k) = \frac{k}{a} + \frac{1}{2a^2}(m_2 + a) + o(1).$$

Доказательство немедленно следует из тождества Вальда $H(k) = E\eta(k) = (k + E\chi(k))/a$ и следствия 3. ■

В каких случаях распределение ξ_1 для однородного процесса восстановления совпадает с распределением ξ_j при $j \geq 2$? Такое совпадение эквивалентно тому, что при $i = 1, 2, \dots$ справедливо

$$p_i = \frac{1}{a} \sum_{j=i}^{\infty} p_j,$$

или, что то же,

$$a(p_i - p_{i-1}) = -p_{i-1}, \quad p_i = \frac{a-1}{a} p_{i-1}, \quad p_i = \frac{1}{a-1} \left(\frac{a-1}{a} \right)^i.$$

Это означает, что процесс восстановления, порожденный последовательностью одинаково распределенных случайных величин ξ_1, ξ_2, \dots , является однородным тогда и только тогда, когда ξ_j (или, точнее, $\xi_j - 1$) имеют геометрическое распределение.

Обозначим γ и χ случайные величины с распределением (6). Пользуясь (5), нетрудно установить, что γ и χ независимы также лишь в случае, когда ξ_j , $j \geq 2$, имеют геометрическое распределение. Для геометрического распределения ξ_j , $j \geq 1$, независимы также $\gamma(n)$ и $\chi(n)$; $\chi(n) = \xi_1$. Доказательство этих фактов проводится точно так же, как для показательного распределения (см. § 4).

Вернемся к общему случаю и напомним, что

$$E\chi = \frac{E\xi_1^2}{2a} + \frac{1}{2}.$$

Это означает, в частности, что если распределение ξ_1 таково, что $E\xi_1^2 > 2a^2 - a$, то среднее значение эксцесса $E\chi(n)$ при больших n станет больше, чем $E\xi_1 = a$.

§ 4. Теорема восстановления и предельное распределение эксцесса и дефекта в нерешетчатом случае

Напомним, что нерешетчатым мы называем случай, когда не существует такого числа $d > 0$, что $P(\bigcup_k \{\xi_2 = kd\}) = 1$, где k пробегает все целочисленные значения.

Чтобы сформулировать основную теорему восстановления в нерешетчатом, случае нам понадобится понятие *непосредственно интегрируемой* функции.

Определение 4. Функция $g(u)$, определенная на $[0, \infty)$, называется *непосредственно интегрируемой*, если

- 1) Функция g интегрируема по Риману³⁾ на любом конечном интервале $[0, N]$.
- 2) $\sum g_k < \infty$, где $g_k = \max_{k \leq u < k+1} |g(u)|$.

³⁾ То есть суммы $\frac{1}{n} \sum g_k$ и $\frac{1}{n} \sum \bar{g}_k$ имеют при $n \rightarrow \infty$ одинаковые пределы, где $\underline{g}_k = \min_{u \in \Delta_k} g(u)$, $\bar{g}_k = \max_{u \in \Delta_k} g(u)$, $\Delta_k = [k\Delta, (k+1)\Delta)$, $\Delta = N/n$.

Очевидно, что любая монотонно убывающая функция $g(t) \downarrow 0$, имеющая конечный интеграл Лебега $\int_0^\infty g(u) du < \infty$, непосредственно интегрируема. Это же справедливо и для разностей таких функций.

Понятие функций, непосредственно интегрируемых по Риману, введенное в [22], несколько отличается от понятия, определенного выше, хотя по сути с ним совпадает. Нам будет удобнее пользоваться определением 4, т. к. оно позволяет несколько упростить изложение и избежать вспомогательных рассуждений (см. Приложение 6).

Теорема 4. Пусть функция g непосредственно интегрируема. Если ξ_j , $j \geq 2$, нерешетчатые, существует $E\xi_2 = a > 0$, распределение ξ_1 произвольно, то при $t \rightarrow \infty$

$$\int_0^t g(t-u) dH(u) \rightarrow \frac{1}{a} \int_0^\infty g(u) du. \quad (9)$$

С функцией H ассоциирована мера на $[0, \infty)$, которую можно обозначить тем же символом H и которая определяется соотношениями $H((x, y)) = H(y) - H(x)$. Тогда наряду с $\int g(t-u) dH(u)$ в (9) можно писать также $\int g(t-u) H(du)$.

Из (9) следует, в частности, что при любом фиксированном u

$$H(t) - H(t-u) \rightarrow \frac{u}{a}. \quad (10)$$

Нетрудно видеть, что это соотношение, называемое *локальной теоремой восстановления*, эквивалентно (9).

Доказательство теоремы 4 технически оказывается довольно сложным, поэтому мы поместили его в Приложение 6. Там же можно найти уточнения теоремы 4 и ее аналог в случае существования плотности распределения ξ_2 .

Другие утверждения § 2, 3 переносятся на нерешетчатый случай без затруднений. Пусть все ξ_j положительны.

Определение 5. Процесс $\eta(t)$ в нерешетчатом случае называется *однородным* (и обозначается $\eta_0(t)$), если распределение первого скачка имеет вид

$$Q(x) = P(\xi_1 > x) = \frac{1}{a} \int_x^\infty P(\xi_2 > t) dt.$$

Если обозначить $p(\lambda)$ х. ф. ξ_2 , то х. ф. ξ_1 будет равна

$$q(\lambda) = E e^{i\lambda \xi_1} = \frac{1}{a} \int_0^\infty e^{i\lambda x} P(\xi_2 > x) dx.$$

Обычное определение интегрируемости по Риману на $[0, \infty)$ предполагает выполнение условия 1) и существование предела $\int_0^N g(u) du$ при $N \rightarrow \infty$. Такой подход охватывает более широкий класс функций, чем в определении 4, допускающих, например, существование последовательности $t_k \rightarrow \infty$ такой, что $g(t_k) \rightarrow \infty$.

Так как интегрирование здесь ведется по $x > 0$, то интеграл существует (как и функция $p(\lambda)$) при всех λ , $\text{Im } \lambda > 0$ (при $\lambda = i\alpha + v$, $-\infty < v < \infty$, $\alpha > 0$, множитель $e^{i\lambda x}$ равен $e^{-\alpha x} e^{ivx}$, см. свойство 6 х. ф.) Поэтому при $\text{Im } \lambda > 0$

$$q(\lambda) = -\frac{1}{i\lambda a} \left[1 + \int_0^{\infty} e^{i\lambda x} dP(\xi_2 > x) \right] = \frac{p(\lambda) - 1}{i\lambda a}.$$

Теорема 5. $H_0(t) \equiv E\eta_0(t) = t/a$.

Доказательство. Эта теорема доказывается так же, как теорема 1. Рассмотрим преобразование Фурье—Стилтьеса над функцией $H_0(t)$:

$$\psi(\lambda) = \int_0^{\infty} e^{i\lambda x} dH_0(x).$$

Отметим, что это преобразование существует при $\text{Im } \lambda > 0$ и для него сохраняется теорема единственности, установленная для х. ф., так как $\varphi(v) = \psi(i\alpha + v)/\psi(i\alpha)$, $-\infty < v < \infty$ (мы положили $\lambda = i\alpha + v$ при фиксированном $\alpha > 0$) можно рассматривать как х. ф. некоторого распределения, полученного «преобразованием Крамера» (см. § 8 гл. 8) над $H_0(t)$.

Имеем в силу (3)

$$\psi(\lambda) = \int_0^{\infty} e^{i\lambda x} dH_0(x) = \sum_{j=0}^{\infty} q(\lambda) p^j(\lambda) = \frac{p(\lambda) - 1}{i\lambda a} \frac{1}{1 - p(\lambda)} = -\frac{1}{i\lambda a}.$$

Очевидно, что это преобразование соответствует функции $H_0(t) = t/a$. ■

Связи между однородным процессом восстановления $\eta_0(t)$ и предельным распределением $\chi(t)$ и $\gamma(t)$ в нерешетчатом случае те же, что и в решетчатом. Так же, как и в § 3, из теоремы восстановления вытекает

Теорема 6. Если ξ_2 нерешетчатая, $E\xi_2 = a$, распределение ξ_1 произвольно, то существует

$$\lim_{t \rightarrow \infty} P(\gamma(t) > u, \chi(t) > v) = \frac{1}{a} \int_{u+v}^{\infty} P(\xi_2 > x) dx. \quad (15)$$

Доказательство. При $t > u$

$$\begin{aligned} P(\gamma(t) > u, \chi(t) > v) &= \\ &= \sum_{j=0}^{\infty} \int_0^{t-u} P(\eta(t) = j, S_j \in dx, \gamma(t) > u, \chi(t) > v) = \\ &= \sum_{j=0}^{\infty} \int_0^{t-u} P(S_j \in dx, \xi_{j+1} > t - x + v) = \\ &= \int_0^{t-u} dH(x) P(\xi_2 > t - x + v). \end{aligned} \quad (16)$$

В силу теоремы восстановления (см. (10)) при $g(x) = P(\xi_2 > x + u + v)$ последний интеграл сходится к $\frac{1}{a} \int_0^{\infty} P(\xi_2 > x + u + v) dx$. ■

Как и в предыдущем параграфе, в случае $E\xi_2^2 = m_2 < \infty$ теорема 6 позволяет уточнить основную теорему восстановления.

Теорема 4А. Если все $\xi_j > 0$ одинаково распределены, $E\xi_j^2 = m_2 < \infty$, то

$$H(t) = \frac{t}{a} + \frac{m_2}{2a^2} + o(1).$$

Доказательство. Из соотношений (16) и леммы 3 следует равномерная интегрируемость $\chi(t)$, т. к.

$$P(\chi(t) > v) = \int_0^t dH(x) P(\xi_j > t - x + v) < (c_1 + c_2) \sum_{k \geq 0} P(\xi_j > k + v),$$

поэтому

$$E\chi(t) \rightarrow \frac{1}{a} \int_0^{\infty} \int_0^{\infty} P(\xi_j > u) dudv = \frac{m_2}{2a}.$$

Остается воспользоваться тождеством Вальда. ■

Пусть, как и прежде, γ и χ — случайные величины, распределение которых совпадает с предельным распределением (15). Из теоремы 6 видно, что каждая из случайных величин γ и χ имеют плотности, равные $1/a P(\xi_2 > x)$. Совместное распределение γ и χ может не иметь плотности. Если существует плотность $f(x)$ случайной величины ξ_2 , то плотность γ и χ существует и равна $1/a f(x + y)$. Из теоремы 6 вытекает также, что γ и χ независимы, если только

$$\int_x^{\infty} P(\xi_2 > u) du = \frac{1}{\alpha} e^{-\alpha x}$$

при каком-нибудь $\alpha > 0$, т. е. независимость имеет место лишь для показательного распределения $\xi_2 \in \Gamma_{\alpha}$.

Кроме того, совпадение $P(\xi_1 > x)$ и $P(\xi_2 > x)$ для однородного процесса восстановления также имеет место лишь для $\xi_2 \in \Gamma_{\alpha}$. Другими словами, процесс восстановления, порожденный последовательностью одинаково распределенных случайных величин ξ_1, ξ_2, \dots будет однородным тогда и только тогда, когда $\xi_j \in \Gamma_{\alpha}$. В этом случае процесс $\eta_0(t)$ называется (см. также § 18.4) пуассоновским. Это связано с тем, что для такого процесса величина $\eta(t) = \eta_0(t)$ при каждом t имеет распределение Пуассона с параметром t/a .

Пуассоновский процесс обладает и рядом других замечательных свойств (см. также § 18.4). Для него, очевидно, $\chi(t) \in \Gamma_{\alpha}$, но, кроме того, случайные величины $\gamma(t)$ и $\chi(t)$ независимы. Действительно, в силу (16) при $u < t$

$$P(\gamma(t) > u, \chi(t) > v) = e^{-\alpha(t+v)} + \alpha \int_0^{t-u} e^{-\alpha(t-x+v)} dx =$$

$$= e^{-\alpha(u+v)} = P(\gamma(t) > u)P(\chi(t) > v);$$

$$P(\gamma(t) = t, \chi(t) > v) = P(\xi_1 > t + v) = e^{-\alpha(t+v)};$$

$$P(\gamma(t) > t) = 0.$$

Факт независимости $\gamma(t)$ и $\chi(t)$ заслуживает внимания с точки зрения его интерпретации. Он означает следующее. Если мы будем следить за сроком службы некоторого прибора, начиная с момента времени t , то этот срок будет иметь то же распределение, что и срок службы нового прибора (распределение $\xi_j \in \Gamma_\alpha$), и не будет зависеть от того, сколько этот прибор уже проработал (на первый взгляд, явление парадоксальное). Так как распределение длительности безотказной работы всех механизмов и приборов, состоящих из большого числа «надежных» элементов, близко к показательному (см. теорему 19.6), то отмеченное обстоятельство представляет значительный практический интерес.

Если ξ_j принимает и отрицательные значения, то задачи, связанные с распределениями $\gamma(n)$ и $\chi(n)$, качественно усложняются. В какой-то мере эти задачи могут быть сведены к случаю неотрицательных величин, поскольку распределение $\chi(n)$ совпадает с распределением величины $\chi^*(n)$, построенной по последовательности $\{\xi_j^* \geq 0\}$, где ξ_j^* распределены так же, как $\chi(0)$. Распределение $\chi(0)$ может быть найдено методами гл. 11.

§ 5. Закон больших чисел и центральная предельная теорема для процесса восстановления

Предположим сначала, что $\xi_j \geq 0$.

Теорема 7. Если $0 < E\xi_2 = a < \infty$, то при $t \rightarrow \infty$

$$\frac{\eta(t)}{t} \xrightarrow{p} \frac{1}{a}.$$

Доказательство. Основным соотношением, которым мы здесь будем пользоваться, является равенство

$$\{\eta(t) > k\} = \{S_k \leq t\}. \quad (17)$$

Так как

$$P\left(\frac{\eta(t)}{t} - \frac{1}{a} > \varepsilon\right) = P\left(\eta(t) > \frac{t}{a} + \varepsilon t\right),$$

то, считая для простоты, что $N = t/a + \varepsilon t$ является целым числом, получим

$$P\left(\frac{\eta(t)}{t} - \frac{1}{a} > \varepsilon\right) = P\left(S_N \leq \frac{N}{1/a + \varepsilon}\right) = P\left(\frac{S_N}{N} \leq a - \varepsilon_1\right),$$

где $\varepsilon_1 = a(1 - 1/(1 + a\varepsilon)) > 0$. По закону больших чисел это означает, что оцениваемая вероятность сходится к 0 при $t \rightarrow \infty$.

Аналогичным образом оценивается вероятность

$$P\left(\frac{\eta(t)}{t} - \frac{1}{a} < -\varepsilon\right).$$

Теорема 8. Если $E\xi_2 = a$, $D\xi_2 = \sigma^2 < \infty$, то при $t \rightarrow \infty$

$$\frac{\eta(t) - t/a}{\sqrt{t\sigma^2/a^3}} \Rightarrow \Phi_{0,1}.$$

Доказательство. На основании (17) имеем

$$P(\eta(t) > k) = P\left(\frac{S_k - ak}{\sigma\sqrt{k}} \leq \frac{t - ak}{\sigma\sqrt{k}}\right). \quad (18)$$

По центральной предельной теореме

$$P\left(\frac{S_k - ak}{\sigma\sqrt{k}} \leq v\right) \rightarrow \Phi(v) \quad \text{при } k \rightarrow \infty.$$

Пусть теперь k при $t \rightarrow \infty$ меняется так, что выполняется $(t - ak)/(\sigma\sqrt{k}) = v$. Считая v фиксированным, можно разрешить это равенство относительно k :

$$k = \frac{t}{a} + \frac{v\sigma(v\sigma \pm 2\sqrt{at}\sqrt{1 + (v^2\sigma^2)/(4ta)})}{2a^2} = \frac{t}{a} \pm \frac{v\sigma}{a^2}\sqrt{at}\left(1 + O\left(\frac{1}{\sqrt{t}}\right)\right)$$

Подставляя найденное значение k в (18), получим

$$P(\eta(t) > k) = P\left(\frac{\eta(t) - t/a}{(\sigma\sqrt{t})/(a^{3/2})} \geq \pm v + O\left(\frac{1}{\sqrt{t}}\right)\right) \rightarrow \Phi(v). \quad (19)$$

Здесь следует, очевидно, выбрать знак минус, так как только в этом случае в обеих частях полученного соотношения будут стоять возрастающие по v функции.

Так как $\Phi(v)$ непрерывна, то наряду с (19)

$$P\left(\frac{\eta(t) - t/a}{(\sigma\sqrt{t})/(a^{3/2})} > -v\right) \rightarrow \Phi(v) = 1 - \Phi(-v)$$

или

$$P\left(\frac{\eta(t) - t/a}{(\sigma\sqrt{t})/(a^{3/2})} \leq v\right) \rightarrow \Phi(v).$$

В следующей главе будет доказан усиленный закон больших чисел, в силу которого $S_k/k \rightarrow a$ п. н. Пользуясь этим фактом, читатель может наряду с теоремой 7 установить, что $\eta(t)/t \rightarrow 1/a$ при $t \rightarrow \infty$ с вероятностью 1.

Обратимся теперь к разнозначным ξ_j и предположим для простоты, что все они одинаково распределены. Из усиленного закона больших чисел будет следовать также, что для заданного $\varepsilon > 0$, $\varepsilon < a$, при всех достаточно больших k будет выполняться $S_k > (a - \varepsilon)k$ и, стало быть, $\underline{S} = \inf S_k$ есть собственная случайная величина (см. об этом также в гл. 11). Отсюда будет вытекать следующее утверждение о $\bar{S}_n = \max_{0 \leq k \leq n} S_k$.

Лемма 4. Случайная величина

$$\zeta_n = \bar{S}_n - S_n \geq 0$$

мажорируется по распределению собственной случайной величиной $-\underline{S} = -\inf S_k$:

$$P(\zeta_n > x) \leq P(-\underline{S} > x).$$

Доказательство. Имеем

$$\begin{aligned}\bar{S}_n &= \max(0, S_1, \dots, S_n) = S_n + \max(-S_n, S_1 - S_n, \dots, -\xi_n, 0) = \\ &= S_n - \min(0, \xi_n, \dots, S_n - S_1, S_n).\end{aligned}$$

Распределение $\zeta_n = -\min(0, \xi_n, \dots, S_n)$ не изменится, если мы перенумеруем случайные величины и заменим ξ_{n-j} на ξ_{j+1} , $j = 0, 1, \dots, n-1$, в результате чего получим $\zeta_n = -\min(0, \xi_1, \dots, S_{n-1}, S_n) \leq -\underline{S}$. ■

Теорема 9. Утверждения теорем 7, 8 справедливы и для разнозначных ξ_j , $E\xi_j = a > 0$.

Доказательство. Теперь вместо (13) мы будем иметь

$$\{\eta(t) > k\} = \{\bar{S}_k \leq t\}.$$

В силу леммы 4 и замечания 6.5 предельные распределения нормированных S_n и \bar{S}_n будут совпадать. Поэтому, повторяя рассуждения теорем 7, 8, мы получим требуемое утверждение. ■

Г л а в а 10

Последовательности независимых случайных величин. Свойства траектории $(0, S_1, S_2, \dots)$ в целом

§ 1. Законы нуля и единицы. Верхние и нижние функции

Первой предельной теоремой, которая была доказана для распределения сумм одинаково распределенных независимых случайных величин, был закон больших чисел: $S_n/n \xrightarrow{p} E\xi$. Можно задать вопрос: будет ли *вся траектория* $S_n/n, S_{n+1}/n + 1, \dots$, начиная с некоторого n , с вероятностью, мало отличающейся от 1, близка к $E\xi$. Т.е. будет ли для любого $\varepsilon > 0$

$$\lim_{n \rightarrow \infty} P\left(\sup_{k \geq n} \left| \frac{S_k}{k} - E\xi \right| < \varepsilon\right) = 1? \quad (1)$$

Очевидно, это есть вопрос о сходимости почти наверное или сходимости с вероятностью 1.

Утверждение (1), называемое *усиленным законом больших чисел*, которое будет доказано в этой главе, является частным случаем так называемых *законов нуля и единицы*. В качестве первого такого закона приведем критерий нуля и единицы Бореля—Кантелли.

Теорема 1. Пусть $\{A_n\}_{n=1}^{\infty}$ — последовательность независимых событий на $(\Omega, \mathfrak{F}, P)$, и пусть A есть событие, состоящее в том, что наступит бесконечно много событий A_k , т.е. $A = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k$ (A составлено из ω , принадлежащих бесконечно многим A_k). Тогда, если $\sum_{k=1}^{\infty} P(A_k) < \infty$, то $P(A) = 0$. Если $\sum_{k=1}^{\infty} P(A_k) = \infty$ и A_k независимы, то $P(A) = 1$.

Доказательство. Пусть $\sum P(A_k) < \infty$. Обозначим $\eta = \sum_{k=1}^{\infty} I(A_k)$ число появлений событий A_k . Тогда $E\eta = \sum P(A_k) < \infty$, что, конечно же, означает, что η есть собственная случайная величина: $P(\eta < \infty) = 1 - P(A) = 1$.

Если же $\sum_{k=1}^{\infty} P(A_k) = \infty$, то в силу независимости A_n

$$\begin{aligned} P(A) &= \lim_{n \rightarrow \infty} P\left(\bigcup_{k=n}^{\infty} A_k\right) = \lim_{n \rightarrow \infty} P\left(\Omega - \bigcap_{k=n}^{\infty} \bar{A}_k\right) = 1 - \lim_{n \rightarrow \infty} P\left(\bigcap_{k=n}^{\infty} \bar{A}_k\right) = \\ &= 1 - \lim_{n \rightarrow \infty} \lim_{m \rightarrow \infty} P\left(\bigcap_{k=n}^m \bar{A}_k\right) = 1 - \lim_{n \rightarrow \infty} \prod_{k=n}^{\infty} (1 - P(A_k)). \end{aligned}$$

Воспользовавшись неравенством $\ln(1-x) \leq -x$, получим, что $\prod_{k=n}^{\infty} (1 - P(A_k)) \leq \exp\left\{-\sum_{k=n}^{\infty} P(A_k)\right\}$. Следовательно,

$$\prod_{k=n}^{\infty} (1 - P(A_k)) \leq e^{-\infty} = 0.$$

Замечание 1. Отметим, что, если первое утверждение теоремы о том, что $E\eta < \infty$ влечет $\eta < \infty$ п. н., выглядит тривиальным, то второе, означающее, что $P(\eta < \infty) = 1$ влечет $E\eta < \infty$, оказывается весьма содержательным. Из него следует также, что, если $\eta < \infty$ с вероятностью 1, но $E\eta = \infty$, то A_k с необходимостью зависимы.

Отметим также, что рассуждения, доказывающие первую часть теоремы, уже использовались нами в тех же целях в доказательстве теоремы 6.1.

Предположим, что на $\langle \Omega, \mathfrak{F}, P \rangle$ задана последовательность независимых случайных величин $\{\xi_n\}_{n=1}^{\infty}$. Пусть, как и прежде, $\sigma(\xi_1, \dots, \xi_n)$ есть σ -алгебра, порожденная первыми n случайными величинами ξ_1, \dots, ξ_n , а $\sigma(\xi_n, \dots)$ — σ -алгебра, порожденная случайными величинами $\xi_n, \xi_{n+1}, \xi_{n+2}, \dots$.

Определение 1. Событие A называется *остаточным*, если $A \in \sigma(\xi_n, \dots)$ при любом $n > 0$.

Например, событие $A = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} \{\xi_k > N\}$, состоящее в том, что бесконечно много раз произошло событие $\{\xi_k > N\}$, очевидно, является остаточным.

Теорема 2 (закон нуля и единицы Колмогорова). Если A — остаточное событие, то $P(A) = 0$ или $P(A) = 1$.

Доказательство. Так как A — остаточное событие, то $A \in \sigma(\xi_{n+1}, \dots)$, $n \geq 0$. Поэтому событие A не зависит от σ -алгебры $\sigma(\xi_1, \dots, \xi_n)$ при любом n . Следовательно (см. теорему 3.4), событие A не зависит от σ -алгебры $\sigma(\xi_1, \dots)$. Так как $A \in \sigma(\xi_1, \dots)$, то A не зависит само от себя:

$$P(A) = P(AA) = P(A)P(A).$$

Но это возможно, если только $P(A) = 0$ или 1.

Обозначим $S = \sup\{0, S_1, S_2, \dots\}$, где $S_n = \sum_{k=1}^n \xi_k$. В качестве примера применения доказанной теоремы сформулируем следующее

Следствие 1. Если ξ_k , $k = 1, 2, \dots$, независимы, то либо $P(S = \infty) = 1$, либо $P(S < \infty) = 1$.

Доказательство следует из того, что событие $\{S = \infty\}$ остаточное. Действительно, для любого n

$$\{S = \infty\} = \{\sup(S_{n-1}, S_n, \dots) = \infty\} = \{\sup(0, S_n - S_{n-1}, \dots) = \infty\} \in \sigma(\xi_n, \dots). \quad \blacksquare$$

Другие примеры остаточных событий получим, если рассмотрим для последовательности независимых величин ξ_1, ξ_2, \dots событие $\left\{\text{ряд } \sum_{k=1}^{\infty} \xi_k \text{ сходится}\right\}$.

Теорема 2 означает, что вероятность этого события может быть только 0 или 1.

Если рассмотреть степенной ряд $\sum_{k=0}^{\infty} z^k \xi_k$, где ξ_k независимы, то обнаружим, что радиус сходимости этого ряда $\rho = \limsup_{k \rightarrow \infty} |\xi_k|^{-1/k}$ есть случайная величина, измеримая относительно σ -алгебры $\sigma(\xi_n, \dots)$ при любом n ($\{\rho < x\} \in \sigma(\xi_n, \dots)$, $-\infty \leq x \leq \infty$). Такие случайные величины также называются *остаточными*. Так как в силу сказанного $F_\rho(x) = P(\rho < x) = 0$ или 1, то отсюда следует, что ρ , как и любая другая остаточная случайная величина, с вероятностью 1 равна постоянной.

В предположении, что элементы последовательности $\{\xi_k\}_{k=1}^{\infty}$ не только независимы, но и одинаково распределены, закон нуля и единицы Колмогорова был распространен Сэвиджем и Хьюиттом на более широкий класс событий.

Пусть $\omega = (x_1, x_2, \dots)$ есть элемент выборочного вероятностного пространства $(\mathbb{R}^\infty, \mathfrak{B}^\infty, P)$ для последовательности $\xi = (\xi_1, \xi_2, \dots)$ (\mathbb{R}^∞ — счетное прямое произведение прямых \mathbb{R}_k , $k = 1, 2, \dots$; $\mathfrak{B}^\infty = \sigma(\xi_1, \dots)$ порождена событиями $\prod_{k=1}^N B_k \subset \sigma(\xi_1, \dots, \xi_N)$, где $B_k \in \sigma(\xi_k)$ — борелевские множества на прямой \mathbb{R}_k).

Определение 2. Событие $A \in \mathfrak{B}^\infty$ назовем *перестановочным*, если

$$(x_1, x_2, \dots, x_{n-1}, x_n, x_{n+1}, \dots) \in A$$

влечет за собой $(x_n, x_2, \dots, x_{n-1}, x_1, x_{n+1}, \dots) \in A$.

Очевидно, это условие перестановочности автоматически распространяется на любые перестановки конечного числа координат. Примером перестановочных событий могут служить остаточные события.

Теорема 3 (закон нуля и единицы для перестановочных событий). Если ξ_k независимы и одинаково распределены и A есть перестановочное событие, то либо $P(A) = 0$, либо $P(A) = 1$.

Доказательство. В силу теоремы аппроксимации (§ 3.5) для любого $A \in \mathfrak{B}^\infty$ найдется последовательность $A_n \in \sigma(\xi_1, \dots, \xi_n)$ такая, что при $n \rightarrow \infty$

$$P(\bar{A}_n A \cup \bar{A} A_n) \rightarrow 0.$$

Введем преобразование

$$T_n \omega = T_n(x_1, x_2, \dots) = (x_{n+1}, \dots, x_{2n}, x_1, \dots, x_n, x_{2n+1}, \dots)$$

и обозначим $B_n = T_n A_n$. Если A перестановочно, то $T_n A = A$, и для любого $B \in \mathfrak{B}^\infty$ $P(T_n B) = P(B)$. Поэтому $P(B_n A) = P(T_n A_n A) = P(A_n A)$, и, следовательно, B_n будут также аппроксимировать A . Но вместе с B_n и A_n свойством аппроксимации будут обладать, очевидно, и множества $C_n = A_n B_n$. В силу независимости A_n и B_n это означает, что

$$P(A) = \lim_{n \rightarrow \infty} P(A_n B_n) = \lim_{n \rightarrow \infty} P^2(A_n) = P^2(A). \quad \blacksquare$$

Из этой теоремы вытекает следующий интересный факт.

Определение 3. Последовательность $\{a_n\}_{n=1}^{\infty}$ называется *верхней* для последовательности случайных величин $\{\eta_n\}_{n=1}^{\infty}$, если события $\{\eta_n > a_n\}$ с вероятностью 1 наступают конечное число раз. Последовательность $\{a_n\}$ называется *нижней*, если с вероятностью 1 происходит бесконечно много событий $\{\eta_n > a_n\}$.

Следствие 2. Если ξ_n независимы и одинаково распределены, то любая последовательность $\{a_n\}$ является либо верхней, либо нижней для последовательности сумм

$$\{S_n\}_{n=1}^{\infty}, \quad S_n = \sum_{k=1}^n \xi_k.$$

Другими словами, нельзя подобрать такую «промежуточную» последовательность $\{a_n\}$, чтобы вероятность события $A = \{\text{бесконечно много раз } S_n > a_n\}$ была бы, скажем, равна $1/2$.

Для доказательства следствия достаточно заметить, что событие A является перестановочным, так как перестановка местами ξ_1 и ξ_n в реализации (ξ_1, ξ_2, \dots) отражается на поведении лишь первых n сумм S_1, \dots, S_n .

Аналогичный факт, конечно, имеет место и для последовательности самих случайных величин $\{\xi_n\}_{n=1}^{\infty}$, но его в отличие от приведенного следствия можно получить проще, так как событие $B = \{\text{бесконечно много раз } \xi_n > a_n\}$ является остаточным.

По поводу свойств верхних и нижних последовательностей для сумм $\{S_n\}$ заметим здесь также следующее. Если $P(\xi_k = c) \neq 1$ и если $\{a_n\}$ является верхней (нижней) последовательностью для $\{S_n\}$, то последовательность $\{b_n = a_{n+k} - t\}_{n=1}^{\infty}$ также является верхней (нижней) при любых $k \geq 0$ и t . При достаточно больших $k \geq k(t)$ это следует из равенств вида¹⁾

$$\begin{aligned} 0 = P(A) &= \sum_{j=-\infty}^{\infty} P(S_k \in [j, j+1)) P(A | S_k \in [j, j+1)) \geq \\ &\geq \sum P(S_k \in [j, j+1)) P(\text{бесконечно много раз } S_n + j > a_{n+k}). \end{aligned}$$

При $k \leq k(t)$ надо, допустив противное, применить доказанную часть утверждения к нижним функциям.

Из сделанного замечания вытекает, в частности, что $\limsup_{n \rightarrow \infty} S_n$ и $\liminf_{n \rightarrow \infty} S_n$ для последовательности сумм одинаково распределенных независимых величин, не равных 0 с вероятностью 1, не могут быть конечными.

В самом деле, событие $B = \{\limsup_{n \rightarrow \infty} S_n \in (a, b)\}$ является перестановочным и, следовательно, в силу закона 0 и 1, $P(B) = 0$ или $P(B) = 1$. Если бы $P(B)$ равнялась 1, то последовательность (b, b, \dots) была бы верхней для $\{S_n\}$, а вместе с ней была бы верхней и последовательность (a, a, \dots) . Это означало бы, что

$$P(\limsup_{n \rightarrow \infty} S_n \leq a) = 1$$

и что предположение $P(B) = 1$ противоречиво.

¹⁾ Пользуясь у. м. о., можно было бы записать точнее: $P(A) = E P(A | S_k) = \int dP(S_k < t) P(\text{бесконечно много раз } S_n + t > a_{n+k})$.

Из теоремы 3 читатель может извлечь также, что для любых последовательностей $\{a_n\}$ и $\{b_n\}$ случайные величины

$$\limsup_{n \rightarrow \infty} \frac{S_n - a_n}{b_n} \quad \text{и} \quad \liminf_{n \rightarrow \infty} \frac{S_n - a_n}{b_n}$$

с вероятностью 1 постоянны.

§ 2. Сходимость рядов независимых случайных величин

В этом параграфе остановимся более подробно на сходимости рядов, составленных из независимых случайных величин. Мы уже знаем, что такие ряды сходятся с вероятностью 1 или 0. Нас будут интересовать условия сходимости.

Получим, в частности, ответ на следующий интересный вопрос. Хорошо известно, что ряд $\sum_{n=1}^{\infty} \frac{1}{n^\alpha}$ при $\alpha \leq 1$ расходится, а знакопередающийся ряд $\sum_{n=1}^{\infty} \frac{(-1)^n}{n^\alpha}$ сходится при любом $\alpha > 0$ (разность соседних элементов имеет порядок малости $\alpha/n^{1+\alpha}$). Как будет вести себя ряд $\sum_{n=1}^{\infty} \frac{\delta_n}{n^\alpha}$, где δ_n одинаково распределены, независимы, $E\delta_n = 0$ (например, $\delta_n = \pm 1$ с вероятностью $1/2$)?

Одним из основных путей исследования этих задач состоит в выяснении связи сходимости п. н. с более простым понятием сходимости по вероятности. Известно, что, вообще говоря, из сходимости по вероятности $\xi_n \xrightarrow{p} \xi$ не следует сходимость п. н. Однако в рассматриваемом случае, когда $\xi_n = S_n = \sum_{k=1}^n \eta_k$ и η_k независимы, это не так. Основным утверждением в этом параграфе является

Теорема 4. Если η_j независимы, $S_n = \sum_{k=1}^n \eta_k$, то сходимость S_n по вероятности влечет сходимость S_n п. н.

Здесь для определенности под сходимостью будем понимать фундаментальность, не предполагая заранее, что сумма ряда известна. Для доказательства теоремы нам понадобится следующее неравенство:

Лемма 1 (первое неравенство Колмогорова). Если η_j независимы и при всех $j \leq n$

$$P(|S_n - S_j| \geq a) \leq p < 1,$$

то

$$P\left(\max_{j \leq n} |S_j| \geq x\right) \leq \frac{1}{1-p} P(|S_n| > x - a). \quad (2)$$

Следствие 3.

$$P\left(\max_{j \leq n} |S_j| \geq x\right) \leq 2P(|S_n| > x - \sqrt{2DS_n}).$$

Собственно, Колмогоровым было установлено имеет это неравенство (лемма 1 является его незначительным обобщением). Оно вытекает из (2), так как в силу неравенства Чебышева

$$P(|S_n - S_j| \geq \sqrt{2DS_n}) \leq \frac{D(S_n - S_j)}{2DS_n} \leq \frac{1}{2}.$$

Доказательство леммы 1. Пусть $\tau = \{\min k \geq 1: |S_k| \geq x\}$. Обозначим $A_j = \{\tau = j\}$, $j = 1, 2, \dots$. Очевидно, что A_j несовместны и, значит,

$$P(|S_n| > x - a) \geq \sum_{j=1}^n P(|S_n| > x - a; A_j) \geq \sum_{j=1}^n P(|S_n - S_j| < a; A_j).$$

(Последнее неравенство верно, так как событие $\{|S_n - S_j| < a\}A_j$ влечет за собой $\{|S_n| > x - a\}A_j$.) Но $A_j \in \sigma(\eta_1, \dots, \eta_j)$, $\{|S_n - S_j| < a\} \in \sigma(\eta_{j+1}, \dots, \eta_n)$. Поэтому эти два события независимы,

$$\begin{aligned} P(|S_n| > x - a) &\geq \sum_{j=1}^n P(A_j)P(|S_n - S_j| < a) \geq \\ &\geq (1 - p) \sum_{j=1}^n P(A_j) = (1 - p)P\left(\max_{j \leq n} |S_j| \geq x\right). \end{aligned}$$

Лемма доказана. ■

Обратимся к доказательству теоремы. Достаточно доказать п. н. фундаментальность $\{S_n\}$, т. е. сходимость для любого $\varepsilon > 0$ при $m \rightarrow \infty$

$$P\left(\sup_{n \geq m} |S_n - S_m| > 2\varepsilon\right) \rightarrow 0. \quad (3)$$

Обозначим

$$A_{n,m}^\varepsilon = \{|S_n - S_m| > \varepsilon\}, \quad A_m^\varepsilon = \bigcup_{n \geq m} A_{n,m}^\varepsilon.$$

Тогда соотношение (3) можно записать в виде

$$P(A_m^{2\varepsilon}) \rightarrow 0 \quad (4)$$

при $m \rightarrow \infty$.

Так как $\{S_n\}$ фундаментальна по вероятности, то

$$p_{m,M} = \sup_{m \leq n \leq M} P(A_{n,m}^\varepsilon) \rightarrow 0$$

при $m \rightarrow \infty$, $M \rightarrow \infty$, так что $p_{m,M} < 1/2$ при всех достаточно больших m и M . Для таких m и M имеем по лемме 1 при $a = \varepsilon$, $x = 2\varepsilon$

$$P\left(\sup_{m < n \leq M} |S_n - S_m| > 2\varepsilon\right) = P\left(\bigcup_{n=m+1}^M A_{m,n}^{2\varepsilon}\right) \leq \frac{1}{1 - p_{m,M}} P(A_{M,m}^\varepsilon) \leq 2P(A_{M,m}^\varepsilon).$$

В силу свойств вероятности

$$P(A_m^{2\varepsilon}) = \lim_{M \rightarrow \infty} P\left(\bigcup_{n=m+1}^M A_{m,n}^{2\varepsilon}\right) \leq 2 \lim_{M \rightarrow \infty} \sup P(A_{M,m}^\varepsilon). \quad (5)$$

Обозначим через S предел (по вероятности) последовательности S_n , $B_n^\varepsilon = \{|S_n - S| > \varepsilon\}$. Тогда $P(B_n^\varepsilon) \rightarrow 0$ при $n \rightarrow \infty$, $A_{M,m} \subset B_M^{\varepsilon/2} \cup B_m^{\varepsilon/2}$, и в силу (5)

$$P(A_m^{2\varepsilon}) \leq 2P(B_m^{\varepsilon/2}) \rightarrow 0 \quad \text{при } m \rightarrow \infty.$$

Соотношение (4), а вместе с ним и утверждение теоремы доказаны. ■

Следствие 4. Если $E\eta_k = 0$, $\sum_{k=1}^{\infty} D\eta_k < \infty$, то S_n п. н. сходится.

Утверждение следует из теоремы 4 и фундаментальности S_n в среднеквадратичном $(E(S_n - S_m))^2 = \sum_{k=m+1}^n D\eta_k \rightarrow 0$ при $m \rightarrow \infty, n \rightarrow \infty$) и, следовательно, по вероятности.

Оказывается, что если $E\eta_k = 0$, $|\eta_k| < c$ при всех k , то условие $\sum D\eta_k < \infty$ будет и необходимым для сходимости S_n п. н.²⁾

В следствии 4 содержится ответ и на вопрос, поставленный в начале параграфа о сходимости $\sum \frac{\delta_n}{n^\alpha}$, где δ_n одинаково распределены, независимы, $E\delta_n = 0$.

Следствие 5. Ряд $\sum \delta_n a_n$ сходится с вероятностью 1, если $D\delta_k = \sigma^2 < \infty$, $\sum a_n^2 < \infty$.

С учетом сделанного выше замечания можно сказать, следовательно, что ряд $\sum \frac{\delta_n}{n^\alpha}$ при $\delta_n = \pm 1$ с вероятностью 1/2, сходится тогда и только тогда, когда $\alpha > 1/2$. Некоторым обобщением следствия 4 является

Следствие 6 (критерий двух рядов). Для сходимости ряда $\sum \eta_n$ п. н. достаточно, чтобы сходились ряды $\sum E\eta_n$, $\sum D\eta_n$.

Доказательство очевидно, так как последовательности $\sum_{k=1}^n E\eta_k$ и $\sum_{k=1}^n (\eta_k - E\eta_k)$ сходятся п. н. по следствию 4. ■

§ 3. Усиленный закон больших чисел

Нетрудно видеть, что если использовать терминологию §1, то усиленный закон больших чисел (1) означает, что последовательность $\{\varepsilon n\}_{n=1}^{\infty}$ при любом $\varepsilon > 0$ является верхней для обеих последовательностей $\{S_n\}$ и $\{-S_n\}$, если только $E\xi_1 = 0$.

Сформулируем предварительно в виде леммы следующее простое утверждение.

Лемма 2. $E|\xi| < \infty$ тогда и только тогда, когда

$$\sum_{j=1}^{\infty} P(|\xi| > j) < \infty.$$

Доказательство следует очевидным образом из равенства

$$E|\xi| = \int_0^{\infty} P(|\xi| > x) dx$$

и неравенств

$$\sum_{j=1}^{\infty} P(|\xi| > j) \leq \int_0^{\infty} P(|\xi| > x) dx \leq 1 + \sum_{j=1}^{\infty} P(|\xi| > j).$$

Усиленный закон больших чисел получим в качестве следствия одного факта, который представляет и самостоятельный интерес.

²⁾ Подробнее см. об этом, например, [25].

Теорема 5. Если ξ_i одинаково распределены и независимы, $E\xi_1 > 0$, то случайная величина $Z = \inf_{k \geq 0} S_k$ является собственной (с вероятностью 1 конечной) :

Доказательство. Пусть $\eta_1 = \eta(1) + 1 = \nu(1)$ (см. обозначения гл. 9) есть номер первой суммы S_k , которая достигла уровня 1. Рассмотрим последовательность $\{\xi_k^* = \xi_{\eta_1+k}\}$, которая в силу леммы 9.1 распределена так же, как $\{\xi_k\}$, и не зависит от $\eta_1, \xi_1, \dots, \xi_{\eta_1}$. Для нее построим индекс $k = \eta_2$, при котором впервые сумма $S_k^* = \sum_{j=1}^k \xi_j^*$ достигла уровня 1. Ясно, что случайные величины η_1 и η_2 одинаково распределены и независимы. Затем для последовательности $\{\xi_k^{**} = \xi_{\eta_1+\eta_2+k}\}$ построим величину η_3 и т.д. В результате получим последовательность марковских моментов η_1, η_2, \dots , которые определяют моменты «регенерации» исходной последовательности $\{S_k\}$, связанные с достижением уровня 1.

Обозначим теперь $Z_1 = \inf_{k < \eta_1} S_k$, $Z_2 = \inf_{k < \eta_2} S_k^*, \dots$. Ясно, что Z_j одинаково распределены и что

$$Z = \inf\{Z_1, S_{\eta_1} + Z_2, S_{\eta_1+\eta_2} + Z_3, \dots\},$$

где по определению $S_{\eta_1} > 1$, $S_{\eta_1+\eta_2} > 2$ и т.д. Поэтому

$$\begin{aligned} \{Z < -N\} &= \bigcup_{k=0}^{\infty} \{Z_{k+1} + S_{\eta_1+\dots+\eta_k} < -N\} \subset \bigcup_{k=0}^{\infty} \{Z_k + k < -N\}, \\ P(Z < -N) &\leq \sum_{k=1}^{\infty} P(Z_k + k < -N) = \sum_{j=N+1}^{\infty} P(Z_1 < -j). \end{aligned}$$

Это выражение сходится к 0 при $N \rightarrow \infty$, если $E|Z_1| < \infty$ (см. лемму 1). Конечность EZ_1 , в которой нам осталось убедиться, следует из конечности $E\eta_1 = E\eta(1)$ (см. теоремы гл. 9) и неравенств

$$E|Z_1| \leq E \sum_{j=1}^{\eta_1} |\xi_j| = E\eta_1 E|\xi_1| < \infty.$$

■

Пусть, как и прежде, ξ_1, ξ_2, \dots независимы и одинаково распределены.

Теорема 6 (усиленный закон больших чисел Колмогорова). Для того, чтобы $S_n/n \xrightarrow{п.н.} a$, необходимо и достаточно, чтобы существовало $E\xi_k = a$.

Доказательство. Достаточность. Предположим, не ограничивая общности, что $E\xi_k = 0$. Тогда из теоремы 5 вытекает, что случайная величина $Z^{(\varepsilon)} = \inf_{k \geq 0} (S_k + \varepsilon k)$ является собственной при любом $\varepsilon > 0$ ($S_k + \varepsilon k$ есть суммы случайных величин $\xi_k + \varepsilon$, $E(\xi_k + \varepsilon) > 0$). Следовательно,

$$P\left(\inf_{k \geq n} \frac{S_k}{k} < -2\varepsilon\right) \leq P\left(\bigcup_{k \geq n} \{S_k + \varepsilon k < -\varepsilon n\}\right) \leq P(Z^{(\varepsilon)} < -\varepsilon n) \rightarrow 0$$

при $n \rightarrow \infty$. Аналогично обнаруживаем, что

$$P\left(\sup_{k \geq n} \frac{S_k}{k} > 2\varepsilon\right) \rightarrow 0 \quad \text{при} \quad n \rightarrow \infty.$$

Так как $P\left(\sup_{k \geq n} \left| \frac{S_k}{k} \right| > 2\varepsilon\right)$ не превосходит суммы двух рассмотренных выше вероятностей, то мы получим, что $\frac{S_n}{n} \xrightarrow{\text{п. н.}} 0$.

Необходимость. Заметим, что $\frac{\xi_n}{n} = \frac{S_n}{n} - \frac{n-1}{n} \frac{S_{n-1}}{n-1} \xrightarrow{\text{п. н.}} 0$, т. е. событие $\left\{\left|\frac{\xi_n}{n}\right| > 1\right\}$ наступает конечное число раз с вероятностью 1. По критерию «нуля или единицы» Бореля—Кантелли это означает, что $\sum_{n=1}^{\infty} P\left(\left|\frac{\xi_n}{n}\right| > 1\right) < \infty$ или, что то же, $\sum P(|\xi_1| > n) < \infty$. Следовательно, по лемме 1 $E\xi_1 < \infty$. ■

Таким образом, условие $E\xi_k = 0$ является необходимым и достаточным для того, чтобы последовательность $\{\varepsilon n\}_{n=1}^{\infty}$ была верхней для обеих последовательностей $\{S_n\}$ и $\{-S_n\}$. В следующей главе мы получим в качестве следствий необходимые и достаточные условия для того, чтобы $\{\varepsilon n\}$ была верхней последовательностью для каждой из траекторий $\{S_n\}$ и $\{-S_n\}$ в отдельности. Разумеется, что, например, для последовательности $\{S_n\}$ это условие будет шире, чем условие $E\xi_k = 0$.

Замечание 2. Мы видели, что в основе приведенного доказательства усиленного закона больших чисел лежит тождество Вальда. Теперь, пользуясь доказанным законом, можно несколько усилить утверждение тождества. Именно, имеет место следующий факт. Если $a = E\xi \neq 0$, то для марковского момента ν условие $E\nu < \infty$ для выполнения тождества Вальда

$$ES_{\nu} = aE\nu$$

излишне. Другими словами, это тождество верно и при $E\nu = \infty$, или, что то же, конечность ES_{ν} влечет за собой конечность $E\nu$. Если $a = 0$, то последнее утверждение неверно. Читателю нетрудно проиллюстрировать этот факт на примере безобидной игры, рассмотренной в § 4.2.

Для доказательства приведенного утверждения достаточно убедиться, что при $a > 0$ равенство $E\nu = \infty$ влечет за собой $ES_{\nu} = \infty$.

Заметим сначала, что для любого марковского момента $ES_{\nu} \geq 0$. Допустив, что $ES_{\nu} < 0$, мы построим так же, как в доказательстве теоремы 4, последовательность независимых моментов остановок ν_1, ν_2, \dots таких, что $S_{\nu_1}, S_{\nu_2} - S_{\nu_1}, \dots$ будут независимы и распределены как S_{ν} . По усиленному закону больших чисел это будет означать, что с вероятностью 1 все S_{ν_N} , начиная с некоторого N , будут отрицательны. Но это, очевидно, противоречит усиленному закону больших чисел, примененному к S_n , $n = 1, 2, \dots$, поскольку $E\xi_1 = a > 0$.

Далее, нетрудно видеть, что $\nu_N = \min(N, \nu)$ вместе с ν есть марковская величина. Имеем

$$\begin{aligned} ES_{\nu} &= E\{S_{\nu}; \nu \leq N\} + E\{S_{\nu}; \nu > N\} = \\ &= ES_{\nu_N} + E\{S_{\nu} - S_{\nu_N}; \nu > N\} = E\nu_N \cdot a + ES_{\nu}^*, \end{aligned}$$

где $S_k^* = \xi_{N+1} + \dots + \xi_{N+k}$, $S_0^* = 0$, $\nu^* = \max(0, \nu - N)$ есть марковская величина относительно последовательности $\{\xi_{N+1}, \xi_{N+2}, \dots\}$. На основании предыдущего получаем

$$ES_{\nu} \geq E\nu_N \cdot a.$$

Так как $E\nu_N \geq E\{\nu; \nu \leq N\} \rightarrow \infty$ при $N \rightarrow \infty$, то утверждение доказано.

Для иллюстрации усиленного закона больших чисел рассмотрим следующий пример.

Пример 1. Пусть $\omega = (\varepsilon_1, \varepsilon_2, \dots)$ есть последовательность независимых случайных величин, принимающих значения 1 и 0 соответственно с вероятностями p и $q = 1 - p$. Поставим в соответствие каждой такой последовательности ω число $\xi = \xi(\omega) = \sum_{k=1}^{\infty} \frac{\varepsilon_k}{2^k}$, так что ω есть двоичное разложение ξ . Ясно, что возможные значения ξ заполняют отрезок $[0, 1]$.

Покажем, что если $p = q = 1/2$, то распределение ξ равномерно. Если же $p \neq 1/2$, то распределение ξ сингулярно. В самом деле, если $x = \sum_{k=1}^{\infty} \delta_k 2^{-k}$, где δ_k принимают значения 0 или 1, то

$$\{\xi < x\} = \{\varepsilon_1 < \delta_1\} \cup \{\varepsilon_1 = \delta_1, \varepsilon_2 < \delta_2\} \cup \{\varepsilon_1 = \delta_1, \varepsilon_2 = \delta_2, \varepsilon_3 < \delta_3\} \cup \dots$$

Так как события в этом объединении не пересекаются, то при $p = 1/2$

$$\begin{aligned} P(\xi < x) &= \sum_{k=0}^{\infty} P(\varepsilon_1 = \delta_1, \dots, \varepsilon_k = \delta_k, \varepsilon_{k+1} < \delta_{k+1}) = \\ &= \sum_{k=0}^{\infty} 2^{-k} P(\varepsilon_{k+1} < \delta_{k+1}) = \sum_{k=0}^{\infty} 2^{-k-1} \delta_{k+1} = x. \end{aligned}$$

Это означает, что распределение ξ равномерно, т.е. для любого измеримого по Борелю множества $B \subset [0, 1]$ определена вероятность $P(\xi \in B) = \text{mes } B$, равная лебеговой мере B . Обозначим $\sigma_n = \sum_{k=1}^n \delta_k$, $s_n = \sum_{k=1}^n \varepsilon_k$. Тогда множество $\left\{x: \lim_{n \rightarrow \infty} \frac{\sigma_n}{n} = p\right\}$ измеримо по Борелю и, стало быть,

$$\text{mes}\left\{x: \lim_{n \rightarrow \infty} \frac{\sigma_n}{n} = \frac{1}{2}\right\} = P\left(\lim_{n \rightarrow \infty} \frac{s_n}{n} = \frac{1}{2}\right).$$

Так как по усиленному закону больших чисел правая часть здесь равна 1, то

$$\text{mes}\left\{x: \lim_{n \rightarrow \infty} \frac{\sigma_n}{n} = \frac{1}{2}\right\} = 1.$$

Другими словами, для почти всех $x \in [0, 1]$ доля единиц в двоичном разложении x равна $1/2$.

Пусть теперь $p \neq 1/2$. Тогда $P\left(\lim_{n \rightarrow \infty} \frac{s_n}{n} = p\right) = 1$, хотя в силу предыдущего

$$\text{mes}\left\{x: \lim_{n \rightarrow \infty} \frac{\sigma_n}{n} = p\right\} = 0,$$

так что «носителем» вероятностной меры на $[0, 1]$ является множество лебеговой меры 0. С другой стороны, распределение случайной величины ξ при $p \neq 1/2$ непрерывно. Это следует из того, что

$$\{\xi = x\} = \bigcap_{k=1}^{\infty} \{\varepsilon_k = \delta_k\},$$

если x двоично-иррационально.

Если x — двоично-рационально, т.е. если при некотором $r < \infty$ все δ_k при $k \geq r$ равны 0 или 1, то непрерывность следует из вложения

$$\{\xi = x\} \subset \bigcap_{k=r}^{\infty} \{\varepsilon_k = 0\} + \bigcap_{k=r}^{\infty} \{\varepsilon_k = 1\},$$

так как вероятности этих событий, очевидно, равны 0. Сингулярность $F_\xi(x)$ при $p \neq 1/2$ доказана. ■

Предлагаем читателю построить график функции распределения ξ .

§ 4. Усиленный закон больших чисел для произвольных независимых слагаемых

Отыскание необходимых и достаточных условий для сходимости

$$\frac{S_n}{b_n} \xrightarrow{\text{п. н.}} a,$$

когда $b_n \uparrow \infty$, а слагаемые ξ_1, ξ_2, \dots распределены не одинаково, представляет собой трудную задачу. Мы ограничимся здесь тем, что докажем следующую теорему.

Теорема 7 (критерий Колмогорова сходимости почти всюду). Пусть ξ_k , $k = 1, 2, \dots$, независимы, $E\xi_k = 0$, $D\xi_k = \sigma_k^2 < \infty$ и, кроме того,

$$\sum_{k=1}^{\infty} \frac{\sigma_k^2}{b_k^2} < \infty. \quad (6)$$

Тогда $S_n/b_n \xrightarrow{\text{п. н.}} 0$.

Доказательство. Из условий теоремы 7 вытекает, что (см. следствие 4) ряд $\sum_{k=1}^{\infty} \frac{\xi_k}{b_k}$ с вероятностью 1 сходится. Поэтому утверждение теоремы 7 будет следствием следующей известной леммы анализа.

Лемма 3. Пусть $b_n \uparrow \infty$, а последовательность x_1, x_2, \dots такова, что ряд $\sum_{k=1}^{\infty} x_k$ сходится. Тогда при $n \rightarrow \infty$

$$\frac{1}{b_n} \sum_{k=1}^{\infty} b_k x_k \rightarrow 0.$$

Доказательство. Обозначим $X_n = \sum_{k=1}^n x_k$, так что $X_n \rightarrow 0$ при $n \rightarrow \infty$, $X = \max_{n \geq 0} |X_n| < \infty$. Используя преобразования Абеля, получим

$$\begin{aligned} \sum_{k=1}^{\infty} b_k x_k &= \sum_{k=1}^n b_k (X_{k-1} - X_k) = \sum_{k=0}^{n-1} b_{k+1} X_k - \sum_{k=1}^n b_k X_k = \\ &= \sum_{k=1}^{n-1} (b_{k+1} - b_k) X_k + b_1 X_0 - b_n X_n, \\ \limsup_{n \rightarrow \infty} \frac{1}{b_n} \sum_{k=1}^n b_k x_k &\leq \limsup_{n \rightarrow \infty} \frac{1}{b_n} \sum_{k=1}^{n-1} (b_{k+1} - b_k) X_k. \end{aligned} \quad (7)$$

Здесь для заданного $\varepsilon > 0$ можно выбрать N так, что $|X_k| < \varepsilon$ при $k \geq N$. Поэтому

$$\sum_{k=1}^{n-1} (b_{k+1} - b_k) X_k \leq \sum_{k=1}^{N-1} (b_{k+1} - b_k) X_k + \varepsilon \sum_{k=N}^{n-1} (b_{k+1} - b_k) = X(b_N - b_1) + \varepsilon(b_n - b_N).$$

Отсюда и из (7) следует, что

$$\limsup_{n \rightarrow \infty} \frac{1}{b_n} \sum_{k=1}^n b_k x_k \leq \varepsilon.$$

Можно было бы доказывать теорему 7 и непосредственно, используя неравенство Колмогорова примерно так же, как это делалось в теореме 4.

Первоначально достаточность в теореме 6 была доказана с помощью теоремы 7 путем использования срезок

$$\xi'_k = \begin{cases} \xi_k & \text{при } |\xi_k| < k, \\ 0 & \text{при } |\xi_k| \geq k, \end{cases}$$

которые имеют дисперсии $D\xi'_k$, удовлетворяющие условию $\sum_{k=1}^{\infty} \frac{D\xi'_k}{k^2} < \infty$, и которые с точки зрения закона больших чисел не отличаются от ξ_k , так как $\sum_{k=1}^{\infty} P(\xi'_k \neq \xi_k) < \infty$ и, следовательно, происходит лишь конечное число событий $\{\xi'_k \neq \xi_k\}$.

Пример 2. Пусть ξ_k , $k = 1, 2, \dots$, — независимые случайные величины, принимающие значения $\xi_k = \pm k^\alpha$ с вероятностями $1/2$. Как мы видели в примере 1 § 8.4 при $\alpha \geq -1/2$ суммы S_n этих величин асимптотически нормальны при соответствующей нормировке. Будут ли они удовлетворять усиленному закону больших чисел? Так как $D\xi_k = \sigma_k^2 = k^{2\alpha}$, то ряд $\sum \frac{D\xi_k}{k^2} = \sum k^{2(\alpha-1)}$ будет сходиться при $\alpha < 1/2$. Таким образом, если $\alpha < 1/2$, то $S_n/n \xrightarrow{\text{п.н.}} 0$. Отметим теперь, что, хотя условие $\alpha < 1/2$ есть всего лишь достаточное, оно точно определяет границы применимости усиленного закона больших чисел. Действительно, $B_n^2 = DS_n \sim n^2/2$ при $\alpha = 1/2$ и, следовательно, величина S_n/n будет асимптотически нормальной с параметрами $(0, 1/\sqrt{2})$ (пример 1 § 8.4). Итак S_n при $\alpha \geq 1/2$ не будет удовлетворять усиленному закону больших чисел, хотя и будет удовлетворять центральной предельной теореме. Этот факт объясняется, очевидно, существенно различными нормировками в этих предельных теоремах.

Г л а в а 11

Факторизационные тождества

В этой главе мы получим ряд замечательных и весьма полезных соотношений, устанавливающих связи между различными характеристиками случайных блужданий, а также связи между возникающими при этом проблемами и простейшими граничными задачами теории функций комплексного переменного.

§ 1. Факторизационные тождества и их первые следствия

Пусть $\langle \Omega, \mathfrak{F}, P \rangle$ — вероятностное пространство, на котором задана последовательность $\{\xi_k\}_{k=1}^{\infty}$ независимых одинаково распределенных случайных величин.

Пусть, как и прежде, $S_n = \sum_{k=1}^n \xi_k$, $S_0 = 0$.

На этом же вероятностном пространстве определим случайную величину

$$\eta_+^0 = \min\{k \geq 1; S_k \geq 0\}.$$

Это есть *время первого достижения нулевого уровня*. На тех элементарных исходах ω , для которых все $S_k < 0$, $k \geq 1$, положим $\eta_+^0 = \infty$. Как и $\nu(0) = \eta(0) + 1$ в § 9.1, величина η_+^0 является марковской.

Случайная величина $\chi_+^0 = S_{\eta_+^0}$ называется *первой неотрицательной суммой*. Она определена лишь на множестве $\{\eta_+^0 < \infty\}$.

Совершенно аналогичными свойствами обладают *время достижения нулевого уровня справа*

$$\eta_-^0 = \min\{k \geq 1; S_k \leq 0\}$$

и *первая неположительная сумма* $\chi_-^0 = S_{\eta_-^0}$.

Исследование свойств введенных случайных величин представляет значительный самостоятельный интерес. Величина η_+^0 , например, есть момент остановки, и знание ее природы весьма существенно для изучения моментов остановки во многих более сложных задачах (см. задачи теории восстановления в гл. 9, задачи о статистическом контроле, описанные в § 4.4, и др.). Кроме того, величины η_+^0 и χ_+^0 потребуются для описания поведения экстремумов

$$\zeta = \sup(S_1, S_2, \dots) \quad \text{и} \quad \gamma = \inf(S_1, S_2, \dots),$$

которые играют важную роль в задачах математической статистики, теории массового обслуживания (см. § 4) и других.

Обозначим $\varphi(\lambda) = \varphi_{\xi_1}(\lambda) = E e^{i\lambda \xi_1}$.

Теорема 1 (факторизационное тождество). Для х. ф. совместных распределений введенных случайных величин при $|z| < 1$, $\text{Im } \lambda = 0$ справедливо тождество

$$1 - z\varphi(\lambda) = [1 - E(e^{i\lambda\chi_+^0} z^{\eta_+^0}; \eta_+^0 < \infty)] D^{-1}(z) [1 - E(e^{i\lambda\chi_-^0} z^{\eta_-^0}; \eta_-^0 < \infty)],$$

где

$$D(z) = 1 - E(z^{\eta_+^0}; \chi_+^0 = 0, \eta_+^0 < \infty) = 1 - E(z^{\eta_-^0}; \chi_-^0 = 0; \eta_-^0 < \infty).$$

Доказательство. Обозначим $\zeta_n = \max\{S_1, \dots, S_n\}$. На множестве $\{\eta_+^0 = k\}$ имеем

$$S_n = S_{\eta_+^0} + S_n - S_{\eta_+^0} = \chi_+^0 + S_n - S_k.$$

Поэтому

$$\begin{aligned} \varphi^n(\lambda) &= E e^{i\lambda S_n} = \sum_{k=1}^n E(e^{i\lambda S_n}; \eta_+^0 = k) + E(e^{i\lambda S_n}; \zeta_n < 0) = \\ &= \sum_{k=1}^n E(e^{i\lambda(S_n - S_k)} e^{i\lambda S_k} I(\eta_+^0 = k)) + M_n, \end{aligned} \quad (1)$$

где $M_n = E(e^{i\lambda S_n}; \zeta_n < 0)$. При каждом фиксированном k случайные величины $S_n - S_k$ и $S_k I(\eta_+^0 = k) = \chi_+^0 I(\eta_+^0 = k)$, где $I(A)$ есть индикатор события A , являются независимыми. Следовательно,

$$\varphi^n(\lambda) = \sum_{k=1}^n \varphi^{n-k}(\lambda) E(e^{i\lambda\chi_+^0}; \eta_+^0 = k) + M_n.$$

Умножим теперь обе части равенства на z^n , $n = 0, 1, \dots$, и просуммируем по n . При этом условимся считать, что при $n = 0$

$$\sum_{k=1}^n = 0, \quad M_n = 1.$$

Для свертки двух последовательностей $c_n = \sum_{k=1}^n a_k b_{n-k}$ справедливо равенство

$$\sum_{n=0}^{\infty} c_n z^n = \sum_{n=1}^{\infty} a_n z^n \sum_{n=0}^{\infty} b_n z^n,$$

если только ряды, фигурирующие в этом равенстве, абсолютно сходятся. Так как $|z| < 1$, $|\varphi(\lambda)| \leq 1$ при $\text{Im } \lambda = 0$, то

$$\begin{aligned} \sum_{n=0}^{\infty} z^n \varphi^n(\lambda) &= \frac{1}{1 - z\varphi(\lambda)} = \sum_{n=0}^{\infty} z^n \sum_{k=1}^n \varphi^{n-k}(\lambda) E(e^{i\lambda\chi_+^0}; \eta_+^0 = k) + \sum_{n=0}^{\infty} z^n M_n = \\ &= \sum_{k=1}^{\infty} z^k E(e^{i\lambda\chi_+^0}; \eta_+^0 = k) \sum_{n=0}^{\infty} z^n \varphi^n(\lambda) + \sum_{n=0}^{\infty} z^n M_n = \\ &= \frac{1}{1 - z\varphi(\lambda)} E(e^{i\lambda\chi_+^0} z^{\eta_+^0}; \eta_+^0 < \infty) + \sum_{n=0}^{\infty} z^n M_n, \end{aligned}$$

или, что то же,

$$1 - z\varphi(\lambda) = \frac{1 - E(e^{i\lambda\chi_+^0} z^{\eta_+^0}; \eta_+^0 < \infty)}{\sum_{n=0}^{\infty} z^n E(e^{i\lambda S_n}; \zeta_n < 0)} = \frac{A_{z+}(\lambda)}{A_{z-}(\lambda)}, \quad (2)$$

где $A_{z\pm}(\lambda)$ означают числитель и знаменатель отношения, полученного для $1 - z\varphi(\lambda)$.

Легко понять, что если мы обозначим

$$\gamma_n = \min(S_1, \dots, S_n),$$

то, повторив приведенные рассуждения, придем к равенству

$$1 - z\varphi(\lambda) = \frac{1 - E(e^{i\lambda\chi_+^0} z^{\eta_+^0}; \eta_+^0 < \infty)}{\sum_{n=0}^{\infty} z^n E(e^{i\lambda S_n}; \gamma_n > 0)} = \frac{B_{z-}(\lambda)}{B_{z+}(\lambda)}. \quad (3)$$

Заметим теперь, что при $|z| < 1$ функции $A_{z+}(\lambda)$ и $B_{z+}(\lambda)$ аналитичны в области $\text{Im } \lambda > 0$, непрерывны, включая границу $\text{Im } \lambda = 0$, и ограничены при $\text{Im } \lambda \geq 0$. Это следует из того, что обе эти функции представляют собой равномерно сходящиеся ряды из функций, обладающих указанными свойствами. Например,

$$E(e^{i\lambda\chi_+^0} z^{\eta_+^0}; \eta_+^0 < \infty) = \sum_{n=1}^{\infty} z^n E(e^{i\lambda\chi_+^0}; \eta_+^0 = n),$$

где $E(e^{i\lambda\chi_+^0}; \eta_+^0 = n)$ есть часть х. ф. неотрицательной случайной величины (см. свойство 6 х. ф.).

Функции $A_{z-}(\lambda)$ и $B_{z-}(\lambda)$ по тем же причинам обладают аналогичными свойствами в полуплоскости $\text{Im } \lambda \leq 0$. Но если сравнить (2) и (3), то получим, что при $\text{Im } \lambda = 0$,

$$A_{z+}(\lambda)B_{z+}(\lambda) = A_{z-}(\lambda)B_{z-}(\lambda).$$

Это означает, что функция $A_{z-}(\lambda)B_{z-}(\lambda)$ ($A_{z+}(\lambda)B_{z+}(\lambda)$) может быть через границу $\text{Im } \lambda = 0$ из нижней (верхней) полуплоскости продолжена до целой функции. Так как эта функция ограничена, то в силу известных теорем она с необходимостью равна постоянной. Далее, $e^{i\lambda x} \rightarrow 0$ при $\text{Im } \lambda \rightarrow -\infty$, $x < 0$ и, следовательно,

$$B_{z-}(-i\infty) = 1 - E(z^{\eta_-^0}; \chi_-^0 = 0, \eta_-^0 < \infty), \quad A_{z-}(-i\infty) = 1,$$

$$A_{z-}(\lambda)B_{z-}(\lambda) = A_{z-}(-i\infty)B_{z-}(-i\infty) = D(z) = 1 - E(z^{\eta_-^0}; \chi_-^0 = 0, \eta_-^0 < \infty).$$

Подставляя в (2) найденное из этого равенства значение $A_{z-}(\lambda) = D(z)/B_{z-}(\lambda)$, получим утверждение теоремы. Второе соотношение для $D(z)$ найдем из равенства $D(z) = A_{z+}(i\infty)B_{z+}(i\infty)$. ■

При доказательстве теоремы 1 мы в формуле (1) использовали разложение $E e^{i\lambda S_n}$ на слагаемые, соответствующие непересекающимся событиям $\left\{ \bigcup_{k=1}^n \{\eta_+^0 = k\} \right\} = \{\zeta_n \geq 0\}$ и $\{\zeta_n < 0\}$. Но схема доказательства полностью сохранится, если мы рассмотрим разбиение Ω на события $\{\zeta_n > 0\}$ и $\{\zeta_n \leq 0\}$. Чтобы сделать это, введем случайные величины

$$\eta_+ = \min\{k: S_k > 0\} \quad (\eta_+ = \infty, \text{ если } \zeta \leq 0). \text{ Заметим, что } \eta_+ = \nu(0) = \eta(0) + 1$$

в обозначениях § 9.1),

$$\chi_+ = S_{\eta_+},$$

$$\eta_- = \min\{k: S_k < 0\} \quad (\eta_- = \infty, \text{ если } \gamma \geq 0),$$

$$\chi_- = S_{\eta_-}.$$

Величина η_+ (η_-) называется *временем появления первой положительной (отрицательной) суммы* χ_+ (χ_-). Теперь наряду с равенствами (2), (3) сможем записать

$$1 - z\varphi(\lambda) = \frac{1 - E(e^{i\lambda\chi_+} z^{\eta_+}; \eta_+ < \infty)}{\sum_{n=0}^{\infty} z^n E(e^{i\lambda S_n}; \zeta_n \leq 0)} = \frac{1 - E(e^{i\lambda\chi_-} z^{\eta_-}; \eta_- < \infty)}{\sum_{n=0}^{\infty} z^n E(e^{i\lambda S_n}; \gamma_n \geq 0)}. \quad (4)$$

Сопоставляя эти соотношения с (2), (3), мы с помощью прежних рассуждений сможем доказать следующее утверждение.

Теорема 2.

$$1 - E(e^{i\lambda\chi_+} z^{\eta_+}; \eta_+ < \infty) = D(z)[1 - E(e^{i\lambda\chi_+} z^{\eta_+}; \eta_+ < \infty)], \quad (5)$$

$$1 - E(e^{i\lambda\chi_-} z^{\eta_-}; \eta_- < \infty) = D(z)[1 - E(e^{i\lambda\chi_-} z^{\eta_-}; \eta_- < \infty)].$$

Здесь функция $D(z)$, определенная в теореме 1, удовлетворяет также соотношениям

$$D^{-1}(z) = \sum_{n=0}^{\infty} z^n P(S_n = 0, \zeta_n \leq 0) = \sum_{n=0}^{\infty} z^n P(S_n = 0, \gamma_n \geq 0). \quad (6)$$

Ясно, что с помощью теоремы 2 можно получить некоторые другие разновидности факторизационного тождества. Например, такое:

$$1 - z\varphi(\lambda) = [1 - E(e^{i\lambda\chi_+} z^{\eta_+}; \eta_+ < \infty)][1 - E(e^{i\lambda\chi_-} z^{\eta_-}; \eta_- < \infty)]. \quad (7)$$

Из представлений (6) для $D(z)$ вытекает, в частности, что

$$P(S_n = 0, \zeta_n \leq 0) = P(S_n = 0, \gamma_n \geq 0)$$

и что $D(z) \equiv 1$, если $P(S_n = 0) = 0$ при всех $n \geq 1$.

Доказательство теоремы 2. Получим, например, первое из соотношений (5). Сравнивая (3) и (4), как и раньше, находим

$$[1 - E(e^{i\lambda\chi_+} z^{\eta_+}; \eta_+ < \infty)]B_{z+}(\lambda) = 1, \quad (8)$$

так как 1 есть значение рассматриваемого произведения в точке $i\infty$. Поэтому в силу (3) получаем (7). Остается сравнить (7) с тождеством теоремы 1.

Выражения (6) для $D(z)$ получаются, если вспомнить, что в правой части (8) должно стоять

$$\left[\sum_{n=0}^{\infty} z^n E(e^{i\lambda S_n}; \zeta_n \leq 0) \right] [1 - E(e^{i\lambda\chi_-} z^{\eta_-}; \eta_- < \infty)].$$

Так как это произведение также равно 1, то полагая $\lambda = -i\infty$, приходим к одному из требуемых равенств. Второе доказывается аналогично. ■

Замечание 1. Важно отметить, что теоремы 1 и 2 наряду с доказательством существования тождеств дают нам также средство отыскания х.ф. совместных распределений величин χ и η . Речь идет об отождествлении с представлениями в теоремах 1 и 2 любого представления функции $1 - z\varphi(\lambda)$ при $\text{Im } \lambda = 0$ в виде $C_{z+}(\lambda)/C_{z-}(\lambda)$ или $C_{z-}(\lambda)/C_{z+}(\lambda)$, где $C_{z\pm}(\lambda)$ аналитичны соответственно в областях $\text{Im } \lambda \geq 0$, непрерывны включая границу $\text{Im } \lambda = 0$ и ограничены при $\text{Im } \lambda \geq 0$. Если, например,

$$1 - z\varphi(\lambda) = \frac{C_{z+}(\lambda)C(z)}{C_{z-}(\lambda)}$$

и $C_{z+}(i\infty) = C_{z-}(-i\infty) = 1$, то, как мы видели, $1 - E(e^{i\lambda\chi_+} z^{\eta_+}; \chi_+ < \infty)$ можно отождествить с $C_{z+}(\lambda)$; $C_{z-}^{-1}(\lambda)$ с $1 - E(e^{i\lambda\chi_-} z^{\eta_-}; \eta_- < \infty)$, а $C(z)$ с $D(z)$.

В общем случае, если некоторую функцию $C_z(\lambda)$, заданную при $\text{Im}(\lambda) = 0$, удастся представить в виде одного из соотношений $C_{z+}(\lambda)/C_{z-}(\lambda)$ или $C_{z-}(\lambda)/C_{z+}(\lambda)$, где $C_{z\pm}$ обладают указанными выше свойствами, то говорят, что $C_z(\lambda)$ допускает факторизацию. Функция $C_{z+}(\lambda)$ ($C_{z-}(\lambda)$) называется *положительной* (*отрицательной*) *компонентой факторизации*. Задача о факторизации представляет собой разновидность так называемой проблемы Коши—Римана в граничных задачах теории функций комплексного переменного.

Получим теперь некоторые следствия теорем 1 и 2.

Следствие 1. Устремляя $z \rightarrow 1$, мы получим

$$1 - \varphi(\lambda) = [1 - E(e^{i\lambda\chi_+}; \eta_+ < \infty)] [1 - E(e^{i\lambda\chi_-}; \eta_-^0 < \infty)].$$

Аналогичным образом можно, очевидно, написать и другие тождества такого типа, соответствующие тождествам, которые можно получить из теорем 1 и 2.

Следствие 2. Если существует $E\xi_k = a < 0$, то $P(\eta_-^0 < \infty) = 1$, существует $E\chi_-^0$ и $P(\zeta \leq 0) = a/E\chi_-^0 > 0$.

Доказательство. Первое соотношение следует из закона больших чисел, так как

$$P(\eta_-^0 > n) < P(S_n > 0) \rightarrow 0$$

при $n \rightarrow \infty$. Следовательно, в рассматриваемом случае

$$E(e^{i\lambda\chi_-}; \eta_-^0 < \infty) = Ee^{i\lambda\chi_-^0}.$$

Существование $E\chi_-^0$ следует из множества Вальда $E\chi_-^0 = aE\eta_-^0$ и теоремы 9.4, в силу которой $E\eta_-^0 < \infty$.

Наконец, поделив обе части тождества в следствии 1 на λ и переходя к пределу $\lambda \rightarrow 0$, получим

$$a = (1 - P(\eta_+ < \infty))E\chi_-^0 = P(\zeta \leq 0)E\chi_-^0.$$

Интересно отметить, что в качестве следствия доказанного утверждения мы можем получить *усиленный закон больших чисел*. Действительно, так как событие $\{\zeta < \infty\}$ является остаточным и $P(\zeta < \infty) \geq P(\zeta \leq 0)$, то из следствия 2 вытекает, что $P(\zeta < \infty) = 1$ при $a < 0$. Это означает, что верно утверждение теоремы 4, из которой и был получен усиленный закон больших чисел.

Ниже, на основании факторизационных тождеств, мы получим некоторое обобщение этого закона.

Во всем дальнейшем в этой главе, чтобы избежать тривиальных затруднений, будем предполагать, что ξ_k с положительной вероятностью принимает как положительные, так и отрицательные значения.

Следствие 3. Если $a = E\xi_k = 0$, то $P(\eta_+ = \infty) = P(\eta_-^0 = \infty) = 0$, так что

$$1 - \varphi(\lambda) = (1 - Ee^{i\lambda\chi_+})(1 - Ee^{i\lambda\chi_-^0}).$$

Доказательство. Рассмотрим последовательность $\tilde{\xi}_k = \xi_k - \varepsilon$, $\varepsilon > 0$. Обозначив $\tilde{\zeta}$, $\tilde{\chi}_-^0$, \tilde{a} , соответствующие характеристики для введенной последовательности, получим

$$P(\zeta \leq 0) < P(\tilde{\zeta} \leq 0) = \frac{\tilde{a}}{E\tilde{\chi}_-^0} = -\frac{\varepsilon}{E\tilde{\chi}_-^0},$$

где

$$E\tilde{\chi}_-^0 \leq E(\tilde{\xi}_1; \tilde{\xi}_1 \leq 0) \leq E(\xi_1 - \varepsilon; \xi_1 \leq \varepsilon) \leq E(\xi_1; \xi_1 \leq 0) < 0.$$

Таким образом, вероятность $P(\zeta \leq 0)$ может быть сделана сколь угодно малой, так что $P(\zeta \leq 0) = P(\eta_+ = \infty) = 0$. Аналогично устанавливается, что $P(\gamma \geq 0) = 0$, и, следовательно,

$$P(\eta_-^0 = \infty) \leq P(\eta_- = \infty) = P(\gamma \geq 0) = 0. \quad \blacksquare$$

§ 2. Факторизационные тождества.

Свойства траектории $(0, S_1, S_2, \dots)$

Найдем теперь еще одно представление для компонент факторизации функции $1 - z\varphi(\lambda)$.

Теорема 3. При $|z| < 1$, $\text{Im } \lambda = 0$ функция $1 - z\varphi(\lambda)$ допускает факторизацию

$$1 - z\varphi(\lambda) = \frac{C_{z+}(\lambda)C(z)}{C_{z-}(\lambda)}, \quad C_{z+}(i\infty) = C_{z-}(-i\infty) = 1, \quad (9)$$

где

$$\begin{aligned} C_{z+}(\lambda) &= \exp \left\{ - \sum_{k=1}^{\infty} \frac{z^k}{k} E(e^{i\lambda S_k}; S_k > 0) \right\}, \\ C_{z-}(\lambda) &= \exp \left\{ \sum_{k=1}^{\infty} \frac{z^k}{k} E(e^{i\lambda S_k}; S_k < 0) \right\}, \\ C(z) &= \exp \left\{ - \sum_{k=1}^{\infty} \frac{z^k}{k} P(S_k = 0) \right\}. \end{aligned} \quad (10)$$

Доказательство. Так как $|z| < 1$, то существует $\ln(1 - z\varphi(\lambda))$, понимаемый в смысле главного значения. Следующие равенства дают нужное разложение

$$\begin{aligned} 1 - z\varphi(\lambda) &= e^{\ln(1 - z\varphi(\lambda))} = \exp \left\{ - \sum_{k=1}^{\infty} \frac{z^k \varphi^k(\lambda)}{k} \right\} = \exp \left\{ - \sum_{k=1}^{\infty} \frac{z^k}{k} E(e^{i\lambda S_k}) \right\} = \\ &= \exp \left\{ - \sum_{k=1}^{\infty} \frac{z^k}{k} E(e^{i\lambda S_k}; S_k > 0) \right\} \exp \left\{ - \sum_{k=1}^{\infty} \frac{z^k}{k} P(S_k = 0) \right\} \times \\ &\quad \times \exp \left\{ - \sum_{k=1}^{\infty} \frac{z^k}{k} E(e^{i\lambda S_k}; S_k < 0) \right\}. \end{aligned}$$

Требуемые свойства компоненты $C_{z+}(\lambda)$ ($C_{z-}(\lambda)$) (аналитичность при $\text{Im } \lambda > 0$ ($\text{Im } \lambda < 0$), непрерывность включая границу $\text{Im } \lambda = 0$, и ограниченность) доказываются точно так же, как для компоненты A_{z+} (A_{z-}) в доказательстве теоремы 1. \blacksquare

Так как в формулах (9), (10), очевидно, $C_{z+}(i\infty) = C_{z-}(-i\infty) = 1$, то в силу замечания, сделанного к теоремам 1 и 2, мы немедленно получаем

Следствие 4.

$$1 - E(e^{i\lambda\chi_+} z^{\eta_+}; \eta_+ < \infty) = \exp \left\{ - \sum_{k=1}^{\infty} \frac{z^k}{k} E(e^{i\lambda S_k}; S_k > 0) \right\};$$

$$D(z) = \exp \left\{ - \sum_{k=1}^{\infty} \frac{z^k}{k} P(S_k = 0) \right\};$$

$$1 - E(e^{i\lambda\chi_-} z^{\eta_-}; \eta_- < \infty) = \exp \left\{ - \sum_{k=1}^{\infty} \frac{z^k}{k} E(e^{i\lambda S_k}; S_k < 0) \right\}.$$

Следствие 5.

1. Всегда $\sum \frac{P(S_k=0)}{k} < \infty$.
2. Три следующие условия эквивалентны:
 - а) $P(\zeta < \infty) = 1$,
 - б) $P(\zeta \leq 0) = P(\eta_+ = \infty) > 0$,
 - в) $\sum_{k=1}^{\infty} \frac{P(S_k > 0)}{k} < \infty$ или $\sum_{k=1}^{\infty} \frac{P(S_k \geq 0)}{k} < \infty$.

Доказательство. Чтобы получить первое утверждение, надо во втором равенстве следствия 4 устремить z к 1 и вспомнить, что $D(1) = 1 - P(\chi_+^0 = 0, \eta_+^0 < \infty) > P(\xi_1 > 0) > 0$. Эквивалентность б) и в) вытекает из равенства

$$1 - P(\eta_+ < \infty) = P(\zeta \leq 0) = \exp \left\{ - \sum_{k=1}^{\infty} \frac{P(S_k > 0)}{k} \right\},$$

которое получится, если в первом тождестве в следствии 4 положить $\lambda = 0, z = 1$.

Установим эквивалентность а) и б). Если $P(\zeta \leq 0) > 0$, то $P(\zeta < \infty) > 0$, и стало быть, $P(\zeta < \infty) = 1$, т. к. событие $\{\zeta < \infty\}$ остаточное. Наоборот, пусть теперь ζ — собственная случайная величина. Выберем N так, чтобы $P(\zeta \leq N) > 0$, и $b > 0$ так, чтобы $k = N/b$ было целое и $P(\xi_1 < -b) > 0$. Тогда —

$$\{\zeta \leq 0\} \supset \left\{ \xi_1 < -1, \dots, \xi_k < -b, \sup_{j \geq 1} (-bk + \xi_{k+1} + \dots + \xi_{k+j}) \leq 0 \right\}.$$

Так как последовательность $\xi_{k+1}, \xi_{k+2}, \dots$ распределена так же, как ξ_1, ξ_2, \dots , то

$$P(\zeta \leq 0) \geq [P(\xi_1 < -b)]^k P(\zeta \leq bk) > 0.$$

Следствие 6.

1. $P(\zeta < \infty, \gamma > -\infty) = 0$.
2. Если существует $E\xi_1 = a < 0$, то

$$P(\zeta < \infty, \gamma = -\infty) = 1,$$

$$\left(\sum_{k=1}^{\infty} \frac{P(S_k > 0)}{k} < \infty, \sum_{k=1}^{\infty} \frac{P(S_k < 0)}{k} = \infty \right).$$

3. Если существует $E\xi_1 = a = 0$, то

$$P(\zeta = \infty, \gamma = -\infty) = 1,$$

$$\left(\sum_{k=1}^{\infty} \frac{P(S_k > 0)}{k} = \infty, \sum_{k=1}^{\infty} \frac{P(S_k < 0)}{k} = \infty \right).$$

Случай $a > 0$ мы здесь не рассматриваем, поскольку он является «симметричным» к случаю $a < 0$.

Доказательство. Первое утверждение следует из того, что по крайней мере один из двух рядов $\sum_{k=1}^{\infty} \frac{P(S_k < 0)}{k}$ или $\sum_{k=1}^{\infty} \frac{P(S_k \geq 0)}{k}$ расходится. Значит, по следствию 5 либо $P(\gamma = -\infty) = 1$, либо $P(\zeta = \infty) = 1$.

Второе и третье утверждения очевидным образом вытекают из следствий 2, 3 и 5. ■

Обобщение усиленного закона больших чисел, о котором мы упоминали выше, состоит в следующем.

Теорема 4 (односторонний закон больших чисел). Сходимость ряда $\sum_{k=1}^{\infty} \frac{P(S_k > \epsilon k)}{k}$ при любом $\epsilon > 0$ является необходимым и достаточным условием для того, чтобы

$$P\left(\limsup_{n \rightarrow \infty} \frac{S_n}{n} \leq 0\right) = 1. \quad (11)$$

Так как $\left\{\limsup_{n \rightarrow \infty} \frac{S_n}{n} \leq \alpha\right\}$ есть остаточное событие и т.к. $S'_n = S_n - \alpha n$ снова образуют последовательность сумм независимых одинаково распределенных величин, то из этой теоремы немедленно получаем

Следствие 7. С вероятностью 1

$$\limsup \frac{S_n}{n} = \alpha,$$

где

$$\alpha = \inf \left\{ b: \sum \frac{P(S_k > bk)}{k} < \infty \right\}.$$

Доказательство теоремы 4.

Достаточность. Если выполнено условие теоремы, то по следствию 5 $P\left(\sup_k \{S_k - \epsilon k\} < \infty\right) = 1$. Следовательно, $\{\epsilon n\}$ есть верхняя последовательность для $\{S_n\}$ и

$$P\left(\limsup_{k \rightarrow \infty} \frac{S_k}{k} \leq \epsilon\right) = 1.$$

А так как ϵ произвольно, то

$$P\left(\limsup_{k \rightarrow \infty} \frac{S_k}{k} \leq 0\right) = 1.$$

Необходимость. Наоборот, если выполнено равенство (11), то с вероятностью 1 с некоторых пор $S_n/n < \epsilon$ для любого $\epsilon > 0$. Но это значит, что $\sup_k (S_k - \epsilon k) < \infty$ с вероятностью 1 и, стало быть, ряд $\sum_{k=1}^{\infty} \frac{P(S_k > \epsilon k)}{k}$ сходится. ■

Чтобы из этой теоремы получить обычный усиленный закон больших чисел, достаточно воспользоваться п. 2 следствия 6.

В заключение этого параграфа сделаем одно замечание.

Замечание 2. В доказательстве следствий 2 и 5 фигурировало равенство $P(\zeta \leq 0) = P(\eta_+ = \infty)$. Более тонким фактом является соотношение

$$P(\zeta < 0) = P(\eta_+^0 = \infty),$$

которое мы сейчас в виде упражнения и докажем. Имеем

$$\{\eta_+^0 = \infty\} = \bigcap_{n=1}^{\infty} \{S_n < 0\} = \{\zeta < 0\} + \bigcap_{n=1}^{\infty} \{S_n < 0, \zeta = 0\}.$$

Надо убедиться, что вероятность последнего события, которое обозначим через A , равна 0. Но это так в силу того, что $A \subset B_0 = \left\{ \limsup_{n \rightarrow \infty} S_n = 0 \right\}$ и $P(B_0) = 0$ (см. § 10.1).

§ 3. Распределение $S = \max(0, \zeta) = \max_{k \geq 0} S_k$

В разного рода приложениях (одному из них мы посвятим следующий параграф) представляет интерес распределение случайной величины

$$S = \max(0, \zeta),$$

разумеется, в том случае, когда $P(\zeta < \infty) = 1$.

Мы установим здесь, как связано распределение S с распределением χ_+ , η_+ и с компонентами факторизации функции $1 - z\varphi(\lambda)$.

Заметим прежде всего, что случайная величина η_+ является марковской. Для таких величин, как легко установить (ср. с леммой 9.1), последовательность $\xi_1^* = \xi_{\eta_++1}, \xi_2^* = \xi_{\eta_++2}, \dots$ на множестве $\{\omega: \eta_+ < \infty\}$ (или при условии $\eta_+ < \infty$) распределена так же, как ξ_1, ξ_2, \dots , и не зависит от $(\eta_+, \xi_1, \dots, \xi_{\eta_+})$. Действительно,

$$\begin{aligned} P(\xi_1^* \in B_1, \dots, \xi_k^* \in B_k | \eta_+ = j, \xi_1 \in A_1, \dots, \xi_{\eta_+} \in A_{\eta_+}) &= \\ &= P(\xi_{j+1} \in B_1, \dots, \xi_{j+k} \in B_k | \xi_1 \in A_1, \dots, \xi_j \in A_j; \eta_+ = j) = \\ &= P(\xi_1 \in B_1, \dots, \xi_k \in B_k). \end{aligned}$$

Рассматривая новую последовательность $\{\xi_k^*\}_{k=1}^{\infty}$, мы опять замечаем, что для нее уровень 0 (уровень χ_+ для исходной последовательности) будет превзойден лишь с вероятностью $p = P(\eta_+ < \infty)$, а распределение $\zeta^* = \sup_k (\xi_1^* + \dots + \xi_k^*)$ будет совпадать с распределением ζ .

Таким образом, обозначив $S^* = \max(0, \zeta^*)$, можно записать, что

$$S = S(\omega) = \begin{cases} 0 & \text{на } \{\omega: \eta_+ = \infty\}, \\ S_{\eta_+} + S^* = \chi_+ + S^* & \text{на } \{\omega: \eta_+ < \infty\}. \end{cases}$$

Так как S^* , как уже отмечалось, не зависит от χ_+ и η_+ , и распределение S^* совпадает с распределением S , то

$$E e^{i\lambda S} = (1 - p) + E(e^{i\lambda(\chi_+ + S^*)}; \eta_+ < \infty) = (1 - p) + E e^{i\lambda S} E(e^{i\lambda \chi_+}; \eta_+ < \infty).$$

Отсюда вытекает

Теорема 5. Если $\sum \frac{P(S_k > 0)}{k} < \infty$ или, что то же, если $p = P(\eta_+ < \infty) < 1$, то

$$E e^{i\lambda S} = \frac{1-p}{1 - E(e^{i\lambda \chi_+}, \eta_+ < \infty)}.$$

Совершенно аналогичным образом можно было бы получить соотношение

$$E e^{i\lambda S} = \frac{1-p_0}{1 - E(e^{i\lambda \chi_+^0}, \eta_+^0 < \infty)}, \quad (12)$$

где $p_0 = P(\eta_+^0 < \infty) < 1$.

Факторизационные тождества в этом случае можно записать в виде

$$1 - \varphi(\lambda) = \frac{(1-p_0)D(1)(1 - E e^{i\lambda \chi_-})}{E e^{i\lambda S}} = \frac{(1-p)(1 - E e^{i\lambda \chi_-})}{E e^{i\lambda S}}.$$

Совершенно аналогично можно найти х.ф. момента времени $\theta = \min\{k: S_k = S\}$, когда впервые достигается максимум $S = S_\theta$. Для этого надо воспользоваться равенством

$$\theta = \begin{cases} 0 & \text{на } \{\omega: \eta_+ = \infty\}, \\ \eta_+ + \theta^* & \text{на } \{\omega: \eta_+ < \infty\}, \end{cases}$$

где θ^* распределена на $\{\eta_+ < \infty\}$ так же, как θ , и от η_+ не зависит. Находим

$$E z^\theta = (1-p) + E z^{\theta^*} E(z^{\eta_+}; \eta_+ < \infty);$$

и окончательно

$$E z^\theta = \frac{1-p}{1 - E(z^{\eta_+}; \eta_+ < \infty)}.$$

Полагая в факторизационном тождестве $\lambda = 0$, получим также соотношение

$$1 - z = [1 - E(z^{\eta_+}; \eta_+ < \infty)] [1 - E z^{\eta_-^0}] = \frac{(1-p)(1 - E z^{\eta_-^0})}{E z^\theta}.$$

§ 4. Системы обслуживания

В этом параграфе мы покажем, что к изучению распределения величины S , рассмотренной в § 3, приводит исследование систем массового обслуживания с очередью.

Представим себе, что в некоторую систему через интервалы времени τ_1, τ_2, \dots поступают «вызовы», которые надо обслуживать. Это могут быть телефонные вызовы; самолеты, прибывающие в аэропорт; клиенты бытовых учреждений; сообщения, содержащие информацию, которую надо обрабатывать, скажем, на электронной машине и др. Пусть на обслуживание k -го вызова (первый вызов пришел в момент времени 0, второй в момент времени τ_1 , и т. д.) тратится время β_k , $k = 1, 2, \dots$. Если в момент прихода k -го вызова система занята обслуживанием одного из предыдущих вызовов, то он становится в «очередь» и ждет своего обслуживания, которое начинается немедленно после того, как система обслужит все предыдущие вызовы. Задача состоит в отыскании распределения времени ожидания w_n , которое n -й вызов ждал начала своего обслуживания.

Предположим для простоты, что $w_1 = 0$ и найдем, как связаны между собой значения w_{n+1} и w_n . $(n+1)$ -й вызов пришел позже n -го на время τ_n , но зато

ему придется ждать дополнительно (по сравнению с n -м вызовом) время β_n обслуживания n -го вызова. Поэтому

$$w_{n+1} = w_n - \tau_n + \beta_n,$$

если только $w_n - \tau_n + \beta_n \geq 0$. Если же $w_n - \tau_n + \beta_n < 0$, то очевидно $w_{n+1} = 0$. Таким образом, если обозначить $\xi_{n+1} = \beta_n - \tau_n$, то

$$w_{n+1} = \max(w_n + \xi_{n+1}, 0), \quad n \geq 1, \quad w_1 = 0.$$

Покажем, что решением этих уравнений является последовательность

$$w_n = S_n - \min(S_1, \dots, S_n) \equiv S_n - \gamma_n,$$

где, как и прежде, $S_n = \sum_{k=1}^n \xi_k$. Так как $S_1 - \gamma_1 = 0$, то достаточно проверить рекуррентное соотношение

$$S_{n+1} - \gamma_{n+1} = \max(0, S_n + \xi_{n+1} - \gamma_n). \quad (13)$$

Если $S_{n+1} > \gamma_n$, то $\gamma_{n+1} = \gamma_n$, и в обеих частях (13) будет стоять $S_{n+1} - \gamma_n$. Если $S_{n+1} \leq \gamma_n$, то $S_{n+1} = \gamma_{n+1}$, и в обеих частях (13) будет стоять 0. Таким образом, равенство $w_n = S_n - \gamma_n$ доказано.

Предположим теперь, что ξ_k независимы и одинаково распределены. Тогда разность

$$S_n - \gamma_n = \max(S_n - S_1, \dots, S_n - S_n) = \max(0, \xi_n, \xi_n + \xi_{n-1}, \dots, \xi_n + \dots + \xi_2)$$

будет распределена, очевидно, так же, как $\bar{S}_{n-1} = \max(0, S_1, \dots, S_{n-1}) = \max(0, \zeta_{n-1})$. Так что распределение w_n совпадает с распределением $\max(0, \zeta_{n-1})$. Так как ζ_n монотонно сходится к ζ при $n \rightarrow \infty$, то всегда существует

$$\lim_{n \rightarrow \infty} P(w_n < x) = P(\max(0, \zeta) < x) = P(S < x).$$

Это распределение называют *стационарным распределением времени ожидания*. Как мы уже знаем, оно будет собственным, если $E\xi_2 = E\beta_1 - E\tau_1 < 0$. Если же $E\beta_1 > E\tau_1$, или $E\beta_1 = E\tau_1$ и $\beta_1 \neq \tau_1$, то «стационарное» время ожидания будет бесконечным.

§ 5. Факторизационные тождества для распределений, связанных с показательной функцией

Рассмотрим теперь ряд примеров, когда факторизационные тождества используются для отыскания явного вида распределений χ_+ , η_+ , S и других.

1. Предположим сначала, что при $x > 0$,

$$P(\xi_1 > x) = be^{-\alpha x}$$

при некоторых $\alpha > 0$ и $1 > \beta > 0$. При $x \leq 0$ распределение ξ_1 может быть произвольным. Этот случай замечателен тем, что для него распределение χ_+ не зависит от η_+ и является показательным. Действительно,

$$P(\chi_+ > x; \eta_+ = n) = \int_{-\infty}^0 P(\eta_+ = n, S_{n-1} \in dz, \chi_+ > x) =$$

$$\begin{aligned}
&= \int_{-\infty}^0 P(\zeta_{n-1} \leq 0, S_{n-1} \in dz, \xi_n > x - z) = \int_{-\infty}^0 P(\xi_n > x - z) P(\zeta_{n-1} \leq 0, S_{n-1} \in dz) = \\
&= e^{-\alpha x} \int_{-\infty}^0 b e^{\alpha z} dP(\zeta_{n-1} \leq 0, S_{n-1} < z) \equiv p_n e^{-\alpha x}.
\end{aligned}$$

Отсюда следует, что

$$\begin{aligned}
p_n &= P(\eta_+ = n), \quad P(\chi_+ > x | \eta_+ = n) = e^{-\alpha x}, \\
P(\chi_+ > x; \eta_+ < \infty) &= p e^{-\alpha x},
\end{aligned}$$

где $p = p_1 + p_2 + \dots$. Следовательно,

$$E(e^{i\lambda\chi_+}; \eta_+ < \infty) = \int_0^\infty e^{i\lambda x} d(p - p e^{-\alpha x}) = p \int_0^\infty \alpha e^{i\lambda x - \alpha x} dx = -\frac{p\alpha}{i\lambda - \alpha},$$

и если $S < \infty$ с вероятностью 1 (т. е. если $p < 1$), то

$$E e^{i\lambda S} = \frac{1 - p}{1 + (p\alpha)/(i\lambda - \alpha)} = (1 - p) + p \frac{-\alpha(1 - p)}{i\lambda - \alpha(1 - p)}.$$

Последнее равенство означает, что

$$P(S = 0) = 1 - p, \quad P(S > x) = p e^{-\alpha(1-p)x} \quad \text{при } x \geq 0.$$

В полученных соотношениях осталось неизвестным лишь число p (знаем только, что $1 - p = a/(E\chi_-^0)$, если существует $a = E\xi_1 < 0$ (см. следствие 2)). Предоставляем читателю с помощью тождества

$$1 - \varphi(\lambda) = \left(1 + \frac{p\alpha}{i\lambda - \alpha}\right) \left[1 - E(e^{i\lambda\chi_-^0}; \eta_-^0 < \infty)\right] \quad (14)$$

убедиться, что $\alpha(1 - p) = \mu(1)$, где $\mu(1) > 0$ — единственный положительный корень уравнения $\psi(\mu) = 1$, где $\psi(\mu) = \varphi(-i\mu) = E e^{\mu\xi_1}$ ($\psi'(0) < 0$, $\psi''(\mu) > 0$, $\psi(\mu) \rightarrow \infty$) при $\mu \rightarrow \alpha$. При $\lambda = -i\mu(1)$ правая часть тождества (14) обязана обратиться в 0).

Отметим также, что в рассматриваемом случае, пользуясь независимостью χ_+ и η_+ , нетрудно найти и распределение η_+ .

2. Предположим теперь, что при $x > 0$,

$$P(\xi_1 < -x) = b e^{-\alpha x}, \quad \alpha > 0, \quad 1 > b > 0,$$

и что S с вероятностью 1 конечно. В этом случае заключения, аналогичные предыдущим, справедливы относительно величин χ_- , η_- . Разница состоит лишь в том, что здесь $P(\eta_- < \infty) = 1$ и, следовательно,

$$P(\chi_- < -x) = e^{-\alpha x}, \quad E\chi_- = -\frac{1}{\alpha}, \quad E e^{i\lambda\chi_-} = \frac{\alpha}{i\lambda + \alpha},$$

Отсюда находим, в частности, что (см. (12))

$$\begin{aligned}
1 - p_0 &= P(\eta_+ = \infty) = \frac{a}{E\chi_-} = -a\alpha; \\
E e^{i\lambda S} &= \frac{1 - p_0}{1 - \varphi(\lambda)} [1 - E e^{i\lambda\chi_-}] = -\frac{a\alpha i\lambda}{(i\lambda + \alpha)(1 - \varphi(\lambda))}.
\end{aligned}$$

В заключение следует отметить, что полученные соотношения позволяют находить стационарное распределение времени ожидания для систем обслуживания с ожиданием (см. § 4), когда *хотя бы одна из случайных величин τ_1 или β_1 имеет показательное распределение* (в силу ряда причин такие предположения можно делать довольно часто. Имеется в виду также, что последовательности $\{\tau_j\}$ и $\{\beta_j\}$ составлены из независимых и одинаково распределенных величин.) В самом деле, если, например, $P(\beta_1 > x) = e^{-\alpha x}$, то для $\xi_1 = \beta_1 - \tau_1$ будем иметь при $x > 0$,

$$\begin{aligned} P(\xi_1 > x) &= P(\beta_1 - \tau_1 > x) = \int_0^{\infty} dF_{\tau}(y) P(\beta > x + y) = \\ &= e^{-\alpha x} \int_0^{\infty} e^{-\alpha y} dF_{\tau}(y) = e^{-\alpha x} E e^{-\alpha \tau_1}. \end{aligned}$$

Это означает, что мы находимся в условиях п. 1 этого параграфа. Аналогично обстоит дело, если показательным является распределение τ_1 .

Важные выводы о распределении рассматриваемых случайных величин (S , η_+ , χ_+ и др.) можно сделать, используя качественные аналитические свойства компонент факторизации $C_{\pm z}(\lambda)$, определяющих нужные нам распределения. Пусть, например, $\varphi(\lambda)$ есть рациональная функция от λ ,

$$1 - z\varphi(\lambda) = \prod_{k=1}^m \frac{\lambda - \lambda_k(z)}{\lambda - \Lambda_k(z)} \quad (15)$$

(это будет, в частности, если $\xi_1 = \beta - \tau$ (см. § 4), где $\beta \in \Gamma_{\alpha_1}$, $\tau \in \Gamma_{\alpha_2}$ и независимы.

Представление (15) возникает и при других предположениях относительно ξ_1 , связанных с показательным распределением). Тогда требуемые аналитические свойства $C_{\pm z}(\lambda)$ (прежде всего аналитичность в соответствующих полуплоскостях) будут иметь место, если только положить¹⁾

$$C_z(\lambda) = \prod_k \frac{\lambda - \lambda_k^-(z)}{\lambda - \Lambda_k^-(z)},$$

где $\lambda_k^-(z)$, $\Lambda_k^-(z)$ — все нули и полюса функции $1 - z\varphi(\lambda)$, лежащие в нижней полуплоскости.

В связи с изложенным можно отметить также, что если некоторая х.ф. $\psi(\lambda)$ есть рациональная функция, то путем разложения ее на простые дроби соответствующее ей распределение можно представить как смесь показательных распределений или их свертки.

Рассмотрение случаев, когда факторизация позволяет находить явные формулы, будет продолжена в § 8.

§ 6. Симметричные непрерывно распределенные случайные величины

Будем говорить, что *распределение ξ_k симметрично*, если оно совпадает с распределением $-\xi_k$, и называть *распределение ξ_k непрерывным*, если функция

¹⁾ Подробнее об этом см. [3].

распределения ξ_k непрерывна. Для таких величин распределения S_n при всех n также симметричны и непрерывны и

$$P(S_n > 0) = P(S_n < 0) = \frac{1}{2}, \quad P(S_n = 0) = 0.$$

Стало быть, здесь $D(z) \equiv 1$, $P(\chi_+ = 0) = 0$, и с вероятностью 1 $\eta_+ = \eta_+^0$, $\chi_+ = \chi_+^0$.

В силу следствия 5 в рассматриваемом случае $\zeta = \infty$, $\gamma = -\infty$ с вероятностью 1 и η_+ есть собственная случайная величина. В силу следствия 4

$$1 - Ez^{\eta_+} = \exp \left\{ - \sum_{k=1}^{\infty} \frac{z^k}{2k} \right\} = \exp \left\{ \frac{1}{2} \ln(1-z) \right\} = \sqrt{1-z},$$

$$Ez^{\eta_+} = 1 - \sqrt{1-z}.$$

Отсюда, разлагая в ряд $\sqrt{1-z}$, находим, что

$$P(\eta_+ = n) = \frac{(2n)!}{(2n-1)(n!)^2 2^{2n}}.$$

При больших n по формуле Стирлинга обнаруживаем, что

$$P(\eta_+ = n) \sim \frac{1}{2\sqrt{\pi n^{3/2}}};$$

$$P(\zeta_n < 0) = \sum_{k=n+1}^{\infty} P(\eta_+ = k) \sim \frac{1}{\sqrt{\pi n}}.$$

Аналогичные формулы справедливы и для η_- , γ_n .

§ 7. Тождество Поллачека—Спитцера

До сих пор мы получали различные факторизационные тождества как соотношения между числителями в представлениях (2), (3) и (4). Обратимся теперь к знаменателям. Получим еще одно тождество, которое играет важную роль при исследовании распределений

$$\bar{S}_n = \max(0, \zeta_n) = \max(0, S_1, \dots, S_n).$$

Это так называемое тождество Поллачека—Спитцера, связывающее х.ф. \bar{S}_n , $n = 1, 2, \dots$, с характеристическими функциями $\max(0, S_n)$, $n = 1, 2, \dots$.

Теорема 6. При $|z| < 1$, $\text{Im } \lambda \geq 0$

$$\sum_{n=0}^{\infty} z^n E e^{i\lambda \bar{S}_n} = \exp \left\{ \sum_{k=1}^{\infty} \frac{z^k}{k} E e^{i\lambda \max(0, S_k)} \right\}.$$

Если использовать обозначения теоремы 3, то правую часть этого тождества можно записать в виде $\frac{C_{z+}(0)}{(1-z)C_{z+}(\lambda)}$.

Доказательство. Из теорем 1–3 и соотношения (2) вытекает, что

$$\sum_{k=0}^{\infty} z^k E(e^{i\lambda S_k}; \zeta_n < 0) = [1 - E(e^{i\lambda \chi_-} z^{\eta_-}; \eta_- < \infty)]^{-1} = C_{z-}(\lambda),$$

где принято $E(e^{i\lambda S_0}; \zeta_0 < 0) = 1$. Обозначим $S_k^* = S_{n-k} - S_n$, $\theta_n^* = \min\{k : S_k^* = \bar{S}_n^* = \max(0, S_1^*, \dots, S_n^*)\}$ (θ_n^* есть первый момент достижения максимума в последовательности $0, S_1^*, \dots, S_n^*$). Тогда событие $\{S_n \in dx, \zeta_n < 0\}$ можно записать в виде $\{S_n^* \in -dx, \theta_n^* = n\}$. Отсюда следует, что

$$E(e^{i\lambda S_n}; \zeta_n < 0) = E(e^{-i\lambda S_n^*}; \theta_n^* = n),$$

$$\sum_{n=0}^{\infty} z^n E(e^{i\lambda S_n^*}; \theta_n^* = n) = C_{z-}(-\lambda). \quad (16)$$

Но последовательность S_1^*, \dots, S_n^* распределена так же, как суммы $\xi_1^*, \xi_1^* + \xi_2^*, \dots, \xi_1^* + \dots + \xi_n^*$, где $\xi_k^* = -\xi_k$. Если положить $\theta_n = \min\{k : S_k = \bar{S}_n\}$, то тождество (16) можно эквивалентным образом записать в виде

$$\sum_{n=0}^{\infty} z^n E(e^{i\lambda S_n}; \theta_n = n) = C_{z-}^*(-\lambda),$$

где $C_{z-}^*(-\lambda)$ есть отрицательная компонента факторизации функции $1 - z\varphi^*(\lambda) = 1 - z\varphi(-\lambda)$, соответствующей случайной величине $-\xi_k$. Так как

$$1 - z\varphi(-\lambda) = D(z) \frac{C_{z+}(-\lambda)}{C_{z-}(-\lambda)}$$

и функция $1/C_{z+}(-\lambda)$ обладает всеми свойствами отрицательной компоненты $C_{z-}^*(-\lambda)$ факторизации функции $1 - z\varphi^*(-\lambda)$, а функция $1/C_{z-}(-\lambda)$ обладает всеми свойствами положительной компоненты, то мы находим, что $C_{z-}^*(\lambda) = C_{z+}^{-1}(-\lambda)$ и

$$\sum_{n=0}^{\infty} z^n E(e^{i\lambda S_n}; \theta_n = n) = \frac{1}{C_{z+}(\lambda)}.$$

Теперь надо заметить, что

$$\begin{aligned} E e^{i\lambda \bar{S}_n} &= \sum_{k=0}^{\infty} E(e^{i\lambda \bar{S}_n}; \theta_n = k) = \\ &= \sum_{k=0}^n E(e^{i\lambda S_k}; \theta_k = k, S_{k+1} - S_k \leq 0, \dots, S_n - S_k \leq 0) = \\ &= \sum_{k=0}^n E(e^{i\lambda S_k}; \theta_k = k) P(\bar{S}_{n-k} = 0). \end{aligned}$$

Так как в правой части стоит свертка двух последовательностей, то

$$\sum_{n=0}^{\infty} z^n E e^{i\lambda \bar{S}_n} = \sum_{n=0}^{\infty} z^n P(\bar{S}_n = 0) \frac{1}{C_{z+}(\lambda)}.$$

Полагая $\lambda = 0$, получим

$$\frac{1}{1-z} = \sum_{n=0}^{\infty} z^n P(\bar{S}_n = 0) \frac{1}{C_{z+}(0)}.$$

Следовательно,

$$\begin{aligned} \sum_{n=0}^{\infty} z^n E e^{i\lambda \bar{S}_n} &= \frac{C_{z+}(0)}{(1-z)C_{z+}(\lambda)} = \\ &= \exp \left\{ -\ln(1-z) + \sum_{k=1}^{\infty} \frac{z^k}{k} E(e^{i\lambda S_k}; S_k > 0) - \sum_{k=1}^{\infty} \frac{z^k}{k} P(S_k > 0) \right\} = \\ &= \exp \left\{ \sum_{k=1}^{\infty} \frac{z^k}{k} E(e^{i\lambda S_k}; S_k > 0) + \sum_{k=1}^{\infty} \frac{z^k}{k} P(S_k \leq 0) \right\} = \exp \left\{ \sum_{k=1}^{\infty} \frac{z^k}{k} E(e^{i\lambda \max(0, S_k)}) \right\}. \end{aligned}$$

Теорема доказана. ■

§ 8*. Явные формулы для дискретных блужданий, непрерывных сверху

Дискретным блужданием, непрерывным сверху, будем называть блуждание частицы, порожденное последовательностью сумм $0, S_1, S_2, \dots$, когда ξ_k целочисленны и $\max_{\omega} \xi_k(\omega) = 1$.

Другими словами, если обозначить $f_k = P(\xi_1 = k)$, то $f_k = 0$ при $k \geq 2$, $\sum_{k=-\infty}^1 f_k = 1$. В этом случае, очевидно, $\chi_+ = 1$ на множестве $\{\eta_+ < \infty\}$, и в силу теоремы 5

$$E e^{i\lambda S} = \frac{1-p}{1-pe^{i\lambda}}, \quad p = P(\chi_+ < \infty).$$

Это означает, что S имеет геометрическое распределение

$$P(S = k) = (1-p)p^k, \quad k = 0, 1, \dots$$

Осталось лишь найти p . Чтобы сделать это, заметим, что по формуле полной вероятности (по первому скачку) при любом $k \geq 2$,

$$P(S = k) = \sum_{j=-\infty}^1 f_j P(S = k-j); \quad 1 = \sum_{j=-\infty}^1 f_j p^{-j}.$$

Таким образом, $z = 1/p > 1$ есть решение уравнения $f(z) = 1$, где $f(z)$ есть производящая функция $f(z) = E z^{\xi_1}$. Так как $f(z)$ всегда выпукла (см. гл. 7), то это уравнение имеет не более двух корней: один из них $z = 1$, другой корень $z > 1$ возможен лишь в случае $E \xi_1 < 0$. Итак, в случае $E \xi_1 < 0$ число $p < 1$ есть единственное решение уравнения $1 = f(p^{-1})$.

Оказывается, для блужданий, непрерывных сверху, могут быть найдены и многие другие явные формулы. Очевидно, что в этом случае последовательность $\{S_n\}$ не может «перескочить» ни один целочисленный уровень $x > 0$: если $S_n > x$, то обязательно найдется $k \leq n$ такое, что $S_k = x$.

Задача, которая нас интересует, по-прежнему состоит в отыскании распределения максимума $\bar{S}_n = \max(0, S_1, \dots, S_n)$, знание которого, как уже отмечалось, бывает важным для многих задач математической статистики, теории массового

обслуживания и др. Отметим, что найти распределение \bar{S}_n — это то же самое, что найти распределение

$$\eta(x) = \min\{k: S_k \geq x\}, \quad \eta(0) = 0,$$

так как

$$\{\bar{S}_n < x\} = \{\eta(x) > n\}. \quad (17)$$

Здесь мы считаем $\eta(x) = \infty$, если $S < x$.

Тождество § 7 дает двойное преобразование над распределением \bar{S}_n . Исследования этого тождества показывают, что само распределение \bar{S}_n (или $\eta(x)$), как правило, в явном виде через распределение ξ_k не выражается. Однако для дискретных блужданий, непрерывных сверху, оказывается, имеют место замечательные «двойственные» соотношения, которые мы сейчас с помощью тождества Поллачека—Спитцера докажем.

Теорема 7. Если ξ_k целочисленны, то условие $P(\xi_k \geq 2) = 0$ является необходимым и достаточным для того, чтобы

$$nP(\eta(x) = n) = xP(S_n = x), \quad x \geq 1. \quad (18)$$

С помощью тождества Вальда нетрудно убедиться также, что если существует $E\xi_1 = a > 0$, то блуждание $\{S_n\}$ будет непрерывным сверху тогда и только тогда, когда $E\eta(x) = x/a$.

Доказательство. Обозначим

$$p_x = P(\eta(x) = \infty) = P(S < x), \quad p_{x,n} = P(\eta(x) = n),$$

$$q_{x,n} = P(\eta(x) > n) = \sum_{k=n+1}^{\infty} p_{x,k} + p_x.$$

Так как для каждого y , $0 \leq y \leq x$,

$$\{\eta(x) = n\} \subset \bigcup_{k=0}^n \{\eta(y) = k\},$$

то, используя непрерывность сверху, можно написать по формуле полной вероятности

$$p_{x,n} = \sum_{k=0}^n p_{y,k} p_{x-y,n-k},$$

где $p_{0,0} = 1$, $p_{y,0} = 0$ при $y > 0$. Следовательно, при $|z| \leq 1$ по правилу свертки

$$p_x(z) = E(z^{\eta(x)}; \eta(x) < \infty) = \sum_{k=0}^{\infty} p_{x,n} z^n = p_y(z) p_{x-y}(z).$$

Полагая $y = 1$, $p_1(z) = p(z)$, получим

$$p_x(z) = p(z) p_{x-1}(z) = p^x(z), \quad x \geq 0.$$

Отсюда можно легко найти производящую функцию последовательности $q_{x,n}$:

$$\begin{aligned} q_x(z) &= \sum_{n=0}^{\infty} z^n \left(p_x + \sum_{k=n+1}^{\infty} p_{x,k} \right) = \frac{p_x}{1-z} + \sum_{n=1}^{\infty} p_{x,n} \sum_{k=0}^{n-1} z^k = \\ &= \frac{p_x}{1-z} + \sum_{n=0}^{\infty} p_{x,n} \frac{1-z^n}{1-z} = \frac{p_x}{1-z} + \frac{p_x(1) - p_x(z)}{1-z} = \frac{1-p_x(z)}{1-z}. \end{aligned}$$

Заметим, что здесь число $p_x(1) = P(\eta(x) < \infty) = P(S \geq x)$ может быть и меньшим, чем 1. Используя (17), получим

$$P(\bar{S}_n = x) = P(\eta(x+1) > n) - P(\eta(x) > n),$$

$$\sum_{n=0}^{\infty} z^n P(\bar{S}_n = x) = \frac{(1 - p^{x+1}(z)) - (1 - p^x(z))}{1 - z} = \frac{p^x(z)(1 - p(z))}{1 - z}.$$

Наконец, пользуясь абсолютной сходимостью фигурирующих ниже рядов, находим при $|v| < 1$, $|z| < 1$,

$$\sum_{n=0}^{\infty} z^n E v^{\bar{S}_n} = \sum_{x=0}^{\infty} v^x \sum_{n=0}^{\infty} z^n P(\bar{S}_n = x) = \frac{1 - p(z)}{(1 - z)(1 - vp(z))}.$$

Обращаясь теперь к формуле Поллачека—Спитцера, можно записать, что

$$\sum_{n=1}^{\infty} \frac{z^n}{n} E v^{\max(0, S_n)} = \ln \frac{1 - p(z)}{1 - z} - \ln(1 - vp(z)) = \ln \frac{1 - p(z)}{1 - z} + \sum_{x=1}^{\infty} \frac{(vp(z))^x}{x}.$$

Сравнивая здесь коэффициенты при v^x , $x \geq 1$, получим

$$\sum_{n=1}^{\infty} \frac{z^n}{n} P(S_n = x) = \frac{p^x(z)}{x}, \quad x \geq 1. \quad (19)$$

Учитывая, что $p^x(z) = p_x(z)$ и сравнивая в (19) коэффициенты при z^n , $n \geq 1$, получим

$$\frac{1}{n} P(S_n = x) = \frac{1}{x} P(\eta_x = n), \quad x \geq 1, \quad n \geq 1.$$

Достаточность утверждения теоремы доказана.

Читателю предлагается проверить, что имеет место также равенство

$$f_x(z) = p^x(z) f_0(z),$$

где $f_x(z) = \sum_{n=0}^{\infty} z^n P(S_n = x)$, и что $p(z)$ при $|z| < 1$ является единственным в области

$|p| \leq 1$ корнем уравнения $z E p^{-\xi_1} = 1$.

Докажем теперь необходимость условия $P(\xi_k \geq 2) = 0$. Обозначим $f_k = P(\xi_1 = k)$. Тогда в силу (18)

$$x f_x = p_{x,1} = \sum_{k=x}^{\infty} f_k.$$

Полагая $f(v) = \sum_{x=1}^{\infty} v^x f_x$, получим отсюда

$$\sum_{x=1}^{\infty} x v^x f_x = v f'(v) = \sum_{x=1}^{\infty} v^x \sum_{k=x}^{\infty} f_k = \sum_{k=1}^{\infty} f_k \frac{v(v^k - 1)}{v - 1} = \frac{v(1 - f(v))}{1 - v}.$$

Мы получили относительно $f(v)$ уравнение

$$f'(v) = \frac{1 - f(v)}{1 - v}.$$

Сделаем замену $y = 1 - f$, придем к уравнению $y' = -y/(1 - v)$, из которого находим $\ln y = \ln(1 - v) + c$, $f(v) = 1 - y = 1 - c(1 - v)$.

Но это означает, что $f_k = 0$ при $k \geq 2$. ■

Пользуясь полученными формулами, можно, например, найти в примере 4.8 распределение для времени разорения игрока при игре с бесконечно богатым противником (суммарный капитал равен ∞). Если начальный капитал первого игрока равен z , то для времени его разорения $\eta(z)$ получим

$$P(\eta(z) = n) = \frac{z}{n} P(S_n = z),$$

где $S_n = \sum_{j=1}^n \xi_j$; $P(\xi_j = 1) = q$, $P(\xi_j = -1) = p$ (p есть вероятность выигрыша первого игрока в одной партии). Следовательно, если n и z вместе либо четны, либо нечетны, то

$$P(\eta(z) = n) = \frac{z}{n} \binom{n}{(n-z)/2} q^{(n+z)/2} p^{(n-z)/2} \quad (20)$$

и $P(\eta(z) = n) = 0$ в противном случае.

Представляет интерес вопрос, как быстро убывает $P(\eta(z) > n)$ с ростом n в том случае, когда игрок с вероятностью 1 разорится, т.е. когда $P(\eta(z) < \infty) = 1$. Как мы уже знаем, это будет тогда и только тогда, когда $p \leq q$. (Такое утверждение будет следовать также и из рассмотрений § 12.3).

Применяя формулу Стирлинга так же, как это делалось в локальной предельной теореме для схемы Бернулли, из (20) уже нетрудно получить, что при каждом фиксированном z при $n \rightarrow \infty$ (n и z одной четности) и $p \leq q$,

$$P(\eta(z) = n) \sim \frac{z}{n^{3/2}} \sqrt{\frac{2}{\pi}} (4pq)^{n/2} \left(\frac{q}{p}\right)^{z/2};$$

$$P(\eta(z) \geq n) \sim \frac{z}{n^{3/2}(p-q)^2} \sqrt{\frac{2}{\pi}} (4pq)^{n/2} \left(\frac{q}{p}\right)^{z/2} \quad \text{при } p < q$$

и

$$P(\eta(z) \geq n) \sim z \sqrt{\frac{2}{\pi n}} \quad \text{при } p = q.$$

Последнее соотношение позволило нам в условиях § 8.9 получить предельное распределение для числа пересечений траекторией S_1, \dots, S_n полосы $[a, b]$ (см. (8.63)). Это утверждение с точностью до нормирующих постоянных сохранится и для произвольного блуждания со свойством $E\xi_k = 0$, $E\xi_k^2 < \infty$. Однако доказательство этого факта даже для блуждания, непрерывного сверху, требует больших усилий, несмотря на то, что для таких блужданий пересечение снизу вверх траекторией $\{S_n\}$ прямой $x = 0$ разбивает эту траекторию, как и в § 8.9, на независимые одинаково устроенные циклы.

Формулу (18) при $P(S_n = x) > 0$ можно переписать также в виде

$$P(\eta(x) = n | S_n = x) = \frac{x}{n}, \quad (21)$$

означающем, что среди траекторий $\{S_k\}$, $k = 0, 1, \dots, n$, со свойством $S_n = x$ доля тех, которые имеют абсолютный максимум в последний момент, составляет x/n

независимо от характера распределения ξ_k (при условии $P(\xi_k \geq 2) = 0$). В таком виде теорема 7 приобретает комбинаторный характер и *может быть доказана комбинаторным путем*.

Действительно, пусть ξ и a означают, соответственно, векторы $\xi = (\xi_1, \dots, \xi_n)$, $a = (a_1, \dots, a_n)$, V_x — совокупность разных векторов a со свойством $\sum_{k=1}^n a_k = x$, V_x^0 — совокупность векторов a из V_x , отличающихся составом координат. Тогда

$$P(S_n = x) = \sum_{a \in V_x} P(\xi = a) = \sum_{a \in V_x^0} c_{x,n}(a) P(\xi = a),$$

где $c_{x,n}(a)$ — соответствующие целочисленные коэффициенты. Далее

$$P(\eta(x) = x) = \sum_{a \in V_x} P(\xi = a) P(\eta(x) = n | \xi = a) = \sum_{a \in V_x^0} b_{x,n}(a) P(\xi = a), \quad (22)$$

где $P(\eta(x) = n | \xi = a)$ равны 1 или 0, так что $b_{x,n}(a)/c_{x,n}(a)$ есть доля таких вероятностей, равных 1. Достаточно показать, что эта доля равна x/n . Мы убедимся в этом не только для каждого $a \in V_x^0$, а более того, для каждого подкласса C_a векторов из V_x , отличающихся от $a \in V_x$ сдвигом координат:

$$T^l a \in C_a, \quad l = 0, \dots, n-1,$$

где $T^l a = (a_{l+1}, \dots, a_n, a_{n+1}, \dots, a_{n+l})$, $a_{n+j} = a_j$ (каждому $a \in V_x^0$ в сумме (22) отвечает более широкий класс, распадающийся на подклассы вида C_a).

Итак, установим, что

$$P(\eta(x) = n | \xi = T^l a) = 1 \quad (23)$$

для x индексов l из n возможных.

Обозначим

$$S_k^l = a_{l+1} + \dots + a_{l+k}, \quad \eta^l(x) = \min\{j: S_j^l = x\}, \quad m = \min\{j: S_j^0 = \max_k S_k^0\}.$$

Траектория $\{S_k^m\}$, $k = 1, \dots, n$, распадается на x «лестничных моментов» (точек возрастания сумм S_k^m): j -й «лестничный момент» равен $\eta^m(j)$. Очевидно, что $\eta^l(x) = n$, если $l = m$, $m + \eta^m(1), \dots, m + \eta^m(x) = m + n$ и что $\eta^l < n$ для других l . Отсюда следует (23) и вместе с этим равенства

$$\frac{b_{x,n}(a)}{c_{x,n}(a)} = \frac{x}{n},$$

$$P(\eta(x) = n) = \frac{x}{n} \sum_{a \in V_x^0} c_{x,n}(a) P(\xi = a) = \frac{x}{n} P(S_n = x).$$

Если $P(\xi_k \geq 2) > 0$, то из приведенных рассуждений видно, что в этом случае число «лестничных моментов» траектории $\{S_m^k\}$ будет $\leq x$, а для некоторых траекторий, имеющих положительную вероятность, будет строго меньше x . Это означает, что в сумме (22) выполняется $b_{x,n}(a)/c_{x,n}(a) \leq x$, а для некоторых a это неравенство будет строгим. Следовательно, в случае $P(\xi \geq 2) > 0$,

$$P(\eta(x) = n) < \frac{x}{n} P(S_n = x).$$

Отсюда следует необходимость в утверждении теоремы 7.

Последовательности зависимых испытаний. Цепи Маркова

§ 1. Счетные цепи Маркова. Определения и примеры. Классификация состояний

До сих пор мы изучали последовательности независимых испытаний. Рассмотрим теперь простейший вариант последовательности зависимых испытаний.

Пусть G есть некоторый эксперимент, который имеет конечное или счетное множество исходов $\{E_1, E_2, \dots\}$. Предположим, что мы неограниченно повторяем эксперимент G . Номер исхода n -го эксперимента обозначим X_n .

Вообще говоря, вероятности различных значений E_{X_n} могут зависеть от того, какие события появились на предыдущих $n-1$ испытаниях. Если же эта вероятность *при фиксированном исходе $E_{X_{n-1}}$ испытания с номером $n-1$* не зависит от исходов остальных $n-2$ испытаний, то эта последовательность испытаний образует цепь Маркова.

Чтобы дать точное определение цепи Маркова, рассмотрим последовательность целочисленных случайных величин $\{X_n\}_{n=0}^{\infty}$. Если в n -м испытании реализовалось событие E_j , то мы будем считать, что $X_n = j$.

Определение 1. Последовательность $\{X_n\}_{n=0}^{\infty}$ образует *цепь Маркова*, если

$$P(X_n = j | X_0 = k_0, X_1 = k_1, \dots, X_{n-2} = k_{n-2}, X_{n-1} = i) = \\ = P(X_n = j | X_{n-1} = i) \equiv p_{ij}^{(n)}. \quad (1)$$

Таким образом, цепь Маркова можно представить себе как некоторую систему с возможными состояниями $\{E_1, E_2, \dots\}$. задается некоторое «начальное» распределение величины X_0 :

$$P(X_0 = j) = p_j^0, \quad \sum p_j^0 = 1;$$

затем в целочисленные моменты времени система меняет свое состояние, при этом вероятность в момент времени n попасть в состояние E_j , если известна вся предыстория системы, зависит только от того, в каком состоянии находилась система в момент $n-1$. Это свойство кратко можно было бы охарактеризовать так: *при фиксированном настоящем будущее и прошлое последовательности $\{X_n\}$ независимы*. Цепи Маркова, описанные в определении 1, называют *счетными* или *дискретными* (счетным является множество возможных состояний).

Например, ветвящийся процесс $\{\zeta_n\}$, описанный в § 7.8, где ζ_n есть количество частиц в n -м поколении, представляет собой цепь Маркова с возможными состояниями $\{0, 1, 2, \dots\}$.

В терминах условных математических ожиданий или условных вероятностей (см. § 4.8) *свойство марковости* (так будем называть свойство (1)) можно записать в виде

$$P(X_n = j | \sigma(X_0, \dots, X_{n-1})) = P(X_n = j | \sigma(X_{n-1})),$$

где $\sigma(\cdot)$ есть σ -алгебра, порожденная случайными величинами, стоящими в аргументе, или, проще,

$$P(X_n = j | X_0, \dots, X_{n-1}) = P(X_n = j | X_{n-1}).$$

Такое определение немедленно допускает обобщение на цепи Маркова с более общим пространством состояний (см. § 6, § 7).

Вопрос существования последовательности $\{X_n\}_{n=0}^{\infty}$, представляющей собой цепь Маркова с данными вероятностями перехода $p_{ij}^{(n)}$ ($p_{ij}^{(n)} \geq 0$, $\sum_j p_{ij}^{(n)} = 1$) и заданным «начальным» распределением $\{p_k^0\}$ значения X_0 , решается, как и в независимом случае, с помощью теоремы Колмогорова (см. Приложение 2) и задания согласованных совместных распределений равенствами

$$P(X_0 = k_0, X_1 = k_1, \dots, X_n = k_n) = p_{k_0}^0 p_{k_0 k_1}^{(1)} p_{k_1 k_2}^{(2)} \dots p_{k_{n-1} k_n}^{(n)},$$

которые, как легко видеть, удовлетворяют марковскому свойству (1).

Определение 2. Марковская цепь $\{X_n\}_{n=0}^{\infty}$ называется *однородной*, если вероятности $p_{ij}^{(n)}$ не зависят от n .

Рассмотрим несколько примеров.

Пример 1. *Блуждания с поглощением и с отражением.* Пусть $a > 0$ — целое число. Рассмотрим блуждание частицы по целым точкам между 0 и a . Если $0 < k < a$, то из точки k с вероятностями $1/2$ частица переходит в $k - 1$ или $k + 1$. Если k равно 0 или a , то частица остается в точке k с вероятностью 1. Это так называемое блуждание с *поглощением*. Если X_n — случайная величина, равная координате частицы в момент n , то последовательность $\{X_n\}$ образует цепь Маркова, так как условное распределение случайной величины X_n зависит только от значения X_{n-1} . Легко видеть, что эта цепь однородна.

Таким блужданием можно описывать безобидную игру (см. пример 4.8) в том случае, когда суммарный капитал обоих игроков равен a . Попадание в точку a будет означать разорение второго игрока.

Если же частица из точки 0 с вероятностью 1 переходит в точку 1, а из точки a с вероятностью 1 в точку $a - 1$, то получим блуждание с *отражением*. Ясно, что и в этом случае положения частицы X_n будут связаны в однородную цепь Маркова.

Пример 2. Пусть $\{\xi_k\}_{k=0}^{\infty}$ — последовательность независимых целочисленных случайных величин и $d > 0$ — целое число. Случайные величины $X_n = \sum_{k=0}^n \xi_k \pmod{d}$, полученные сложением ξ_k по модулю d ($X_n = \sum_{k=0}^n \xi_k - jd$, где j таково, что $0 \leq X_n < d$), образуют цепь Маркова. Действительно, $X_n = X_{n-1} + \xi_n \pmod{d}$, поэтому условное относительно X_1, X_2, \dots, X_{n-1} распределение X_n зависит только от значения X_{n-1} .

Если к тому же $\{\xi_k\}$ одинаково распределены, то эта цепь будет однородной.

Разумеется, все сказанное останется справедливым и при $d = \infty$, т.е. для обычного сложения. Только в этом случае множество возможных состояний системы будет бесконечным.

Итак, для однородной цепи Маркова вероятности $p_{ij}^{(n)}$ перехода из состояния E_i в E_j на n -м шаге не зависят от n . Обозначим эти вероятности p_{ij} . Они образуют матрицу $P = \|p_{ij}\|$ со свойствами:

$$p_{ij} \geq 0, \quad \sum_j p_{ij} = 1.$$

Второе свойство есть следствие того, что из состояния E_i система с вероятностью 1 попадает в одно из состояний E_1, E_2, \dots .

Матрицы с указанными двумя свойствами называются *стохастическими*.

Матрица P полностью описывает изменение состояния системы за один шаг. Рассмотрим теперь изменение состояния системы за k шагов. Введем обозначение $p_{ij}(k) = P(X_k = j | X_0 = i)$. При $k > 1$ по формуле полной вероятности получаем

$$p_{ij}(k) = \sum_s P(X_{k-1} = s | X_0 = i) p_{sj} = \sum_s p_{is}(k-1) p_{sj}.$$

Суммирование здесь ведется по всем состояниям. Если обозначить $P(k) = \|p_{ij}(k)\|$ матрицу, образованную вероятностями $p_{ij}(k)$, то написанное выше равенство будет означать, что $P(k) = P(k-1)P$ или, что то же, $P(k) = P^k$. Таким образом, матрица P однозначно определяет переходные вероятности за любое число шагов. (Здесь следует добавить, что для однородной цепи

$$P(X_{n+k} = j | X_n = i) = P(X_k = j | X_0 = i) = p_{ij}(k).$$

Из сказанного видно, что «распределение» цепи будет полностью определяться заданием матрицы P и начального распределения $p_k^0 = P(X_0 = k)$.

Предоставляем читателю в качестве упражнения проверить, что для любых $k \geq 1$ и множеств B_1, \dots, B_{n-k}

$$P(X_n = j | X_{n-k} = i; X_{n-k-1} \in B_1, \dots, X_0 \in B_{n-k}) = p_{ij}(k).$$

Для доказательства этого соотношения можно сначала проверить его при $k = 1$, а затем воспользоваться индукцией.

Последовательность независимых целочисленных одинаково распределенных случайных величин X_n очевидно образует цепь Маркова, для которой $p_{ij} = p_j = P(X_n = j)$. Здесь $P(k) = P^k = P$.

*Классификация состояний*¹⁾.

Определение 3.

K1. Состояние E_i называется *несущественным*, если существует такое состояние E_j и целое число $t_0 > 0$, что $p_{ij}(t_0) > 0$, $p_{ji}(t) = 0$ для любого целого t .

В противном случае E_i называется *существенным* состоянием.

K2. Существенные состояния E_i и E_j называются *сообщающимися*, если существуют такие целые числа $t > 0$ и $s > 0$, что $p_{ij}(t) > 0$ и $p_{ji}(s) > 0$.

¹⁾ Здесь и в § 2 будем следовать в основном работе А. Н. Колмогорова [12].

Пример 3. Пусть система может находиться в одном из четырех состояний $\{E_1, E_2, E_3, E_4\}$ с матрицей переходных вероятностей

$$P = \begin{pmatrix} 0 & 1/2 & 1/2 & 0 \\ 1/2 & 0 & 0 & 1/2 \\ 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 1/2 & 1/2 \end{pmatrix}$$

(на рис. 9 состояния обозначены точками, переходы из состояний в состояния — стрелками, числа на стрелках означают соответствующие вероятности). В этой цепи состояния E_1 и E_2 несущественны, а E_3 и E_4 являются существенными сообщающимися состояниями.

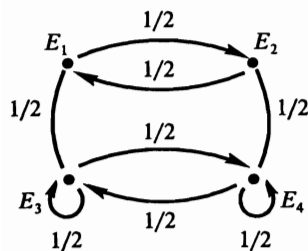


Рис. 9.

В примере 1, описывающем блуждание с поглощением, состояния $1, 2, \dots, a-1$ будут несущественными. Состояния 0 и a существенны, но не сообщающиеся (их естественно называть *поглощающими*). В блуждании с отражением все состояния существенны и сообщающиеся.

Пусть $\{X_n\}_{n=0}^{\infty}$ — однородная цепь Маркова. Выделим класс S^0 всех несущественных состояний. Пусть теперь E_i — какое-нибудь существенное состояние. Обозначим S_{E_i} класс состояний, включающий в себя E_i и все состояния, с ним сообщающиеся. Если $E_j \in S_{E_i}$, то E_j существенно, сообщается с E_i и $E_j \in S_{E_i}$. Следовательно, $S_{E_i} = S_{E_j}$. Таким образом, все множество существенных состояний разбивается на непересекающиеся классы сообщающихся состояний, которые будем обозначать S^1, S^2, \dots .

Определение 4. Если класс S_{E_i} состоит из одного состояния E_i , то это состояние называется *поглощающим*.

Ясно, что если система попала в существенное состояние E_i , то она уже никогда не выйдет из класса S_{E_i} .

Определение 5. Цепи Маркова, состоящие из одного класса существенных сообщающихся состояний, называются *неразложимыми (неприводимыми)*. Если цепь содержит более одного класса, то она называется *разложимой (приводимой)*.

Если состояния перенумеровать так, чтобы сначала шли состояния из S^0 , затем из S^1 и т.д., то матрица переходных вероятностей будет иметь вид (см. рис. 10). Здесь подматрицы, заштрихованные в клеточку, являются стохастическими. Каждая

	S^0	S^1	S^2				
S^0							
S^1	0		0	0	0	0	
S^2	0	0		0	0	0	

Рис. 10.

из них соответствует некоторой неразложимой цепи. Подматрицы, отмеченные нулями, содержат одни нули.

Одним из основных объектов изучения здесь являются *неразложимые цепи Маркова*. Ими и займемся.

Введем следующие обозначения:

$$f_j(n) = P(X_n = j, X_{n-1} \neq j, \dots, X_1 \neq j | X_0 = j),$$

$$F_j = \sum_{n=1}^{\infty} f_j(n);$$

$f_j(n)$ есть вероятность того, что система, выйдя из j -го состояния, впервые вернется в него через n шагов. Вероятность того, что система, выйдя из j -го состояния, вновь когда-нибудь вернется в него, равна F_j .

Определение 6.

К3. Состояние E_j называется *возвратным*, если $F_j = 1$, и *невозвратным*, если $F_j < 1$.

К4. Состояние E_j называется *нулевым*, если $p_{jj}(n) \rightarrow 0$ при $n \rightarrow \infty$, и *ненулевым* в противном случае.

К5. Состояние E_j называется *периодическим* с периодом d_j , если возвращение с положительной вероятностью в E_j возможно только за число шагов, кратное $d_j > 1$, и d_j есть наибольшее число, обладающее этим свойством.

Другими словами, d_j есть наибольший общий делитель (н.о.д.) множества чисел $\{n : f_j(n) > 0\}$. Заметим, что из этого множества всегда можно выбрать конечное подмножество $\{n_1, \dots, n_k\}$ такое, что d_j есть наибольший общий делитель этих чисел. Ясно также, что $p_{jj}(n) = f_j(n) = 0$, если $n \not\equiv 0 \pmod{d_j}$.

Пример 4. Рассмотрим блуждание частицы по целым точкам на вещественной прямой, задаваемое следующим образом: частица либо с вероятностью $1/2$ сдвигается на единицу вправо, либо с той же вероятностью остается на месте. Здесь $f_j(1) = 1/2$, а для $n > 1$ $f_j(n) = 0$ для любой точки j . Поэтому $F_j < 1$, и все состояния невозвратны. Легко видеть, что $p_{jj}(n) = 1/2^n \rightarrow 0$, поэтому все состояния нулевые.

Если же точка с вероятностью $1/2$ смещается вправо и с той же вероятностью влево, то мы получим цепь с периодом 2, так как вернуться в любое фиксированное состояние можно лишь за четное число шагов.

§ 2. Необходимые и достаточные условия возвратности состояний.

Теорема об однотипности состояний неразложимой цепи, структура цепи в периодическом случае

Напомним, что производящей функцией последовательности $\{a_n\}_{n=0}^{\infty}$ называется функция

$$a(z) = \sum_{n=0}^{\infty} a_n z^n,$$

где z — комплексное число. Если последовательность $\{a_n\}$ ограничена, то этот ряд сходится при $|z| < 1$.

Теорема 1. Состояние E_j возвратно тогда и только тогда, когда $P_j = \sum_{n=1}^{\infty} p_{jj}(n) = \infty$. Для невозвратного E_j

$$F_j = \frac{P_j}{1 + P_j}. \quad (2)$$

Утверждение этой теоремы есть своего рода распространение леммы Бореля—Кантелли на случай зависимых событий $A_n = \{X_n = j\}$. События A_n происходят с вероятностью 1 бесконечное число раз тогда и только тогда, когда $\sum_{n=1}^{\infty} P(A_n) = P_j = \infty$.

Доказательство. По формуле полной вероятности имеем

$$p_{jj}(n) = f_j(1)p_{jj}(n-1) + f_j(2)p_{jj}(n-2) + \dots + f_j(n-1)p_{jj}(1) + f_j(n) \cdot 1.$$

Введем производящие функции последовательностей $\{p_{jj}(n)\}$ и $\{f_j(n)\}$:

$$P_j(z) = \sum_{n=1}^{\infty} p_{jj}(n)z^n, \quad F_j(z) = \sum_{n=1}^{\infty} f_j(n)z^n.$$

Оба ряда сходятся внутри единичного круга и являются там аналитическими функциями. Выписанная выше формула для $p_{jj}(n)$ после умножения обеих частей на z^n и суммирования по n приводит к равенству

$$P_j(z) = z f_j(1)(1 + P_j(z)) + z^2 f_j(2)(1 + P_j(z)) + \dots = (1 + P_j(z))F_j(z).$$

Таким образом,

$$F_j(z) = \frac{P_j(z)}{1 + P_j(z)}, \quad P_j(z) = \frac{F_j(z)}{1 - F_j(z)}.$$

Пусть $P_j = \infty$. Тогда $P_j(z) \rightarrow \infty$ при $z \uparrow 1$ и, стало быть, $F_j(z) \rightarrow 1$. Из того, что $F_j(z) < F_j$ для вещественных $z < 1$, получаем, что $F_j = 1$.

Пусть теперь $F_j = 1$. Тогда $F_j(z) \rightarrow 1$ при $z \uparrow 1$, а поэтому $P_j(z) \rightarrow \infty$. Значит, $P_j = \infty$. Если $P_j < \infty$, то при $z = 1$ получаем равенство (2). ■

Число $P_j = \sum_{n=1}^{\infty} p_{jj}(n)$ можно интерпретировать как среднее число попаданий в состояние E_j , если начальным состоянием тоже было E_j . Это следует из того, что число возвращений в E_j можно представить в виде $\sum_{n=1}^{\infty} I(X_n = j)$ (здесь, как и прежде, $I(A)$ — индикатор события A). Поэтому математическое ожидание числа возвращений равно

$$\mathbb{E} \sum_{n=1}^{\infty} I(X_n = j) = \sum_{n=1}^{\infty} \mathbb{E} I(X_n = j) = \sum_{n=1}^{\infty} p_{jj}(n) = P_j.$$

Из теоремы 1 вытекает

Следствие 1. Невозвратное состояние всегда является нулевым.

Это очевидно, поскольку из сходимости ряда $\sum p_{jj}(n) < \infty$ следует $p_{jj}(n) \rightarrow 0$.

Таким образом, на основании определений КЗ–К5 мы могли бы выделить в неразложимой цепи 8 возможных типов состояний (каждый из трех признаков

может быть или не быть). Но на самом деле их оказывается меньше, так как невозвратные состояния автоматически являются нулевыми, а ненулевые — возвратными. Поэтому возможны лишь 6 типов состояний, которые порождаются.

1) Классификацией по асимптотическим свойствам вероятностей $p_{jj}(n)$ (невозвратные состояния, возвратные нулевые, ненулевые).

2) Классификацией по арифметическим свойствам вероятностей $p_{jj}(n)$ или $f_j(n)$ (периодичность и непериодичность).

Теорема 2 (теорема солидарности). *В неразложимой цепи Маркова все состояния принадлежат одному типу: если хоть одно возвратно, то и все возвратны; если хоть одно нулевое, то и все нулевые; если хоть одно периодически с периодом d , то и все периодически с периодом d .*

Доказательство. Пусть E_k и E_j — два различных состояния. Существуют числа N и M такие, что

$$p_{kj}(N) > 0, \quad p_{jk}(M) > 0.$$

Из формулы полной вероятности

$$p_{kk}(N + M + n) = \sum_{l,s} p_{kl}(N) p_{ls}(n) p_{sk}(M)$$

следует неравенство

$$p_{kk}(N + M + n) \geq p_{kj}(N) p_{jj}(n) p_{jk}(M) = \alpha \beta p_{jj}(n).$$

Здесь $n > 0$ — произвольное целое число, $\alpha = p_{kj}(N) > 0$, $\beta = p_{jk}(M) > 0$. Точно так же можно получить неравенство

$$p_{jj}(N + M + n) \geq \alpha \beta p_{kk}(n).$$

Следовательно,

$$\frac{1}{\alpha \beta} p_{kk}(n + M + N) \geq p_{jj}(n) \geq \alpha \beta p_{kk}(n - M - N). \quad (3)$$

Из этих неравенств видно, что асимптотические свойства у $p_{kk}(n)$ и $p_{jj}(n)$ одинаковы. Если E_k — нулевое, т.е. $p_{kk}(n) \rightarrow 0$, то и $p_{jj}(n) \rightarrow 0$. Если E_k — возвратное или, что то же, $P_k = \sum_{n=1}^{\infty} p_{kk}(n) = \infty$, то

$$\sum_{n=M+N+1}^{\infty} p_{jj}(n) \geq \alpha \beta \sum_{n=M+N+1}^{\infty} p_{kk}(n - M - N) = \infty$$

и, значит, E_j — тоже возвратное.

Предположим теперь, что E_k — периодическое состояние с периодом d_k . Если $p_{kk}(n) > 0$, то d_k делит n . Обозначать этот факт будем символом d_k/n . Так как

$$p_{kk}(M + N) \geq \alpha \beta > 0, \quad \text{то} \quad \frac{d_k}{M + N}.$$

Покажем, что состояние E_j тоже периодическое, и период его d_j равен d_k . Действительно, если $p_{jj}(n) > 0$ для некоторого n , то в силу (3) $p_{kk}(n + M + N) > 0$. Следовательно, $d_k/(n + M + N)$, а из того, что $d_k/(M + N)$, мы получаем d_k/n и, следовательно, $d_k \leq d_j$. Теперь аналогичным способом можно доказать, что $d_j \leq d_k$. Таким образом, $d_j = d_k$. ■

Если состояния неразложимой цепи Маркова — периодические с периодом $d > 1$, то цепь называется *периодической*.

Покажем теперь, что изучение периодических цепей в значительной мере может быть сведено к изучению неперiodических. Для этого нам понадобится

Теорема 3. Если цепь Маркова — периодическая с периодом d , то множество состояний разбивается на d подклассов $\Psi_0, \Psi_1, \dots, \Psi_{d-1}$ таких, что с вероятностью 1 за один шаг система из класса Ψ_k переходит в класс Ψ_{k+1} , из класса Ψ_{d-1} система переходит в Ψ_0 .

Доказательство. Выберем какое-нибудь состояние, например, E_1 . С помощью этого состояния построим подклассы $\Psi_0, \Psi_1, \dots, \Psi_{d-1}$ следующим образом: $E_i \in \Psi_\alpha$, $0 \leq \alpha \leq d-1$, если существует целое число $k > 0$ такое, что $p_{1i}(kd + \alpha) > 0$.

Покажем, что никакое состояние не может принадлежать сразу двум подклассам. Для этого достаточно доказать следующее утверждение: если $E_i \in \Psi_\alpha$ и для некоторого s $p_{1i}(s) > 0$, то $s \equiv \alpha \pmod{d}$.

Действительно, существует число $t_1 > 0$, для которого $p_{11}(t_1) > 0$. Тогда в силу определения Ψ_α имеем $p_{11}(kd + \alpha + t_1) > 0$. Кроме того, $p_{11}(s + t_1) > 0$. Следовательно, $d/(kd + \alpha + t_1)$ и $d/(s + t_1)$. Это влечет за собой $\alpha \equiv s \pmod{d}$.

Так как из состояния E_1 можно с положительной вероятностью попасть в любое состояние E_i , то $\bigcup_{\alpha} \Psi_\alpha$ содержит все множество состояний.

Докажем теперь, что система из Ψ_α за 1 шаг с вероятностью 1 переходит в $\Psi_{\alpha+1}$ ($\alpha + 1$ понимается по \pmod{d}). Для этого надо показать, что для $E_i \in \Psi_\alpha$

$$\sum_{E_j \in \Psi_{\alpha+1}} p_{ij} = 1.$$

Для этого в свою очередь достаточно показать, что $p_{ij} = 0$, если $E_i \in \Psi_\alpha$, $E_j \notin \Psi_{\alpha+1}$. Если допустить противное, то, учитывая неравенство $p_{1i}(kd + \alpha) > 0$, получим $p_{1j}(kd + \alpha + 1) > 0$, и, следовательно, $E_j \in \Psi_{\alpha+1}$. Полученное противоречие заканчивает доказательство теоремы. ■




	Ψ_0	Ψ_1	Ψ_{d-1}
Ψ_0	0		0
Ψ_1	0	0	
	0	0	0
Ψ_{d-1}		0	0

Рис. 11.

Из доказанной теоремы видно, что матрица периодической цепи будет иметь вид (рис. 11), где отличны от нуля лишь элементы заштрихованных клеток.

Из периодической цепи Маркова с периодом d можно образовать d новых цепей Маркова. Состояниями α -й цепи будут состояния из подкласса Ψ_α . Вероятности перехода задаются равенствами

$$p_{ij}^\alpha = p_{ij}(d).$$

В силу только что доказанной теоремы $\sum_{E_j \in \Psi_\alpha} p_{ij}^\alpha = 1$.

Новая цепь, к которой в известном смысле сводится изучение исходной, уже не будет иметь подклассов.

§ 3. Теоремы о случайных блужданиях по решетке

1. Случайное блуждание по целым точкам на прямой. Представим себе частицу, которая движется по целым точкам на прямой. Перемещение из одной точки в другую происходит через равные промежутки времени. За один шаг частица из точки k

с положительной вероятностью p перемещается в точку $k + 1$ и с положительной вероятностью $q = 1 - p$ перемещается в точку $k - 1$. Как уже отмечалось, описанной физической системе соответствует цепь Маркова

$$X_n = X_{n-1} + \xi_n = X_0 + S_n,$$

где ξ_n принимает значения 1 и -1 с вероятностями p и q , $S_n = \sum_{k=1}^n \xi_k$. Состояниями здесь служат целые точки на прямой.

Легко видеть, что вернуться в какую-либо точку с положительной вероятностью можно только за четное число шагов. При этом $f_0(2) = 2pq > 0$. Поэтому это периодическая цепь с периодом 2.

Исследуем теперь, когда случайное блуждание образует возвратную цепь. Имеет место следующая

Теорема 4. *Описанное случайное блуждание образует возвратную цепь Маркова тогда и только тогда, когда $p = q = 1/2$.*

Доказательство. Так как $0 < p < 1$, то блуждание представляет собой неразложимую цепь Маркова. Поэтому в силу теоремы 2 достаточно изучить тип одной точки, например, нуля.

Воспользуемся теоремой 1. Для этого надо исследовать сходимость ряда $\sum_{n=1}^{\infty} p_{00}(n)$. Так как цепь — периодическая с периодом 2, то $p_{00}(2k+1) = 0$. Остается

подсчитать $\sum_{k=1}^{\infty} p_{00}(2k)$. Сумма S_n есть координата блуждающей частицы ($X_0 = 0$) после n шагов. Поэтому $p_{00}(2k) = P(S_{2k} = 0)$. Чтобы $S_{2k} = 0$, необходимо, чтобы k суммируемых случайных величин равнялись 1, а оставшиеся k случайных величин равнялись -1 (k шагов вправо и k шагов влево). Поэтому по теореме 2 гл. 5

$$P(S_{2k} = 0) \sim \frac{1}{\sqrt{\pi k}} e^{-2kH(1/2)} = \frac{1}{\sqrt{\pi k}} (4pq)^k.$$

Вясним поведение функции $\varphi(p) = 4pq = 4p(1-p)$ на отрезке $[0, 1]$. В точке $p = 1/2$ функция $\varphi(p)$ достигает единственного экстремума, а $\varphi(1/2) = 1$. Во всех остальных точках отрезка $[0, 1]$ $\varphi(p) < 1$. Значит, $4pq < 1$ для $p \neq 1/2$, а это влечет за собой сходимость ряда $\sum_{k=1}^{\infty} p_{00}(2k)$ и невозвратность цепи. Если же $p = 1/2$, то $p_{00}(2k) \sim 1/\sqrt{\pi k}$, и ряд $\sum_{k=1}^{\infty} p_{00}(2k)$ расходится, что, в свою очередь, влечет за собой возвратность всех состояний цепи. ■

Доказанная теорема позволяет сделать также следующее замечание. Если $p \neq 1/2$, то среднее число возвращений в 0 конечно, ибо оно равно $\sum_{k=1}^{\infty} p_{00}(2k)$. Это означает, что, начиная с некоторого шага, частица уже никогда больше не вернется в 0. Частицу «снесет» вправо или влево от нуля в зависимости от того, $p > 1/2$ или $p < 1/2$. Это легко получить и из усиленного закона больших чисел.

Если $p = 1/2$, то среднее число возвращений в 0 бесконечно; частицу не «сносит». Интересно отметить, что при этом среднее число возвращений растет не пропорционально числу шагов. Действительно, среднее число возвращений за

первые $2n$ шагов есть $\sum_{k=1}^n p_{00}(2k)$. Из доказательства теоремы 4 нам известно, что $p_{00}(2k) \sim 1/\sqrt{\pi k}$. Поэтому

$$\sum_{k=1}^n p_{00}(2k) \sim \sum_{k=1}^n \frac{1}{\sqrt{\pi k}} \sim \frac{2\sqrt{n}}{\sqrt{\pi}}.$$

Таким образом, в безобидной игре, рассмотренной нами в примере 7 гл. 4, доля «ничьих» результатов с увеличением числа шагов быстро падает, а отклонения по величине и по длительности возрастают.

2. Симметричные случайные блуждания в \mathbb{R}^k , $k \geq 2$. Будем рассматривать следующую модель случайного блуждания в k -мерном евклидовом пространстве \mathbb{R}^k . Если частица находится в точке (m_1, \dots, m_k) , то она может переместиться с вероятностью $1/2^k$ в любую из 2^k вершин куба $|x_j - m_j| = 1$, т. е. в точки с координатами $(m_1 \pm 1, \dots, m_k \pm 1)$. Такое блуждание естественно называть симметричным. Если обозначить X_n положение частицы после n -го скачка, то мы, как и прежде, получим последовательность k -мерных случайных величин, связанных в однородную неразложимую цепь Маркова. Покажем, что все состояния этой цепи для блуждания на плоскости, как и в одномерном случае, возвратны. В трехмерном пространстве состояния окажутся уже невозвратными. Будет доказана, таким образом,

Теорема 5. *Симметричное случайное блуждание возвратно в пространствах одного и двух измерений и невозвратно в пространстве трех и более измерений.*

По этому поводу В. Феллер весьма удачно заметил, что поговорка «все дороги ведут в Рим» верна только для двумерной поверхности. Сформулированное нами утверждение примыкает к известной теореме Пойа о невозвратности при $k \geq 2$ симметричных блужданий в \mathbb{R}^k , когда частица за один шаг перемещается на ± 1 по одной из осей координат (ξ_j принимают $2k$ значений с вероятностями $1/2k$).

Обратимся к доказательству теоремы 5. Пусть $k = 2$. Нетрудно видеть, что наше блуждание X_n можно представить в виде суммы двух независимых компонент

$$X_n = (X_n^1, 0) + (0, X_n^2), \quad (X_0^1, X_0^2) = X_0,$$

где X_n^i , $i = 1, 2$ — уже скалярные (одномерные) последовательности, описывающие симметричные независимые случайные блуждания на прямых. Это ясно, так как двумерная последовательность допускает представление

$$X_{n+1} = X_n + \xi_n, \quad (4)$$

где ξ_n принимает 4 значения $(\pm 1, 0) + (0, \pm 1) = (\pm 1, \pm 1)$ с вероятностью $1/4$ каждое.

С помощью представления (4) мы можем исследовать асимптотическое поведение переходных вероятностей $p_{ij}(n)$. Пусть состояние X_0 совпадает с началом координат $(0, 0)$. Тогда

$$\begin{aligned} p_{00}(2n) &= P(X_{2n} = (0, 0) | X_0 = (0, 0)) = \\ &= P(X_{2n}^1 = 0 | X_0^1 = 0) P(X_{2n}^2 = 0 | X_0^2 = 0) \sim \left(\frac{1}{\sqrt{\pi n}} \right)^2 = \frac{1}{\pi n}. \end{aligned}$$

Отсюда следует, что ряд $\sum_{n=0}^{\infty} p_{00}(n)$ расходится, и все состояния нашей цепи возвратны.

В случае $k = 3$ следует поступить аналогичным образом, представив последовательность X_n в виде суммы трех независимых компонент

$$X_n = (X_n^1, 0, 0) + (0, X_n^2, 0) + (0, 0, X_n^3),$$

где X_n^i , как и прежде, — симметричные случайные блуждания на прямой. Если положить $X_0 = (0, 0, 0)$, то получим

$$p_{00}(2n) = (P(X_{2n}^1 = 0 | X_0^1 = 0))^3 \sim \frac{1}{(\pi n)^{3/2}}.$$

Здесь ряд $\sum_{n=1}^{\infty} p_{00}(n)$ будет уже сходящимся, а состояния цепи невозвратными.

В отличие от случая прямой и плоскости здесь частица, выйдя из начала координат, может с положительной вероятностью никогда туда не вернуться.

Ясно, что эта картина сохранится и для блуждания в любом k -мерном пространстве при $k \geq 3$, так как $\sum_{n=1}^{\infty} (\pi n)^{-k/2} < \infty$ при $k \geq 3$. ■

3. Произвольное симметричное случайное блуждание на прямой. Пусть, как и прежде,

$$X_n = X_0 + \sum_{j=1}^n \xi_j, \quad (5)$$

но ξ_j здесь уже произвольные независимые одинаково распределенные целочисленные случайные величины. Теорема 4 может быть обобщена следующим образом:

Теорема 6. Если ξ_j симметричны, существует $E\xi_j$ ($E\xi_j = 0$), то случайное блуждание X_n образует возвратную цепь Маркова с нулевыми состояниями.

Доказательство. Достаточно убедиться, что

$$\sum_{n=1}^{\infty} P(S_n = 0) = \infty,$$

где $S_n = \sum_{j=1}^n \xi_j$, и что $P(S_n = 0) \rightarrow 0$ при $n \rightarrow \infty$. Обозначим $p(z) = Ez^{\xi_1} = \sum_{k=-\infty}^{\infty} z^k P(\xi_1 = k)$. Тогда производящая функция S_n будет равна $Ez^{S_n} = p^n(z)$, и по формуле обращения (см. § 7.8)

$$P(S_n = 0) = \frac{1}{2\pi i} \int_{|z|=1} p^n(z) z^{-1} dz, \quad (6)$$

$$\sum_{n=0}^{\infty} P(S_n = 0) = \frac{1}{2\pi i} \int_{|z|=1} \frac{dz}{z(1-p(z))} = \frac{1}{\pi} \int_0^\pi \frac{dt}{1-p(e^{it})}.$$

Последнее равенство справедливо в силу четности вещественной функции $p(z)$ и получено путем замены $z = e^{it}$.

Так как $E\xi_1 = 0$, то $1 - p(e^{it}) = o(t)$ при $t \rightarrow 0$, и при достаточно малом δ и $0 \leq t < \delta$

$$0 \leq 1 - p(e^{it}) < t.$$

Отсюда следует

$$\int_0^{\pi} \frac{dt}{1 - p(e^{it})} \geq \int_0^{\delta} \frac{dt}{t} = \infty.$$

Сходимость $P(S_n = 0) \rightarrow 0$ следует из (6), так как для всех z на окружности $|z| = 1$, за исключением, быть может, конечного числа точек, $p(z) < 1$, $p^n(z) \rightarrow 0$ при $n \rightarrow \infty$. ■

Теорему 6 можно дополнить следующим образом.

Теорема 7. Если н. о. д. возможных значений ξ_j равен 1, то множество значений $\{X_n\}$ образует один класс существенных сообщающихся состояний, совпадающий с множеством всех целых чисел.

Утверждение теоремы вытекает из следующей леммы:

Лемма 1. Если н. о. д. целых чисел $a_1 > 0, \dots, a_r > 0$ равен 1, то существует N такое, что любое натуральное число $k \geq N$ можно представить в виде

$$k = x_1 a_1 + \dots + x_r a_r,$$

где $x_i \geq 0$ — целые числа.

Доказательство. Рассмотрим функцию $L(x) = x_1 a_1 + \dots + x_r a_r$, где $x = (x_1, \dots, x_r)$ есть вектор с целыми координатами. Пусть $d > 0$ — минимальное натуральное число, для которого существует такой вектор x^0 , что

$$d = L(x^0).$$

Покажем, что любое натуральное число, представимое в виде $L(x)$, делится на d . Пусть это не так. Тогда существуют x, k и $0 < \alpha < d$, для которых

$$L(x) = kd + \alpha.$$

Но, так как функция $L(x)$ линейна, то

$$L(x - kx^0) = kd + \alpha - kd = \alpha < d,$$

что противоречит минимальности d в множестве положительных целых значений $L(x)$.

Все числа a_1, \dots, a_r также являются значениями функции $L(x)$, и поэтому все делятся на d . Так как, по предположению, наибольший делитель этих чисел равен единице, то $d = 1$.

Пусть k — любое натуральное число. Тогда, обозначив через $\theta < A$ остаток от деления k на $A = a_1 + \dots + a_r$, мы можем записать

$$\begin{aligned} k &= t(a_1 + \dots + a_r) + \theta = t(a_1 + \dots + a_r) + \theta L(x^0) = \\ &= a_1(t + \theta x_1^0) + a_2(t + \theta x_2^0) + \dots + a_r(t + \theta x_r^0), \end{aligned}$$

где $t + \theta x_i^0 > 0$, $i = 1, \dots, r$, при достаточно больших k . ■

Следствие 2. Если выполнены условия теорем 6, 7, то цепь (5) при любом начальном состоянии X_0 с вероятностью 1 побывает бесконечное число раз в любом состоянии k . В частности, случайная величина $\nu = \min\{n : X_n = k\}$ для любых X_0 и k будет собственной.

Если интересоваться периодичностью цепи (5), то нужны более полные сведения о множестве возможных значений ξ_j . Мы предоставляем читателю

убедиться, например, в том, что если множество возможных значений ξ_j имеет вид $\{a + a_k d\}$, $k = 1, 2, \dots, d \geq 1$, н. о. д. $(a_1, a_2, \dots) = 1$, н. о. д. $(a, d) = 1$, то цепь будет периодической с периодом d .

§ 4. Эргодические теоремы

Нам понадобятся следующие условия:

(I) Существует состояние E_s такое, что время $\tau^{(s)}$ возвращения в E_s ($P(\tau^{(s)} = n) = f_s(n)$) имеет конечное математическое ожидание $E\tau^{(s)} < \infty$.

(II) Цепь неразложима.

(III) Цепь неперiodична.

Введем в рассмотрение так называемые «табу-вероятности» $P_i(n, j)$ перехода из E_i в E_j за n шагов, минуя возвращение в «запретное» состояние E_i :

$$P_i(n, j) = P\left(X_n = j; X_1 \neq i, \dots, X_{n-1} \neq \frac{i}{X_0} = i\right).$$

Теорема 8 (эргодическая). Условия (I)–(III) необходимы и достаточны для того, чтобы при любых i и j существовали не зависящие от i положительные пределы

$$\lim_{n \rightarrow \infty} p_{ij}(n) = \pi_j > 0, \quad i, j = 0, 1, 2, \dots \quad (7)$$

Числа $\{\pi_i\}$ являются единственным в классе последовательностей, образующих абсолютно сходящиеся ряды, решением системы уравнений

$$\begin{cases} \sum_{n=0}^{\infty} \pi_j = 1, \\ \pi_j = \sum_{n=0}^{\infty} \pi_k p_{kj}, \quad j = 0, 1, 2, \dots \end{cases} \quad (8)$$

Кроме того, $E\tau^{(j)} < \infty$ при всех j и числа $\pi_j = (E\tau^{(j)})^{-1}$ допускают представление

$$\pi_j = (E\tau^{(j)})^{-1} = (E\tau^{(s)})^{-1} \sum_{k=1}^{\infty} P_s(k, j) \quad (9)$$

при любом s .

Цепь, обладающая свойством (7), называется эргодической.

Числа π_j есть, по существу, вероятности попадания системы в состояние E_j через большой интервал времени. При этом оказывается, что эти вероятности не зависят от начального состояния системы. Система «забывает», откуда началось движение. Распределение $\{\pi_j\}$ называют *стационарным или инвариантным*. Свойство (8) выражает *инвариантность* распределения относительно переходных вероятностей p_{ij} . Другими словами, если $P(X_n = k) = \pi_k$, то $P(X_{n+1} = k) = \sum_j \pi_j p_{jk}$ также равны π_k .

Доказательство. Достаточность. Рассмотрим «траекторию» цепи Маркова, начинающуюся в фиксированном состоянии E_s , и рассмотрим интервалы времени $\tau_1 \geq 1, \tau_2 \geq 1, \dots$, через которые система возвращается в E_s . Так как каждый раз эволюция системы начинается из одного и того же фиксированного состояния, то длительности циклов τ_k , образованных попаданиями системы в E_s , будут в силу

(см. теорему 3). Если же $t = kd$, то из только что доказанной теоремы и из теоремы 3 следует, что существует предел $\lim_{k \rightarrow \infty} p_{ij}(kd) = \pi_j > 0$ и не зависит от i .

Проверка условий (II), (III) теоремы 8 обычно не представляет собой сколь угодно существенных трудностей: основные трудности бывают связаны с проверкой условия (I). Для конечных цепей Маркова это условие выполнено всегда.

Теорема 9. Пусть цепь Маркова имеет конечное число состояний и удовлетворяет условиям (II), (III). Тогда существуют постоянные $c > 0$ и $q < 1$ такие, что для времени возвращения τ в любое фиксированное состояние справедливы неравенства

$$P(\tau > n) < cq^n. \quad (12)$$

Эти неравенства означают, очевидно, что условие (I) для конечных цепей всегда выполнено и, следовательно, эргодическая теорема для них верна тогда и только тогда, когда выполнены условия (II), (III).

Доказательство. Рассмотрим некоторое состояние E_0 и обозначим

$$r_j(n) = P(X_k \neq 0, \quad k = 1, 2, \dots, n | X_0 = j).$$

Тогда, если цепь имеет m состояний, то $r_j(m) < 1$ при любом j . Действительно, $r_j(n)$ не возрастает с ростом n . Пусть N есть наименьшее число, для которого $r_j(N) < 1$. Это означает, что существует цепочка состояний $E_j, E_{j_1}, \dots, E_{j_N}$ такая, что $E_{j_N} = E_0$, и вероятность этой цепочки $p_{jj_1} \cdots p_{j_{N-1}j_N}$ положительна. Но легко видеть, что $N \leq m$, так как в противном случае указанная цепочка содержала бы по крайней мере два одинаковых состояния, а следовательно, и цикл, заключенный между этими состояниями, который можно было бы из цепочки выбросить, лишь увеличив ее вероятность. Таким образом, $r_j(m) < 1$,

$$r(m) = \max_j r_j(m) < 1.$$

Кроме того, $r_j(n_1 + n_2) \leq r_j(n_1)r(n_2) \leq r(n_1)r(n_2)$.

Остается заметить, что если τ есть время возвращения в E_0 , то $P(\tau > nm) = r_0(nm) \leq r(m)^n$. Отсюда следует утверждение теоремы. ■

Замечание 1. Выполнение условия (12) влечет за собой экспоненциальную скорость сходимости к нулю разностей $|p_{ij}(n) - \pi_j|$. В этом можно убедиться, используя аналитичность функции $F_0(z) = \sum_{n=1}^{\infty} f_0(n)z^n$ в области $|z| < q^{-1}$, $q^{-1} > 1$ и равенство

$$P_0(z) = \sum p_{00}(n)z^n = \frac{1}{1 - F_0(z)} - 1 \quad (13)$$

(см. теорему 1; мы считаем, что τ в условии (12) относится к состоянию E_0 , так что $f_0(n) = P(\tau = n)$). Так как $F_0'(1) = E\tau = 1/\pi_0$, то $F_0(z) = 1 + (z-1)/\pi_0 + \dots$, и из полученного равенства будет следовать аналитичность в области $|z| \leq 1 + \varepsilon$, $\varepsilon > 0$ функции $P_0(z) - (z\pi_0)/(1-z) = \sum_{n=1}^{\infty} (p_{00}(n) - \pi_0)z^n$, которая, очевидно, и будет означать, что

$$|p_{00}(n) - \pi_0| < c(1 + \varepsilon)^{-n}, \quad c = \text{const}.$$

Приведем теперь два примера конечных цепей Маркова.

Действительно, состояние E_i , как и E_s , возвратно. Рассмотрим циклы, образованные возвращением системы в состояние E_i , и обозначим A_k событие, состоящее в том, что в течение k -го цикла система хоть раз попадает в E_s . Тогда в силу марковости события A_k независимы и $P(A_k) = p > 0$. Следовательно, по критерию Бореля—Кантелли (см. § 10.1) с вероятностью 1 произойдет бесконечно много событий A_k , и тем более $P(\bigcup A_k) = 1$.

По формуле полной вероятности находим

$$p_{ij}(n) = \sum_{k=1}^n f_{is}(k) p_{sj}(n-k).$$

Использование теоремы мажорируемой сходимости дает

$$\lim_{n \rightarrow \infty} p_{ij}(n) = \sum_{k=1}^{\infty} f_{is}(k) \pi_j = \pi_j.$$

Так как состояние E_j , так же как и E_s , является непериодическим и ненулевым, то

$$\pi_j = \lim_{n \rightarrow \infty} p_{jj}(n) > 0.$$

Представление (9) следует из (11).

Докажем теперь *необходимость*. Необходимость условий (II), (III) очевидна, так как $p_{ij}(n) > 0$ для каждого i и j при всех достаточно больших n . Необходимость условия (I) следует из справедливости для E_s равенств (10).

Осталось доказать вторую часть теоремы. Так как

$$\sum_j p_{sj}(n) = 1,$$

то $\sum \pi_j \leq 1$. Поэтому в последнем равенстве и в равенстве $p_{sj}(n+1) = \sum_{k=0}^{\infty} p_{sk}(n) p_{kj}$ в силу неравенств $p_{sj}(n) \leq \pi_j E \tau_1$ можно воспользоваться теоремой о мажорируемой сходимости, откуда следует, что

$$\sum \pi_j = 1, \quad \pi_j = \sum_{k=0}^{\infty} \pi_k p_{kj}.$$

Остается показать, что полученная система имеет единственное решение. Пусть числа $\{q_j\}$ также удовлетворяют (8) и ряд $\sum |q_j|$ сходится. Тогда, меняя порядок суммирования, получим

$$\begin{aligned} q_j &= \sum_k q_k p_{kj} = \sum_k p_{kj} \left(\sum_l p_{lk} q_l \right) = \sum_l q_l \sum_k p_{lk} p_{kj} = \sum_l q_l p_{lj}(2) = \\ &= \sum_l p_{lj}(2) \left(\sum_s p_{sl} q_s \right) = \sum_s q_s p_{sj}(3) = \dots = \sum_k q_k p_{kj}(n) \end{aligned}$$

для любого n . Так как $\sum q_k = 1$, то, переходя к пределу при $n \rightarrow \infty$, получим

$$q_j = \sum_k q_k \pi_j = \pi_j.$$

Если цепь Маркова — периодическая с периодом d , то $p_{ij}(t) = 0$, если $t \neq kd$ для любой пары состояний E_i и E_j , принадлежащих одному подклассу

(см. теорему 3). Если же $t = kd$, то из только что доказанной теоремы и из теоремы 3 следует, что существует предел $\lim_{k \rightarrow \infty} p_{ij}(kd) = \pi_j > 0$ и не зависит от i .

Проверка условий (II), (III) теоремы 8 обычно не представляет собой сколь угодно существенных трудностей: основные трудности бывают связаны с проверкой условия (I). Для конечных цепей Маркова это условие выполнено всегда.

Теорема 9. Пусть цепь Маркова имеет конечное число состояний и удовлетворяет условиям (II), (III). Тогда существуют постоянные $c > 0$ и $q < 1$ такие, что для времени возвращения τ в любое фиксированное состояние справедливы неравенства

$$P(\tau > n) < cq^n. \quad (12)$$

Эти неравенства означают, очевидно, что условие (I) для конечных цепей всегда выполнено и, следовательно, эргодическая теорема для них верна тогда и только тогда, когда выполнены условия (II), (III).

Доказательство. Рассмотрим некоторое состояние E_0 и обозначим

$$r_j(n) = P(X_k \neq 0, \quad k = 1, 2, \dots, n | X_0 = j).$$

Тогда, если цепь имеет m состояний, то $r_j(m) < 1$ при любом j . Действительно, $r_j(n)$ не возрастает с ростом n . Пусть N есть наименьшее число, для которого $r_j(N) < 1$. Это означает, что существует цепочка состояний $E_j, E_{j_1}, \dots, E_{j_N}$ такая, что $E_{j_N} = E_0$, и вероятность этой цепочки $p_{jj_1} \cdots p_{j_{N-1}j_N}$ положительна. Но легко видеть, что $N \leq m$, так как в противном случае указанная цепочка содержала бы по крайней мере два одинаковых состояния, а следовательно, и цикл, заключенный между этими состояниями, который можно было бы из цепочки выбросить, лишь увеличив ее вероятность. Таким образом, $r_j(m) < 1$,

$$r(m) = \max_j r_j(m) < 1.$$

Кроме того, $r_j(n_1 + n_2) \leq r_j(n_1)r(n_2) \leq r(n_1)r(n_2)$.

Остается заметить, что если τ есть время возвращения в E_0 , то $P(\tau > nm) = r_0(nm) \leq r(m)^n$. Отсюда следует утверждение теоремы. ■

Замечание 1. Выполнение условия (12) влечет за собой экспоненциальную скорость сходимости к нулю разностей $|p_{ij}(n) - \pi_j|$. В этом можно убедиться, используя аналитичность функции $F_0(z) = \sum_{n=1}^{\infty} f_0(n)z^n$ в области $|z| < q^{-1}$, $q^{-1} > 1$ и равенство

$$P_0(z) = \sum p_{00}(n)z^n = \frac{1}{1 - F_0(z)} - 1 \quad (13)$$

(см. теорему 1; мы считаем, что τ в условии (12) относится к состоянию E_0 , так что $f_0(n) = P(\tau = n)$). Так как $F'_0(1) = E\tau = 1/\pi_0$, то $F_0(z) = 1 + (z - 1)/\pi_0 + \dots$, и из полученного равенства будет следовать аналитичность в области $|z| \leq 1 + \varepsilon$, $\varepsilon > 0$ функции $P_0(z) - (z\pi_0)/(1 - z) = \sum_{n=1}^{\infty} (p_{00}(n) - \pi_0)z^n$, которая, очевидно, и будет означать, что

$$|p_{00}(n) - \pi_0| < c(1 + \varepsilon)^{-n}, \quad c = \text{const}.$$

Приведем теперь два примера конечных цепей Маркова.

Пример 5. Допустим, что поведение двух шахматистов A и B в турнирной игре описывается следующим образом. Шахматист A каждую партию независимо от исходов предыдущих партий выигрывает с вероятностью p , проигрывает с вероятностью q и делает ничью с вероятностью $r = 1 - p - q$. Шахматист B менее уравновешен, он выигрывает партию с вероятностями $p + \varepsilon$, p и $p - \varepsilon$ соответственно, если предыдущая партия была им выиграна, сыграна вничью, проиграна. Аналогично ведет себя вероятность проигрыша: она равна в этих трех случаях соответственно $q - \varepsilon$, q , $q + \varepsilon$. Кто из игроков A и B в длительном турнире наберет больше очков?

Чтобы ответить на этот вопрос, надо вычислить для каждого из шахматистов стационарные вероятности π_1, π_2, π_3 попадания в состояния E_1, E_2, E_3 , которые мы отождествим соответственно с выигрышем, ничьей, проигрышем (ср. с законом больших чисел в конце параграфа).

Для игрока A цепь Маркова с состояниями E_1, E_2, E_3 , описывающая ход его игры, будет иметь матрицу переходных вероятностей

$$P_A = \begin{pmatrix} p & r & q \\ p & r & q \\ p & r & q \end{pmatrix}.$$

Очевидно, что здесь $\pi_1 = p$, $\pi_2 = r$, $\pi_3 = q$.

Для игрока B матрица переходных вероятностей будет равна

$$P_B = \begin{pmatrix} p + \varepsilon & r & q - \varepsilon \\ p & r & q \\ p - \varepsilon & r & q + \varepsilon \end{pmatrix}.$$

Уравнения для стационарных вероятностей в этом случае будут иметь вид

$$\pi_1(p + \varepsilon) + \pi_2 p + \pi_3(p - \varepsilon) = \pi_1,$$

$$\pi_1 r + \pi_2 r + \pi_3 r = \pi_2,$$

$$\pi_1 + \pi_2 + \pi_3 = 1.$$

Решая эту систему, находим

$$\pi_2 - r = 0, \quad \pi_1 - p = \varepsilon \frac{p - q}{1 - 2\varepsilon}.$$

Таким образом, доли ничьих при большом числе партий у игроков A и B будут совпадать, доля выигрышей у игрока B будет больше, чем у A , если $\varepsilon > 0$, $p > q$ или $\varepsilon < 0$, $p < q$. Если $p = q$, то стационарные распределения для исходов партий у игроков A и B одинаковы.

Пример 6. Рассмотрим сложение независимых целочисленных случайных величин ξ_1, ξ_2, \dots по некоторому модулю $d > 1$ (см. пример 2). Мы положим ($|x|$ означает целую часть x) $X_0 = 0$, $X_1 = \xi_1 - [\xi_1/d]d$, $X_2 = X_1 + \xi_2 - [(X_1 + \xi_2)/d]d$ и т.д., так что X_n есть остаток от деления на d суммы $X_{n-1} + \xi_n$. Такое суммирование иногда называют суммированием на окружности (точки 0 и d склеены в одну). Не ограничивая общности, можем считать, очевидно, что ξ_k принимают лишь значения $0, 1, \dots, d - 1$. Если $P(\xi_k = j) = q_j$, то

$$p_{ij} = P(X_n = j | X_{n-1} = i) = \begin{cases} q_{j-i}, & \text{если } j \geq i, \\ q_{d+j-i}, & \text{если } j < i. \end{cases}$$

Допустим, что множество индексов k , для которых $q_k > 0$, имеет н.о.д., равный 1. Тогда ясно, что цепь $\{X_n\}$ будет представлять собой один класс существенных состояний без подклассов и будут существовать

$$\lim_{n \rightarrow \infty} p_{ij}(n) = \pi_j,$$

удовлетворяющие системе уравнений $\sum_i \pi_i p_{ij} = \pi_j$, $\sum_j \pi_j = 1$, $j = 0, \dots, d-1$.

Заметим теперь, что стохастическая матрица переходных вероятностей $\|p_{ij}\|$ в нашем случае обладает свойствами

$$\sum_i p_{ij} = \sum_j p_{ij} = 1.$$

Такие матрицы называют *дважды стохастическими*. Для них стационарное распределение всегда является равномерным, так как $\pi_j = 1/d$ удовлетворяют системе уравнений для финальных вероятностей.

Таким образом, сложение произвольных случайных величин на окружности приводит к *равномерному предельному распределению*. При этом скорость сходимости $p_{ij}(k)$ к стационарному распределению будет показательной.

Нетрудно заметить, что при суммировании по модулю композиция двух равномерных распределений снова будет равномерной. В этом смысле равномерное распределение является устойчивым. Более того, композиция любого распределения с равномерным также будет равномерной. Действительно, если η равномерно распределена и не зависит от ξ_1 , то (сложение и вычитание здесь происходит по модулю d , $p_j = P(\xi_1 = j)$)

$$P(\xi_1 + \eta = k) = \sum_{j=0}^{d-1} \pi_j P(\eta = k - j) = \sum_{j=0}^{d-1} \pi_j \frac{1}{d} = \frac{1}{d}.$$

Таким образом, если передается некоторый сигнал, принимающий d возможных значений (например, буквы), и на него накладывается «случайный» (равномерный) шум, то принимаемый сигнал также будет иметь равномерное распределение и, стало быть, не будет содержать никакой информации о передаваемом сигнале.

Рассмотренный пример заслуживает внимания и как простейшая иллюстрация законов, которые возникают при «суммировании» случайных величин со значениями уже не на числовой оси, а в некоторой группе (множество чисел $0, 1, \dots, d-1$ с операцией сложения по модулю d образуют конечную абелеву группу). Оказывается, что явление, обнаруженное нами в примере, — равномерность предельного распределения — сохраняется и для значительно более широкого класса групп.

Вернемся к произвольным счетным цепям. Мы уже отмечали, что основные трудности при проверке условий теоремы 8 бывают связаны с условием (I). Подробно этот вопрос будет рассмотрен в § 7 для более широкого класса цепей (см. теоремы 14, 15 и их следствия). В некоторых случаях это можно сделать без труда с помощью результатов глав 9, 11.

Пример 7. В § 11.4 мы видели, что времена ожидания в системе обслуживания удовлетворяют соотношениям

$$X_{n+1} = \max(X_n + \xi_{n+1}, 0), \quad X_1 = 0,$$

где ξ_n независимы и одинаково распределены. Ясно, что X_n связаны в однородную цепь Маркова с множеством состояний $\{0, 1, \dots\}$, если ξ_k целочисленны. Последовательность X_n можно интерпретировать также как *блуждание с задерживающим экраном* в точке 0. Если $E\xi_k < 0$, то из теорем гл. 9 нетрудно получить (подробнее см. в § 7), что время возвращения в состояние 0 имеет конечное математическое ожидание. Поэтому, применяя эргодическую теорему, мы можем независимо от § 11.4 сделать заключение о существовании предельного (стационарного) распределения для X_n при $n \rightarrow \infty$ (или, принимая во внимание сказанное в § 11.4, — заключение о конечности $\sup_{k \geq 0} S_k$, $S_k = \sum_{j=1}^k \xi_j$, что составляет содержание теоремы 10.4).

Сделаем теперь ряд замечаний, позволяющих сформулировать еще один критерий эргодичности цепи, связанный с существованием решения уравнения (8).

Заметим прежде всего, что теорема 2 (солидарности) может быть дополнена теперь следующим образом. Состояние E_j мы назовем *эргодическим*, если при любом i $p_{ij}(n) \rightarrow \pi_j > 0$. Состояние E_j назовем *положительно возвратным*, если оно возвратно и ненулевое (в этом случае время $\tau^{(j)}$ возвращение в E_j имеет конечное математическое ожидание $E\tau^{(j)} < \infty$). Из теоремы 8 следует тогда, что для неразложимой неперiodической цепи *состояние E_j эргодично тогда и только тогда, когда оно положительно возвратно. Если хоть одно состояние цепи эргодично, то и все эргодичны.*

Теорема 8А. Пусть цепь неразложима и неперiodична (удовлетворяет условиям (II), (III)). Тогда возможно лишь одно из двух: либо все состояния цепи являются нулевыми, либо эргодическими. Существование абсолютно сходящегося решения уравнений (8) необходимо и достаточно для эргодичности цепи.

Доказательство. Первое утверждение теоремы следует из того, что в силу локальной теоремы восстановления для случайного блуждания, порожденного моментами попадания цепи в состояние E_j , предел $\lim_{n \rightarrow \infty} h(n) = \lim_{n \rightarrow \infty} p_{ij}(n)$ всегда существует и равен $(E\tau^{(j)})^{-1}$.

Поэтому для доказательства достаточности во втором утверждении (необходимость следует из теоремы 8) надо в случае существования абсолютно сходящегося решения $\{\pi_j\}$ исключить существование нулевых состояний. Предположим противное, что $p_{ij}(n) \rightarrow 0$. Выберем j такое, что $\pi_j > 0$. Тогда

$$0 < \pi_j = \sum \pi_i p_{ij}(n) \rightarrow 0$$

при $n \rightarrow \infty$ по теореме о мажорируемой сходимости. Это противоречие заканчивает доказательство теоремы. ■

В заключение этого параграфа мы приведем два утверждения о предельном при $n \rightarrow \infty$ поведении числа $m_j(n)$ попаданий (возвращений) системы в фиксированное состояние E_j за время n . Эти утверждения являются непосредственным следствием теорем 7 и 8 гл. 9, поскольку величины $m_j(n)$ просто совпадают со значениями в точке n процесса восстановления $\eta(n)$, порожденного независимыми одинаково распределенными величинами τ_1, τ_2, \dots , где τ_1 есть время возвращения системы в состояние E_j .

Именно, справедлива

Теорема 10. Если $E\tau_1 = a < \infty$, то при $n \rightarrow \infty$

$$\frac{m_j(n)}{n} \rightarrow \frac{1}{p \cdot a}. \quad (14)$$

Если цепь эргодична, то отсюда следует, что

$$\frac{m_j(n)}{n} \xrightarrow{p} \pi_j.$$

Если $D\tau_1 = \sigma^2 < \infty$, то

$$P\left(\frac{m_j(n) - n/a}{\sqrt{n\sigma^2/a^3}} < x | X_0 = j\right) \rightarrow \Phi(x), \quad (15)$$

где $\Phi(x)$, как и прежде, есть функция нормального распределения с параметрами $(0, 1)$.

Предлагаем читателю в виде упражнения установить, что если

$$P\left(\bigcup_{n=1}^{\infty} \{X_n = j\} | X_0 = k\right) = 1, \quad (16)$$

то соотношение (14) (как и (15)) останется справедливым и в том случае, когда начальное значение $X_0 = j$ заменено на значение $X_0 = k$.

В доказательстве теоремы 6 отмечалось, что равенство (16) для возвратной неразложимой цепи всегда выполнено.

Резюмируя содержание этого параграфа, можно заметить, что изучение последовательности зависимых испытаний, связанных в цепь Маркова с дискретным множеством состояний, во многом может быть осуществлено с помощью результатов для последовательностей независимых случайных величин. Изучение других видов зависимых испытаний требует, как правило, иных подходов.

§ 5*. Поведение переходных вероятностей для разложимых цепей

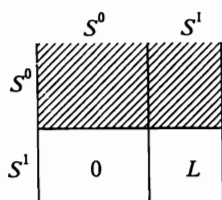


Рис. 12.

Рассмотрим теперь *конечную* цепь Маркова *общего типа*. Как было выяснено, множество ее состояний состоит из класса S^0 несущественных состояний и нескольких классов S^1, \dots, S^l существенных состояний. Для выяснения природы асимптотического поведения $p_{ij}(n)$ для таких цепей достаточно рассмотреть случай, когда существенные состояния образуют один класс без подклассов. Матрица вероятностей $p_{ij}(n)$ в этом случае имеет вид, изображенный на рис. 12. В силу эргодической теоремы для элементов подматрицы L существуют положительные пределы π_j . Таким образом, осталось выяснить, как ведут себя элементы заштрихованной части матрицы.

Теорема 11. Пусть $E_i \in S^0$. Тогда

$$\lim_{t \rightarrow \infty} p_{ij}(t) = \begin{cases} 0, & \text{если } E_j \in S^0, \\ \pi_j > 0, & \text{если } E_j \in S^1. \end{cases}$$

Доказательство. Пусть $E_j \in S^0$. Обозначим

$$A_j(t) = \max_{E_i \in S^0} p_{ij}(t).$$

Для любого существенного состояния E_r существует такое целое t_r , что $p_{ir}(t_r) > 0$. Так как переходные вероятности в S^1 с некоторыми пор все положительны, то

существует s такое, что $p_{ii}(s) > 0$ для $E_i \in S^0$ и для всех $E_i \in S^1$. Поэтому для достаточно больших t имеют место следующие соотношения:

$$p_{ij}(t) = \sum_{E_k \in S^0} p_{ik}(s) p_{kj}(t-s) \leq A_j(t-s) \sum_{E_k \in S^0} p_{ik}(s),$$

где $h(i) = \sum_{E_k \in S^0} p_{ik}(s) = 1 - \sum_{E_k \in S^1} p_{ik}(s) < 1$. Если $h = \max_{E_i \in S^0} h(i)$, то из полученного неравенства следует

$$A_j(t) \leq h A_j(t-s) \leq \dots \leq h^{\lfloor t/s \rfloor}.$$

Таким образом, $\lim_{t \rightarrow \infty} p_{ij}(t) \leq \lim_{t \rightarrow \infty} A_j(t) = 0$.

Пусть теперь $E_i \in S^0$, а $E_j \in S^1$. Имеют место равенства

$$p_{ij}(t+s) = \sum_k p_{ik}(t) p_{kj}(s) = \sum_{E_k \in S^0} p_{ik}(t) p_{kj}(s) + \sum_{E_k \in S^1} p_{ik}(t) p_{kj}(s).$$

Устремив t и s к ∞ , мы видим, что первая сумма в последнем выражении есть $o(1)$. Во второй сумме

$$\sum_{E \in S^1} p_{ik}(t) = 1 + o(1); \quad p_{kj}(t) = \pi_j + o(1).$$

Таким образом,

$$p_{ij}(t+s) = \pi_j \sum_{E_k \in S} p_{ik}(t) + o(1) = \pi_j + o(1)$$

при $t, s \rightarrow \infty$.

С помощью доказанной теоремы нетрудно убедиться, что существования $\lim_{t \rightarrow \infty} p_{ij}(t) = \pi_j \geq 0$ необходимо и достаточно для того, чтобы цепь состояла из двух классов S^0 и S^1 и чтобы S^1 не содержал подклассов.

§ 6. Цепи Маркова с произвольным множеством состояний. Эргодичность цепей, имеющих положительный атом

Цепи Маркова $X = \{X_n\}$, которые рассматривались в предыдущих разделах, принимали значения в счетном множестве $\{1, 2, \dots\}$ или $\{0, 1, \dots\}$; их называют *счетными* или *дискретными*. Мы рассмотрим теперь цепи Маркова со значениями в произвольном пространстве состояний \mathcal{X} , на котором определена некоторая σ -алгебра \mathfrak{B} подмножеств из \mathcal{X} . Пара $\langle \mathcal{X}, \mathfrak{B} \rangle$ образует *фазовое* измеримое пространство цепи $\{X_n\}$. Пусть далее $\langle \Omega, \mathfrak{F}, P \rangle$ — исходное вероятностное пространство. Измеримое отображение Y пространства $\langle \Omega, \mathfrak{F} \rangle$ в $\langle \mathcal{X}, \mathfrak{B} \rangle$ мы будем называть *\mathcal{X} -значным случайным элементом*. Если $\mathcal{X} = \mathbb{R}$, \mathfrak{B} есть σ -алгебра борелевских множеств, то Y будет обычной случайной величиной. Отображение Y может быть и тождественным, тогда пространство $\langle \Omega, \mathfrak{F} \rangle = \langle \mathcal{X}, \mathfrak{B} \rangle$ называют также *выборочным*.

Рассмотрим последовательность $\{X_n\}$ \mathcal{X} -значных случайных элементов и обозначим $\mathfrak{F}_{k,m}$, $m \geq k$, σ -алгебру, порожденную элементами X_k, \dots, X_m (т. е. событиями $\{X_k \in B_k\}, \dots, \{X_m \in B_m\}$, $B_i \in \mathfrak{B}$, $i = k, \dots, m$). Очевидно, что $\mathfrak{F}_n = \mathfrak{F}_{0,n}$ образуют неубывающую последовательность $\mathfrak{F}_0 \subset \mathfrak{F}_1 \subset \dots \subset \mathfrak{F}_n \subset \dots$. Условное математическое ожидание $E(\xi | \mathfrak{F}_{k,m})$ мы иногда будем обозначать также $E(\xi | X_k, \dots, X_m)$.

Определение 7. \mathcal{X} -значная цепь Маркова есть такая последовательность \mathcal{X} -значных случайных элементов X_n , что для любого $B \in \mathfrak{B}$

$$P(X_{n+1} \in B | \mathfrak{F}_n) = P(X_{n+1} \in B | X_n) \quad \text{п. н.} \quad (17)$$

В дальнейшем слова «почти наверное» мы, как правило, будем опускать.

Соотношение (17) в силу свойств у. м. о., очевидно, эквивалентно тому, что для любой измеримой функции $f: \mathcal{X} \rightarrow \mathbb{R}$ выполняется

$$E(f(X_{n+1}) | \mathfrak{F}_n) = E(f(X_{n+1}) | X_n). \quad (18)$$

Определение 7 эквивалентно следующему определению.

Определение 8. Последовательность $X = \{X_n\}$ образует цепь Маркова, если для любых $A \in \mathfrak{F}_{n+1, \infty}$

$$P(A | \mathfrak{F}_n) = P(A | X_n) \quad (19)$$

или, что то же, для любой $\mathfrak{F}_{n+1, \infty}$ -измеримой функции $f(\omega)$

$$E(f(\omega) | \mathfrak{F}_n) = E(f(\omega) | X_n). \quad (20)$$

Доказательство эквивалентности. Нам надо доказать, что из (18) следует (19). Пусть сначала $A = \{X_{n+1} \in B_1, X_{n+2} \in B_2\}$. Тогда в силу (18)

$$\begin{aligned} P(A | \mathfrak{F}_n) &= E[I(X_{n+1} \in B_1)P(X_{n+2} \in B_2 | \mathfrak{F}_{n+1}) | \mathfrak{F}_n] = \\ &= E[I(X_{n+1} \in B_1)P(X_{n+2} \in B_2 | X_{n+1}) | \mathfrak{F}_n] = E(A | X_n). \end{aligned}$$

Отсюда следует равенство (19) для любого $A \in \mathcal{A}_{n+1, n+2}$, где $\mathcal{A}_{k, m}$ — алгебра, порожденная множествами $\{X_k \in B_{k, \dots}, X_m \in B_m\}$. Ясно, что $\mathcal{A}_{n+1, n+2}$ порождает $\mathfrak{F}_{n+1, n+2}$. Пусть теперь $A \in \mathfrak{F}_{n+1, n+2}$. Тогда по теореме об аппроксимации найдутся $A_k \in \mathcal{A}_{n+1, n+2}$ такие, что $d(A, A_k) \rightarrow 0$ (см. § 3.4). Отсюда следует $I(A_k) \xrightarrow{p} I(A)$, и по свойствам у. м. о. (см. п. 4 § 4.8)

$$P(A_k | \mathfrak{F}^*) \xrightarrow{p} P(A | \mathfrak{F}^*),$$

где $\mathfrak{F}^* \subset \mathfrak{F}$ — некоторая σ -алгебра. Положим $P_A = P_A(\omega) = P(A | X_n)$. Нам известно, что для $A_k \in \mathcal{A}_{n+1, n+2}$

$$E(P_{A_k}; B) = P(A_k B) \quad (21)$$

для любого $B \in \mathfrak{F}_n$ (это и означает, что $P_{A_k}(\omega) = P(A_k | \mathfrak{F}_n)$). Пользуясь снова свойствами у. м. о. (теорема о мажорируемой сходимости; п. 4 § 4.8) и переходя к пределу в (21), получим $E(P_A; B) = P(AB)$. Это доказывает (19) для $A \in \mathfrak{F}_{n+1, n+2}$.

Повторяя m раз приведенные выше рассуждения, мы докажем (19) для $A \in \mathfrak{F}_{n+1, m}$. Переход к $A \in \mathfrak{F}_{n+1, \infty}$ также происходит по схеме, изложенной выше. ■

Отметим, что (19) без труда распространяется также на события $A \in \mathfrak{F}_{n, \infty}$. Можно было бы все рассмотрения в доказательстве эквивалентности с самого начала проводить для $A \in \mathfrak{F}_{n, \infty}$ (сначала для $A \in \mathcal{A}_{n, n+2}$ и т. д.).

Приведем еще одно эквивалентное определение марковости.

Определение 9. Последовательность $\{X_n\}$ образует цепь Маркова, если для любых $A \in \mathfrak{F}_n$, $B \in \mathfrak{F}_{n, \infty}$

$$P(AB | X_n) = P(A | X_n)P(B | X_n). \quad (22)$$

Это есть свойство условной независимости будущего от прошлого при фиксированном настоящем (условная независимость \mathfrak{F}_n и $\mathfrak{F}_{n, \infty}$ при фиксированной X_n).

Доказательство эквивалентности. Пусть выполнено (20). Тогда для $A \in \mathfrak{F}_n$, $B \in \mathfrak{F}_{n,\infty}$

$$\begin{aligned} P(AB|X_n) &= E[E(I_A I_B | \mathfrak{F}_n) | X_n] = E[I_A E(I_B | \mathfrak{F}_n) | X_n] = \\ &= E[I_A E(I_B | X_n) | X_n] = E(I_B | X_n) E(I_A | X_n). \end{aligned}$$

Обратно, пусть теперь выполнено (22). Тогда

$$\begin{aligned} P(AB) &= EP(AB|X_n) = EP(A|X_n)P(B|X_n) = \\ &= EE[I_A P(B|X_n) | X_n] = EI_A P(B|X_n). \end{aligned} \quad (23)$$

С другой стороны,

$$P(AB) = EI_A I_B = EI_A P(B|\mathfrak{F}_n). \quad (24)$$

Так как (23), (24) верны для любых $A \in \mathfrak{F}_n$, то это означает, что $P(B|X_n) = P(B|\mathfrak{F}_n)$. ■

Итак, пусть $\{X_n\}$ есть \mathcal{X} -значная цепь Маркова. Тогда в силу свойств у.м.о. $P(X_{n+1} \in B | X_n) = P_{(n)}(X_n, B)$, где функция $P_{(n)}(x, B)$ при каждом $B \in \mathfrak{B}$ измерима по x относительно σ -алгебры \mathfrak{B} . Везде в дальнейшем мы будем считать, что функции $P_{(n)}(x, B)$ являются условными распределениями (см. определение 4.12), т.е. при каждом $x \in \mathcal{X}$ $P_{(n)}(x, B)$ есть распределение вероятностей по B . Условные распределения $P_{(n)}(x, B)$ всегда существуют, если σ -алгебра \mathfrak{B} является счетно-порожденной, т.е. порождена счетным набором подмножеств из \mathcal{X} (см. [26]). Это условие всегда выполнено, если $\mathcal{X} = \mathbb{R}^k$, \mathfrak{B} есть σ -алгебра борелевских множеств. В нашем случае некоторая дополнительная проблема состоит еще и в том, что множество $\mathcal{N} \subset \mathcal{X}$ «нулевой вероятности», на котором мы можем произвольно менять $P_{(n)}(x, B)$ может зависеть от распределения X_n , т.к. «нулевая вероятность» имеется в виду относительно распределения X_n .

Определение 10. Цепь $X = \{X_n\}$ называется *однородной*, если существуют условные распределения $P_{(n)}(x, B) = P(x, B)$, не зависящие от n и начального распределения X_0 (или от распределений X_n). Функция $P(x, B)$ называется *переходной вероятностью* (или *переходной функцией*) однородной цепи Маркова. Ее можно для наглядности записывать в виде

$$P(x, B) = P(X_1 \in B | X_0 = x). \quad (25)$$

Если цепь Маркова счетна, $\mathcal{X} = \{1, 2, \dots\}$, то в обозначениях § 1 $P(i, \{j\}) = p_{ij} = p_{ij}(1)$.

Переходная вероятность и начальное распределение X_0 полностью определяют совместное распределение X_0, \dots, X_n для любого n . Действительно, в силу формулы полной вероятности и свойства марковости

$$P(X_0 \in B_0, \dots, X_n \in B_n) = \int_{y_0 \in B_0} \dots \int_{y_n \in B_n} P(X_0 \in dy_0) P(y_0, dy_1) \dots P(y_{n-1}, dy_n). \quad (26)$$

Цепь Маркова с начальным значением $X_0 = x$ мы будем обозначать $\{X_n(x)\}$.

В приложениях цепи Маркова, обычно задаются с помощью задания условных распределений $P(x, B)$ или в более «сильной» форме — с помощью явных формул, связывающих X_{n+1} со значениями X_n , и некоторых «управляющих» элементов (см. примеры 6–12), которые позволяют сразу же выписать переходные вероятности. В этих случаях, как уже отмечалось, совместное распределение (X_0, \dots, X_n) можно

определить с помощью начального распределения X_0 и переходной функции $P(x, B)$ по формуле (26). Легко видеть, что последовательность $\{X_n\}$ с так определенными совместными распределениями удовлетворяет всем определениям цепи Маркова и имеет переходную вероятность $P(x, B)$. В дальнейшем везде, где это потребуется, мы будем считать выполненным соотношение (26). Его можно считать еще одним определением цепи Маркова, но более «сильным», чем определения 8–10, т. к. оно задает (или использует) в явном виде переходную функцию $P(x, B)$.

Одним из основных объектов изучения будет асимптотическое поведение переходной вероятности за n шагов

$$P(x, n, B) = P(X_n(x) \in B) = P(X_n \in B | X_0 = x).$$

Для нее справедливо рекуррентное соотношение, вытекающее из формулы полной вероятности (или из (26))

$$P(X_{n+1} \in B) = EE(I(X_{n+1} \in B) | \mathfrak{F}_n) = \int P(X_n \in dy) P(y, B),$$

$$P(x, n+1, B) = \int P(x, n, dy) P(y, B). \quad (27)$$

Заметим теперь, что марковское свойство (19) однородных цепей можно записать также в виде

$$P(X_{n+k} \in B_k | \mathfrak{F}_n) = P(X_n, k, B_k)$$

или более общо:

$$P(X_{n+1} \in B_1, \dots, X_{n+k} \in B_k | \mathfrak{F}_n) = P(X_1^{new}(X_n) \in B_1, \dots, X_k^{new}(X_n) \in B_k), \quad (28)$$

где $\{X_k^{new}(x)\}$ — цепь Маркова, не зависящая от $\{X_n\}$, имеющая те же переходные функции, что и $\{X_n\}$, и начальное значение x . Свойство (28) допускает распространение на случайное время n .

Определение 11. Случайная величина $\nu \geq 0$ называется *марковской* или *моментом остановки* относительно $\{\mathfrak{F}_n\}$, если $\{\nu \leq n\} \in \mathfrak{F}_n$. Другими словами, тот факт, произошло событие $\{\nu \leq n\}$ или нет, полностью определяется значением траектории X_0, X_1, \dots, X_n .

Отметим, что часто под \mathfrak{F}_n в определении 11 понимают более широкие σ -алгебры, существенными остаются лишь соотношения $\{\nu \leq n\} \in \mathfrak{F}_n$ и измеримость X_0, \dots, X_n относительно \mathfrak{F}_n .

Обозначим \mathfrak{F}_ν σ -алгебру событий B таких, что $B \cap \{\nu = k\} \in \mathfrak{F}_k$. Другими словами, \mathfrak{F}_ν можно представлять себе как σ -алгебру, порожденную множествами $\{\nu = k\} B_k$, $B_k \in \mathfrak{F}_k$, т. е. траекторией $\{X_n\}$ до момента ν .

Лемма 2 (Строго марковское свойство). Для любого $k \geq 1$

$$P(X_{\nu+1} \in B_1, \dots, X_{\nu+k} \in B_k | \mathfrak{F}_\nu) = P(X_1^{new}(X_\nu) \in B_1, \dots, X_k^{new}(X_\nu) \in B_k),$$

где процесс $\{X_k^{new}\}$ определен в (28).

Таким образом, после случайного момента остановки ν траектория $X_{\nu+1}, X_{\nu+2}, \dots$ развивается по тем же законам, что и X_1, X_2, \dots , но с начальным условием X_ν . Это свойство называют *строго марковским*. Оно будет использоваться ниже для моментов $\nu = \tau_V$ первого попадания $\{X_n\}$ в некоторое множество $V \subset \mathcal{X}$. Мы уже использовали это свойство неявно в § 4, когда множество V совпадало с точкой и позволяло делить траекторию $\{X_n\}$ на независимые циклы.

Доказательство леммы 2. Для простоты рассмотрим одномерные распределения. Нам надо доказать, что

$$P(X_{\nu+1} \in B_1 | \mathfrak{F}_\nu) = P(X_\nu, B_1).$$

Для любого $A \in \mathfrak{F}_\nu$ имеем

$$\begin{aligned} E(P(X_\nu, B_1); A) &= \sum_n E(P(X_n, B_1); A\{\nu = n\}) = \\ &= \sum_n EE(I(A\{\nu = n\}\{X_{n+1} \in B_1\}) | \mathfrak{F}_n) \\ &= \sum_n P(A\{\nu = n\}\{X_{n+1} \in B_1\}) = P(A\{X_{\nu+1} \in B_1\}). \end{aligned}$$

Но это и означает, что $P(X_\nu, B_1)$ есть требуемое у. м. о. Рассмотрение многомерных распределений происходит совершенно аналогично, и мы предоставляем его читателю. ■

Перейдем теперь к изучению асимптотических свойств распределений $P(x, n, B)$ при $n \rightarrow \infty$.

Определение 12. Распределение $\pi(\cdot)$ на (x, B) называется *инвариантным*, если оно удовлетворяет уравнению

$$\pi(B) = \int \pi(dy)P(y, B).$$

Из (27) следует, что если $X_n \in \pi$, то $X_{n+1} \in \pi$. Распределение π называют также *стационарным*.

Для цепей Маркова в произвольном фазовом пространстве \mathcal{X} та простая и полная классификация состояний, которая была дана для счетных цепей в § 1, не представляется возможной, хотя некоторые понятия удастся перенести и на общий случай.

Такие естественные и важные для счетных цепей понятия, как неразложимость цепи, в общем случае приобретают иную форму.

Пример 8. Пусть $X_{n+1} = X_n + \xi_n \pmod{1}$ (X_{n+1} есть дробная часть $X_n + \xi_n$), ξ_n независимы, одинаково распределены и принимают с положительными вероятностями два значения: 0 и $\sqrt{2}$. Здесь в зависимости от начального значения x цепь разбивается на континуальное множество «подцепей» с множествами состояний вида $M_x = \{x + k\sqrt{2} \pmod{1}, k = 0, 1, 2, \dots\}$. Ясно, что если $x_1 - x_2$ не кратно $\sqrt{2} \pmod{1}$, то M_{x_1} и M_{x_2} не пересекаются, $P(X_n(x_1) \in M_{x_2}) = 0$, $P(X_n(x_2) \in M_{x_1}) = 0$ при всех n . Таким образом, цепь является явно разложимой. Тем не менее оказывается, что такая цепь эргодична в следующем смысле: для любого x $X_n(x) \Leftrightarrow U_{0,1}(P(x, n, [0, t]) \rightarrow t)$ (см., например, [6, 10]). О наиболее употребимых условиях неразложимости см. в § 7.

Определение 13. Цепь называется *периодической*, если существуют целое $d \geq 2$ и множество $\mathcal{X}_1 \subset \mathcal{X}$, такое, что для $x \in \mathcal{X}_1$ $P(x, n, \mathcal{X}_1) = P(X_n(x) \in \mathcal{X}_1) = 1$ для $n = kd$, $k = 1, 2, \dots$, и $P(x, n, \mathcal{X}_1) = 0$ для $n \neq kd$.

Периодичность означает, что все множество состояний \mathcal{X} разбивается на подклассы $\mathcal{X}_1, \dots, \mathcal{X}_d$, такие, что $P(X_1(x) \in \mathcal{X}_{k+1}) = 1$ при $x \in \mathcal{X}_k$, $k = 1, \dots, d$, $\mathcal{X}_{d+1} = \mathcal{X}_1$. В случае отсутствия такого свойства цепь будет называться *непериодической*.

Состояние $x_0 \in \mathcal{X}$ называется *атомом* цепи X , если для любого $x \in \mathcal{X}$

$$P\left(\bigcup_{n=1}^{\infty} \{X_n(x) = x_0\}\right) = 1.$$

Пример 9. Пусть $X_0 \geq 0$, и при $n \geq 0$

$$X_{n+1} = \begin{cases} (X_n + \xi_{n+1})^+, & \text{если } X_n > 0; \\ \eta_{n+1}, & \text{если } X_n = 0, \end{cases}$$

где $\xi_n, \eta_n \geq 0$, $n = 1, 2, \dots$ две последовательности независимых и одинаково распределенных в каждой последовательности случайных величин. Ясно, что $\{X_n\}$ образует цепь Маркова, и эта цепь при $E\xi_k < \infty$ имеет в силу усиленного закона больших чисел атом в точке $x_0 = 0$:

$$P\left(\bigcup_{n=1}^{\infty} \{X_n(x) = 0\}\right) = P\left(\inf_k S_k \leq -x\right) = 1,$$

$$\text{где } S_k = \sum_{j=1}^k \xi_j.$$

Цепи Маркова в произвольном фазовом пространстве \mathcal{X} образуют весьма сложный объект изучения. Однако, если цепь обладает атомом, то положение может сильно упроститься и эргодическая теорема об асимптотическом поведении $P(x, n, B)$ при $n \rightarrow \infty$ может быть доказана, не выходя за рамки подходов, изложенных в предыдущих параграфах.

Пусть x_0 — атом цепи $\{X_n\}$. Обозначим

$$\tau = \min\{k : X_k(x_0) = x_0\}.$$

Это есть собственная случайная величина ($P(\tau < \infty) = 1$).

Определение 14. Атом x_0 называется *положительным*, если $E\tau < \infty$.

В терминологии § 4 x_0 есть положительное, возвратное состояние.

Для характеристики сходимости распределений в произвольном пространстве нам понадобятся понятия расстояния по вариации и сходимости по вариации. Если P и Q — два распределения на $(\mathcal{X}, \mathcal{B})$, то расстоянием по вариации называется число

$$\|P - Q\| = 2 \sup_{B \in \mathcal{B}} |P(B) - Q(B)|.$$

Говорят, что последовательность распределений P_n на $(\mathcal{X}, \mathcal{B})$ сходится по вариации к P ($P_n \xrightarrow{\text{var}} P$), если $\|P_n - P\| \rightarrow 0$ при $n \rightarrow \infty$. Подробнее об этом см. § 6 Приложения 4.

Обозначим как и в § 4 $P_{x_0}(k, B)$ «табу-вероятность»

$$P_{x_0}(k, B) = P(X_k(x_0) \in B, \quad X_1(x_0) \neq x_0, \dots, X_{k-1}(x_0) \neq x_0)$$

перехода за k шагов из x_0 в B , не заходя при этом в «запретное» состояние x_0 .

Теорема 12. Если цепь $\{X_n\}$ имеет положительный атом и н. о. д. возможных значений τ равен 1, то цепь эргодична в смысле сходимости по вариации: существует единственное инвариантное распределение π такое, что для любого $x \in \mathcal{X}$

$$\|P(x, n, \cdot) - \pi(\cdot)\| \rightarrow 0. \quad (29)$$

При этом для любого $B \in \mathfrak{B}$

$$\pi(B) = \frac{1}{E\tau} \sum_{k=1}^{\infty} P_{x_0}(k, B). \quad (30)$$

Если обозначить $X_n(\mu_0)$ цепь Маркова с начальным распределением μ_0 ($X_0 \in \mu_0$), $P(\mu_0, n, B) = P(X_n(\mu_0) \in B) = \int \mu_0(dx) P(x, n, B)$, то наряду с (29) будет справедливо также

$$\|P(\mu_0, n, \cdot) - \pi(\cdot)\| \rightarrow 0 \quad (31)$$

для любого начального распределения μ_0 .

Условие существования положительного атома есть аналог условий (I), (II) в теореме 8. Условие, касающееся н.о.д. возможных значений τ , есть условие неперIODичности.

Доказательство теоремы 12 по существу повторяет доказательство эргодической теоремы 8. Пусть сначала $X_0 = x_0$. Как и в теореме 8 (обозначения этой теоремы сохраняются) находим

$$\begin{aligned} P(x_0, n, B) &= \\ &= \sum_{k=1}^n P(\gamma(n) = k) P(X_n \in B | X_{n-k} = x_0, X_{n-k+1} \neq x_0, \dots, X_{n-1} \neq x_0) = \\ &= \sum_{k=1}^n \frac{P(\gamma(n) = k)}{P(\tau \geq k)} P(\tau \geq k) P(X_k \in B | X_0 = x_0, X_1 \neq x_0, \dots, X_{k-1} \neq x_0) = \\ &= \sum_{k=1}^n \frac{P(\gamma(n) = k)}{P(\tau \geq k)} P_{x_0}(k, B). \end{aligned}$$

Для меры π , определенной в (30), будем иметь

$$P(x_0, n, B) - \pi(B) = \sum_{k=1}^n \left(\frac{P(\gamma(n) = k)}{P(\tau \geq k)} - \frac{1}{E\tau} \right) P_{x_0}(k, B) - \frac{1}{E\tau} \sum_{k>n} P_{x_0}(k, B).$$

Так как $P(\gamma(n) = k) \leq P(\tau \geq k)$ и $P_{x_0}(k, B) \leq P(\tau \geq k)$ (см. доказательство теоремы 8), то при любом N

$$\sup_B |P(x_0, n, B) - \pi(B)| \leq \sum_{k=1}^N \left(\frac{P(\gamma(n) = k)}{P(\tau \geq k)} - \frac{1}{E\tau} \right) + 2 \sum_{k>N} P(\tau \geq k). \quad (32)$$

Далее, так как $P(\gamma(n) = k) \rightarrow (P(\tau \geq k))/(E\tau)$, $\sum P(\tau \geq k) = E\tau < \infty$, то правая часть в (32) выбором N и затем n может быть сделана сколь угодно малой. Стало быть,

$$\lim_{n \rightarrow \infty} \sup_B |P(x_0, n, B) - \pi(B)| = 0.$$

Рассмотрим теперь произвольное начальное состояние $x \in \mathcal{X}$, $x \neq x_0$. Так как x_0 есть атом, то для вероятностей первого попадания в x_0 на k шаге $F(x, k, x_0) = P(X_k(x) = x_0, X_1 \neq x_0, \dots, X_{k-1} \neq x_0)$ имеем

$$\sum_k F(x, k, x_0) = 1,$$

$$P(x, n, B) = \sum_{k=1}^n F(x, k, x_0) P(x_0, n-k, B),$$

$$\|P(x, n, \cdot) - \pi(\cdot)\| \leq \sum_{k \leq n/2} F(x, k, x_0) \|P(x_0, n-k, \cdot) - \pi(\cdot)\| + 2 \sum_{k > n/2} F(x, k, x_0) \rightarrow 0$$

при $n \rightarrow \infty$.

Соотношение (31) следует из того, что

$$\|P(\mu_0, n, \cdot) - \pi(\cdot)\| \leq \int \mu_0(dx) \|P(x, n, \cdot) - \pi(\cdot)\| \rightarrow 0$$

по теореме о мажорируемой сходимости.

Далее, из сходимости $P(x, n, \cdot)$ по вариации следует, что

$$\int P(x, n, dy) P(y, B) \rightarrow \int \pi(dy) P(y, B).$$

Так как левая часть этого соотношения, равная в силу (27) $P(x, n+1, B)$, сходится к $\pi(B)$, то выполнено (28) и, стало быть, π есть инвариантная мера.

Допустим теперь, что π_1 — другое инвариантное распределение. Тогда $\pi_1(\cdot) = P(\pi_1, n, \cdot) \xrightarrow{var} \pi(\cdot)$, $\pi_1 = \pi$. ■

Вернемся к примеру 9. Покажем, что условия теоремы 12 выполнены, если $E\xi_k < 0$, $E\eta_k < \infty$. Действительно, обозначим $\theta(x) = \min\{k \geq 1 : S_k = \sum_{j=1}^k \xi_j \leq -x\}$.

Согласно теореме восстановления § 9.4

$$H(x) = E\theta(x) \sim \frac{x}{|E\xi_1|}$$

при $E\xi_1 < 0$, $x \rightarrow \infty$ и, стало быть, найдутся постоянные c_1 , c_2 , такие, что $H(x) < c_1 + c_2 x$ при всех $x \geq 0$. Поэтому для атома $x_0 = 0$ получим

$$E\tau = \int_0^\infty P(\eta_1 \in dx) H(x) \leq c_1 + c_2 \int_0^\infty x P(\eta_1 \in dx) = c_1 + c_2 E\eta_1 < \infty.$$

Отметим также, что в этом примере

$$P_{x_0}(k, B) = P\left(S_k \in B, \min_{1 \leq j < k} S_j > 0\right)$$

определяется совместным распределением S_k и $\underline{S}_k = \min_{j < k} S_j$, которое можно эффективно изучать с помощью факторизационных тождеств.

§ 7*. Эргодичность харрисовых цепей Маркова

Исследованию эргодичности цепей Маркова в произвольных пространствах состояний посвящено большое число исследований. Важнейшие из них связаны с именами Деблина, Дуба, Харриса, Орея. До недавнего времени эти исследования по праву относились к весьма не простым. Однако после появления конструкции искусственного положительного атома, предложенной Атреа, Неем и Нуммелином

(см., например, [6, 18, 27]) оказалось возможным доказывать теоремы эргодичности, избегая сложных построений и редуцировать задачу к случаю, рассмотренному в предыдущем параграфе.

Важную роль в дальнейшем будет играть понятие «харрисовости» цепи. Для некоторого фиксированного множества $V \in \mathfrak{B}$ определим случайную величину

$$\tau_V(x) = \min\{k \geq 1 : X_k(x) \in V\},$$

являющуюся временем первого попадания в V ($\tau_V(x) = \infty$, если все $X_k(x) \notin V$).

Определение 15. Цепь Маркова $\{X_n\}$ в $(\mathcal{X}, \mathcal{B})$ называется *харрисовой* (или *неприводимой по Харрису*), если существуют множество $V \in \mathfrak{B}$, вероятностная мера μ на $(\mathcal{X}, \mathcal{B})$ и числа $n_0 \geq 1$, $p \in (0, 1)$ такие, что

$$(I_0) \quad P(\tau_V(x) < \infty) = 1 \quad \forall x \in \mathcal{X},$$

$$(II) \quad P(x, n_0, B) \geq p\mu(B) \quad \text{для всех } x \in V, B \in \mathfrak{B}.$$

Условие (I_0) выполняет роль условия неразложимости цепи: из любой точки $x \in \mathcal{X}$ траектория X_n рано или поздно попадает в V . Условие (II) гарантирует, что через n_0 шагов после попадания в V «блуждающая частица» будет иметь распределение, которое «минорируется» одним и тем же «распределением» $p\mu(\cdot)$. Это условие иногда называют условием «перемешивания», оно обеспечивает «частичную утерю памяти» о прошлом траектории. Этого не происходит для цепей в примере 8, для которых условие (II) не выполнено ни для каких V, μ, n_0 ($P(x, \cdot)$ есть набор взаимно сингулярных распределений, сингулярных относительно меры Лебега).

Если цепь имеет атом, то условия $(I_0), (II)$ всегда выполнены при $V = \{x_0\}$, $n_0 = 1$, $p = 1$, $\mu(\cdot) = P(x_0, \cdot)$, так что такая цепь является харрисовой.

В качестве множества V обычно рассматривают «компактные» множества (если $\mathcal{X} = \mathbb{R}^k$, то это будут ограниченные множества), иначе обеспечить выполнение неравенств в (II) , как правило, не удастся. Если само пространство \mathcal{X} «компактно» (конечное множество или ограниченное множество в \mathbb{R}^k), то условие (II) может быть выполнено и для $V = \mathcal{X}$ (условие (I_0) при этом выполнено всегда). Например, если цепь $\{X_n\}$ конечна, неразложима и непериодична, то согласно теореме 9 найдется n_0 такое, что $P(i, n_0, j) \geq p > 0$ при всех i и j . Стало быть, условие (II) при $V = \mathcal{X}$ выполнено, если в качестве μ взять равномерное распределение на \mathcal{X} .

Условие (II) можно интерпретировать как наличие у всех распределений $P(x, n_0, \cdot)$ при $x \in V$ абсолютно непрерывной компоненты относительно меры μ :

$$\inf_{x \in V} \frac{P(x, n_0, dy)}{\mu(dy)} \geq p > 0.$$

Нам понадобятся еще условие «положительности» множества V (или «положительности цепи»):

$$(I) \quad \sup_{x \in V} E\tau_V(x) < \infty$$

и условие непериодичности, которое мы запишем в следующем виде. Пусть $X_k(\mu)$ — цепь Маркова с начальным значением $X_0 \in \mu$, где μ — из условия (II) .

Положим

$$\tau_V(\mu) = \min\{k \geq 0 : X_k(\mu) \in V\}.$$

Очевидно, что $\tau_V(\mu)$ в силу (I_0) есть собственная случайная величина. Обозначим n_1, n_2, \dots возможные значения $\tau_V(\mu)$, т.е. значения, при которых $P(\tau_V(\mu) = n_k) > 0$, $k = 1, 2, \dots$. Условие непериодичности имеет вид:

(III) Существует k такое, что

$$\text{о. н. д. } (n_0 + n_1, n_0 + n_2, \dots, n_0 + n_k) = 1,$$

где n_0 из условия (II).

Условие (III) всегда выполнено, если выполнено (II) при $n_0 = 1$, $\mu(V) > 0$ (тогда $n_1 = 0$, $n_0 + n_1 = 1$).

Проверка условия (I), как правило, связана (см. примеры ниже) с получением оценок для $E_{TV}(x)$ при $x \notin V$, из которых автоматически будет следовать (I₀).

Теорема 13. Пусть выполнены условия (I₀), (I), (II), (III), т.е. цепь Маркова является неперiodической положительной харрисовой цепью. Тогда существует единственное инвариантное распределение π такое, что для любого начального распределения μ_0

$$\|P(\mu_0, n, \cdot) - \pi(\cdot)\| \rightarrow 0. \quad (33)$$

Основной элемент доказательства состоит в использовании упомянутой выше конструкции «искусственного атома» и в редукции задачи к теореме 12. Это позволяет получить в ходе доказательства представление для инвариантной меры π , аналогичное (30) (см. (37)).

Замечательный факт состоит в том, что условия теоремы 13 оказываются необходимыми для сходимости (33) (подробнее об этом см. [6]).

Доказательство теоремы 13. Предположим для простоты, что $n_0 = 1$. Построим сначала наряду с X «расширенную» цепь Маркова $X^* = \{X_n^*\} = \{\tilde{X}_n, \omega(n)\}$, где $\omega(n)$ есть последовательность независимых одинаково распределенных случайных величин с $P(\omega(n) = 1) = p$ и $P(\omega(n) = 0) = 1 - p$. Совместное распределение $(\tilde{X}(n), \omega(n))$ в фазовом пространстве $\mathcal{X}^* = \mathcal{X} \times \{0, 1\} = \{x^* = (x, \delta) : x \in \mathcal{X}; \delta = 0, 1\}$ и переходная функция P^* цепи X^* определяются следующим образом (обозначение $X_n^*(x^*)$ имеет тот же смысл, что и $X_n(x)$):

$$P(X_1^*(x^*) \in (B, \delta)) \equiv P^*(x^*, (B, \delta)) = P(x, B)P(\omega(1) = \delta) \quad \text{при } x \notin V$$

(при $\tilde{X}_n \notin V$ компоненты X_{n+1}^* «разыгрываются» независимо с соответствующими маргинальными распределениями). Если же $x \in V$, то распределение $X^*(x^*, 1)$ определяется равенствами

$$P(X_1^*((x, 1) \in (B, \delta)) = P^*((x, 1), (B, \delta)) = \varphi(B)P(\omega(1) = \delta),$$

$$P(X_1^*((x, 0) \in (B, \delta)) = P^*((x, 0), (B, \delta)) = Q(x, B)P(\omega(1) = \delta),$$

где

$$Q(x, B) = \frac{P(x, B) - p\varphi(B)}{1 - p},$$

так что для любого $B \in \mathfrak{B}$

$$p\varphi(B) + (1 - p)Q(x, B) = P(x, B). \quad (34)$$

Таким образом, $P(\omega(n+1) = 1 | X_n^*) = p$ при любых значениях X_n^* . Однако при «выборе» \tilde{X}_{n+1} происходит частичная (лишь при $\tilde{X}_n \in V$) рандомизация (или расщепление): при $\tilde{X}_n \in V$ полагаем $P(\tilde{X}_{n+1} \in B | X_n^*)$ равной величине $\varphi(B)$ (не зависящей от $\tilde{X}_n \in V!$), если $\omega_n = 1$. Если же $\omega(n) = 0$, то названную вероятность полагаем равной $Q(\tilde{X}_n, B)$. Очевидно, что в силу условия (II) (при $n_0 = 1$) $\varphi(B)$ и $Q(x, B)$ являются распределениями, а в силу равенства (34)

первая координата $\tilde{X}(n)$ определенного выше процесса X_n^* обладает свойством $P(\tilde{X}_{n+1} \in B | \tilde{X}_n) = P(\tilde{X}_n, B)$, и, значит, распределения последовательностей X и \tilde{X} совпадают.

Как уже отмечалось, определенный выше процесс $X^*(n)$ обладает следующим свойством: распределение $P(X_{n+1}^* \in (B, \delta) | X_n^*)$ на множестве $X_n^* \in V^* = (V, 1)$ не зависит от $X^*(n)$ и является известным заранее распределением $\varphi(B)P(\omega(1) = \delta)$. Это и означает, что попадания цепи в состояние V^* делят траекторию X^* на независимые циклы, так же, как это происходит при наличии положительного атома.

Выше было описано, как распределение X^* строится по распределению X . Приведем также очевидные соотношения, которые восстанавливают распределение X по распределению X^* :

$$P(X_n(x) \in B) = p P(X_n^*(x, 1) \in B^*) + (1 - p) P(X_n^*(x, 0) \in B^*), \quad (35)$$

где $B^* = (B, 0) \cup (B, 1)$. Отметим также, что если $X_n = \tilde{X}_n$ рассматривается как компонента X_n^* , то она должна записываться как функция от начального значения $x^* \in \mathcal{X}^*$ в форме $X_n(x^*)$.

Положим

$$\tau^* = \min\{k \geq 1 : X_k^*(x^*) \in V^*\}, \quad x^* \in V^*.$$

Ясно, что τ^* не зависит от значения $x^* = (x, 1)$, так как $X_1(x^*)$ при любом $x \in V$ имеет распределение φ . Это свойство позволяет отождествить множество V^* с одной точкой. Другими словами, надо рассмотреть еще одно фазовое пространство \mathcal{X}^{**} , которое получается из \mathcal{X}^* заменой множества $V^* = (V, 1)$ на точку, которую мы обозначим x_0 . В этом новом фазовом пространстве построим цепь X^{**} , эквивалентную X^* , с помощью очевидных соотношений для переходной вероятности P^{**} :

$$P^{**}(x^*, (B, \delta)) = P^*(x^*, (B, \delta)) \quad \text{при} \quad x^* \neq (V, 1) = V^*, \quad (B, \delta) \neq V^*,$$

$$P^{**}(x_0, (B, \delta)) = P(\omega(1) = \delta)\mu(B), \quad P^{**}(x^*, x_0) = P^*(x^*, V^*).$$

Итак, мы построили цепь X^{**} с переходной функцией P^{**} , и эта цепь имеет атом x_0 . Очевидно, что $\tau^* = \min\{k \geq 1 : X_k^{**}(x_0) = x_0\}$. Докажем теперь, что этот атом положителен. Обозначим

$$E = \sup_x E\tau_V(x).$$

Лемма 3.

$$E\tau^* \leq \frac{2E}{p}.$$

Доказательство. Рассмотрим эволюцию первой координаты $X_k(x^*)$ процесса $X_k^*(x^*)$, $x^* \in V^*$. Разобьем ось времени $k \geq 0$ на интервалы по попаданию $X_k(x^*)$ в множество V . Пусть $\tau_1 \geq 1$ — время первого такого попадания (напомним, что $X_1(x^*) = X_0(\mu)$ и имеет распределение μ , так что $\tau_1 = 1$, если $\mu(V) = 1$). До момента τ_1 (в случае $\tau_1 > 1$) переходы $X_k(x^*)$, $k \geq 2$, управлялись переходной функцией $P(y, B)$, $y \in V^c = \mathcal{X} \setminus V$. В момент τ_1 в соответствии с определением X^* производится независимое от предыстории испытание в схеме Бернулли с вероятностью «успеха» (т.е. события $\omega(\tau_1) = 1$), равной p . Если $\omega(\tau_1) = 1$, то $\tau^* = \tau_1$. Если же $\omega(\tau_1) = 0$, то переход от $X_{\tau_1}(x^*)$ к $X_{\tau_1+1}(x^*)$ происходит с помощью переходной функции $Q(y, B) = (P(y, B) - p\mu(B))/(1 - p)$, $y \in V$. Далее все происходит

аналогично предыдущему: пусть $\tau_1 + \tau_2$ есть момент второго попадания $X(x^*, k)$ в V (в случае $\omega(\tau_1) = 0$). В промежутке $[\tau_1 + 1, \tau_2]$ переходы $X(x^*, k)$ осуществляются согласно переходной функции $P(y, B)$, $y \in V^c$. В момент $\tau_1 + \tau_2$ производится новое независимое испытание в схеме Бернулли с результатом $\omega(\tau_1 + \tau_2)$. Если $\omega(\tau_1 + \tau_2) = 1$, то $\tau^* = \tau_1 + \tau_2$. Если $\omega(\tau_1 + \tau_2) = 0$, то переход от $X(x^*, \tau_1 + \tau_2)$ к $X(x^*, \tau_1 + \tau_2 + 1)$ происходит согласно $Q(y, B)$, и т. д.

Другими словами, эволюцию координаты $X_k(x^*)$ процесса $X_k^*(x^*)$ можно представить так. Пусть $\tilde{X} = \{\tilde{X}_k\}$, $k = 1, 2, \dots$, есть цепь Маркова с распределением μ в момент $k = 1$ и переходной вероятностью $Q(x, B)$ в моменты $k \geq 2$

$$Q(x, B) = \begin{cases} \frac{P(x, B) - p\mu(B)}{1 - p}, & x \in V, \\ P(x, B), & x \in V^c. \end{cases}$$

Определим T_i следующим образом:

$$T_0 = 0, \quad T_1 = \tau_1 = \min\{k \geq 1 : \tilde{X}_k \in V\}, \\ T_i = \tau_1 + \dots + \tau_i = \min\{k > T_{i-1} : \tilde{X}_k \in V\}, \quad i \geq 2.$$

Пусть, далее, ν — не зависящая от \tilde{X} случайная величина с геометрическим распределением

$$P(\nu = k) = (1 - p)^{k-1}p, \quad k \geq 1, \\ (\nu = \min\{k \geq 1 : \omega(T_k) = 1\}). \quad (36)$$

Тогда из сказанного выше следует, что распределение $X(x^*, 1), \dots, X(x^*, \tau^*)$ совпадает с распределением $\tilde{X}(1), \dots, \tilde{X}(T_\nu)$; в частности, $\tau^* = T_\nu$,

$$E\tau^* = \sum_{k=1}^{\infty} p(1 - p)^{k-1} E T_k.$$

Далее, так как $\mu(B) \leq P(x, B)/p$ при $x \in V$, то при любом $x \in V$

$$E\tau_1 = \mu(V) + \int_{V^c} \mu(du)(1 + E\tau_V(u)) \leq \\ \leq \frac{1}{p} \left[P(x, V) + \int_{V^c} P(x, du)(1 + E\tau_V(u)) \right] = \frac{E\tau_V(x)}{p} \leq \frac{E}{p}.$$

Чтобы оценить $E\tau_i$ при $i \geq 2$, заметим, что $Q(x, B) \leq (1 - p)^{-1}P(x, B)$ при $x \in V$. Поэтому, если σ -алгебру, порожденную $\{\tilde{X}_k, \omega(\tau_k)\}$ при $k \leq T_i$, обозначить $\mathfrak{F}_{(i)}$, то

$$E(\tau_i | \mathfrak{F}_{(i-1)}) \leq \sup_{x \in V} \left[Q(x, V) + \int_{V^c} Q(x, du)(1 + E\tau_V(u)) \right] \leq \\ \leq \frac{1}{1 - p} \sup_{x \in V} \left[P(x, V) + \int_{V^c} P(x, du)(1 + E\tau_V(u)) \right] = \\ = (1 - p)^{-1} \sup_{x \in V} E\tau_V(x) = E(1 - p)^{-1}.$$

Отсюда следует неравенство $ET_k \leq E(1/p + (k-1)/(1-p))$, из которого получаем, что

$$E\tau^* \leq E\left(\frac{1}{p} + p \sum_{k=1}^{\infty} (k-1)(1-p)^{k-2}\right) = 2 \frac{E}{p}.$$

Лемма доказана. ■

Вернемся к доказательству теоремы. Чтобы воспользоваться теоремой 12, нам надо доказать теперь, что $P(\tau^*(x^*) < \infty) = 1$ для любых $x^* \in \mathcal{X}^*$, где $\tau^*(x^*) = \min\{k \geq 1 : X_k^*(x^*) \in V^*\}$. Но цепь X попадает в V с вероятностью 1. Через ν возвращений в V (ν определена в (36)) процесс $X^* = (X(n), \omega(n))$ окажется в множестве V^* .

Условие непереродности при $n_0 = 1$ будет выполнено, если $\mu(V) > 0$. В этом случае мы получаем в силу теоремы 12, что существует единственная инвариантная мера π^* такая, что для любого $x^* \in \mathcal{X}^*$

$$\begin{aligned} \|P^*(x^*, n, \cdot) - \pi^*(\cdot)\| &\rightarrow 0, \\ \pi^*((B, \delta)) &= \frac{1}{E\tau^*} \sum_{k=1}^{\infty} P_{V^*}^*(k, (B, \delta)), \end{aligned} \quad (37)$$

$$P_{V^*}^*(k, (B, \delta)) = P(X_k^*(x^*) \in (B, \delta), X_1^*(x^*) \notin V^*, \dots, X_{k-1}^*(x^*) \notin V^*).$$

В последнем равенстве имеется в виду любая точка $x^* \in V^*$; вероятность от выбора $x^* \in V^*$ не зависит.

Отсюда и из «формулы обращения» (35) получаем утверждение (33) для начального значения x^* и представление для инвариантной меры π процесса X .

Доказательство сходимости $\|P(\mu_0, n, \cdot) - \pi(\cdot)\| \rightarrow 0$ и единственности инвариантной меры происходит точно так же, как в теореме 12 (они следуют также из соответствующих утверждений для X^*).

Проверка условий теоремы 12 в случае, когда $n_0 > 1$ или $\mu(V) = 0$, не вызывает никаких дополнительных затруднений, и мы предоставляем ее читателю. ■

Отметим, что аналогично тому, как это делалось в доказательстве теоремы 8, можно установить также единственность решения интегрального уравнения для инвариантной меры (см. определение 12) в более широком классе обобщенных конечных мер.

Основными и наиболее трудными для проверки условиями теоремы 13 несомненно являются условия (I), (II). Условие (I₀) обычно получается «автоматически» в ходе проверки условия (I), т. к. эта проверка требует оценивать $E\tau_V(x)$ при любых x . Проверка условия (III) непереродности, как правило, не вызывает затруднений. Если, например, возвращение в множество V возможно за m_1 и m_2 шагов, н. о. д. $(m_1, m_2) = 1$, то цепь непереродна.

Итак, мы остановимся теперь более подробно на основных условиях (I), (II). Условие (II) выражается непосредственно в терминах локальных характеристик цепи (переходных вероятностей за один или за фиксированное число шагов $n_0 > 1$) и в этом смысле его можно считать окончательным. Надо уметь только «правильно угадать» наиболее подходящие множества V и меру μ (если таковые существуют, конечно). Например, для многомерных цепей Маркова в $\mathcal{X} = \mathbb{R}^d$ условие (II) будет выполнено, если выполнено одно из следующих двух условий.

(II_a). Распределение $X_{n_0}(x)$ имеет при некотором n_0 абсолютно непрерывную относительно меры Лебега (или относительно сумм мер Лебега на \mathbb{R}^d

и его «координатных» подпространств) компоненту, равномерно положительную по $x \in V_N = \{y : |y| \leq N\}$ при некотором $N > 0$ на множестве V_N . В этом случае роль μ будет играть равномерное распределение на V_N .

(II). $\mathcal{X} = \mathbb{Z}^d$ есть целочисленная решетка в \mathbb{R}^d . В этом случае цепь становится счетной и все упрощается.

Мы уже отмечали, что в тех случаях, когда цепь имеет положительный атом, как это имело место в примере 9, никаких условий на структуру (гладкость) распределения $X_{n_0}(x)$ не требуется.

Иначе обстоит дело с условием «положительности» (I). Оно выражено в терминах весьма сложных характеристик ($E\tau_V(x)$) и требует дополнительных исследований и отыскания условий, также выраженных в терминах локальных характеристик, которые обеспечивали бы выполнение (I). Этому во многом и будет посвящено дальнейшее содержание этого параграфа.

Предварительно приведем одно «промежуточное» утверждение, которое будет полезно для дальнейшего. Мы уже пользовались такого типа утверждением в рамках примера 9.

Теорема 14. Пусть существует такая неотрицательная измеримая функция $g : \mathcal{X} \rightarrow \mathbb{R}$, что выполнены следующие условия (I^g):

$$\begin{aligned} (I^g)_1 \quad & E\tau_V(x) \leq c_1 + c_2 g(x) \quad \text{при } x \in V^c = \mathcal{X} \setminus V, \quad c_1, c_2 = \text{const.} \\ (I^g)_2 \quad & \sup_{x \in V} E g(X_1(x)) < \infty. \end{aligned}$$

Тогда выполнено условие (I).

Функцию в теореме 14 часто называют «пробной» функцией или функцией Ляпунова. Положим для краткости $\tau_V(x) = \tau(x)$.

Доказательство. Пусть выполнено (I^g). Тогда для $x \in V$

$$\begin{aligned} E\tau(x) &\leq 1 + E[\tau(X_1(x)); X_1(x) \in V^c] \leq 1 + E(E[\tau(X_1(x)) | X_1(x)]; X_1(x) \in V^c) \leq \\ &\leq 1 + E(c_1 + c_2 g(X_1(x)); X_1(x) \in V^c) \leq 1 + c_1 + c_2 \sup_{x \in V} E g(X_1(x)) < \infty. \end{aligned}$$

Отметим, что неравенство (I^g)₂ относится к «локальным» характеристикам системы (как и условие (II)) и в этом смысле его также можно считать «окончательным» с точностью до проблемы выбора функции g .

Рассмотрим теперь условия, обеспечивающие выполнение (I^g)₁. Процессы $\{X_n\} = \{X_n(x)\}$, $X_0(x) = x$, которые будут рассматриваться в последующих утверждениях (например, в теореме 15) не обязательно должны быть марковскими. Мы будем использовать только те их свойства, которые будут оговорены в условиях утверждений.

Мы будем опять использовать неотрицательные пробные функции $g : \mathcal{X} \rightarrow \mathbb{R}$ и рассматривать множество V , «индуцированное» функцией g и некоторым множеством U , которое чаще всего будет ограниченным фиксированным отрезком прямой:

$$V = g^{-1}(U) = \{x \in \mathcal{X} : g(x) \in U\}.$$

Обозначение $\tau(x) = \tau_V(x)$ сохраняет свой смысл

$$\tau = \tau(x) = \min\{k \geq 1 : g(X_k(x)) \in U\} = \min\{k \geq 1 : X_k(x) \in V\}.$$

Теорема 15. Если при некотором $\varepsilon > 0$ и $n = 0, 1, 2, \dots$

$$\mathbb{E}(g(X_{n+1}) - g(X_n) | \tau > n) \leq -\varepsilon, \quad (38)$$

то

$$\mathbb{E}\tau(x) \leq \frac{g(x)}{\varepsilon}, \quad x \in V^c,$$

и, стало быть, выполнено $(I^g)_1$.

Доказательство. Обозначим $\tau_{(N)} = \min(\tau, N)$, $\Delta(n) = g(X_{n+1}) - g(X_n)$. Имеем

$$\begin{aligned} -g(x) &= -\mathbb{E}g(X_0) \leq \mathbb{E}(g(X_{\tau_{(N)}}) - g(X_0)) = \mathbb{E} \sum_{n=0}^{\tau_{(N)}} \Delta(n) = \\ &= \sum_{n=0}^N \mathbb{E} \Delta(n) I(\tau > n) = \sum_{n=0}^N \mathbb{P}(\tau > n) \mathbb{E}(\Delta(n) | \tau > n) \leq -\varepsilon \sum_{n=0}^N \mathbb{P}(\tau > n). \end{aligned}$$

Отсюда следует, что при любом N

$$\sum_{n=0}^N \mathbb{P}(\tau > n) \leq \frac{g(x)}{\varepsilon}.$$

Следовательно, это неравенство справедливо и при $N = \infty$, $\mathbb{E}\tau \leq g(x)/\varepsilon$. ■

Следствие 2. Если $\{X_n\}$ — цепь Маркова и при $x \in V^c$

$$\mathbb{E}g(X_1(x)) - g(x) \leq -\varepsilon, \quad (39)$$

то $\mathbb{E}\tau(x) \leq g(x)/\varepsilon$.

Доказательство очевидным образом следует из того, что в силу (39) и однородности цепи на множестве $\{X_n \in V^c\}$ выполнено $\mathbb{E}(g(X_{n+1}) - g(X_n) | X_n) \leq -\varepsilon$ и того, что $\{\tau > n\} \subset \{X_n \in V^c\}$ и, стало быть,

$$\mathbb{E}(g(X_{n+1}) - g(X_n); \tau > n) = \mathbb{E}[\mathbb{E}(g(X_{n+1}) - g(X_n) | X_n); \tau > n] \leq -\varepsilon \mathbb{P}(\tau > n). \quad \blacksquare$$

Теорема 15 представляет собой некоторую модификацию критерия положительной возвратности, известного как критерий Фостера—Мустафы—Твиди (см., например, [6, 27]).

Рассмотрим некоторые применения полученных результатов. Пусть X — цепь Маркова на вещественной полуоси $\mathbb{R}_+ = [0, \infty)$. Обозначим для краткости $X_1(x) - x = \xi(x)$. Это есть приращение за 1 шаг цепи, стартовой в точке x . $\xi(x)$ можно определить также как случайную величину с распределением $\mathbb{P}(\xi(x) \in B) = P(x, B - x)$ ($B - x = \{y \in \mathcal{X} : y + x \in B\}$).

Следствие 3. Если при некоторых $N \geq 0$, $\varepsilon > 0$

$$\sup_{x \leq N} \mathbb{E}\xi(x) < \infty, \quad \sup_{x > N} \mathbb{E}\xi(x) \leq -\varepsilon, \quad (40)$$

то выполнены условия (I_0) , (I) при $V = [0, N]$.

Доказательство. Воспользуемся теоремами 14, 15 и следствием 2 при $g(x) \equiv x$, $V = [0, N]$. Условия $(I^g)_2$ и (39) выполнены очевидным образом. ■

Таким образом, наличие «отрицательного сноса» цепи в области $x > N$ гарантирует положительность цепи. Однако выполнение условия (I) можно обеспечить и в том случае, когда «снос» $\mathbb{E}\xi(x)$ неограниченно убывает при $x \rightarrow \infty$.

Следствие 4. Пусть $\sup E\xi^2(x) < \infty$ и

$$E\xi^2(x) \leq \sigma^2, \quad E\xi(x) \leq -\frac{c}{x} \quad \text{при } x > N.$$

Тогда, если $2c > \sigma^2$, то выполнено (II) при $V = [0, N]$.

Доказательство. Воспользуемся вновь теоремой 14 и следствием 2, но при $g(x) = x^2$. Получим для $x > N$

$$Eg(X_1(x)) - g(x) = E(2x\xi(x) + \xi^2(x)) \leq -2c + \sigma^2 < 0.$$

Прежде чем переходить к рассмотрению примеров, связанных с эргодичностью, отметим следующее обстоятельство. Чем «больше» множество V , тем легче проверять условие (I) и чем «меньше» это множество, тем легче проверять условие (II) (и труднее — (I)). В связи с этим возникает вопрос, когда можно рассматривать два множества «маленькое» W и «большое» $V: W \subset V$, такие, что выполнение (I) для V и (II) для W влечет за собой выполнение (I), (II) для W ?

Лемма 4. Пусть

$$E \equiv \sup_{x \in V} E\tau_V(x) < \infty \quad (I)_V$$

и существует t такое, что

$$\inf_{x \in V} P\left(\bigcup_{j=1}^m \{X_j(x) \in W\}\right) \geq q > 0.$$

Тогда выполнено $(I)_W$:

$$\sup_{x \in W} E\tau_W(x) \leq \sup_{x \in V} E\tau_W(x) \leq \frac{Em}{q}.$$

Таким образом, если выполнено условие (I) для множества V , условие (II) для множества W и условия леммы 4, то выполнены условия (I), (II) для множества W .

Для доказательства леммы 4 нам понадобится следующее утверждение, обобщающее (как неравенство) известное тождество Вальда.

Пусть дана последовательность неотрицательных случайных величин τ_1, τ_2, \dots , измеримых соответственно относительно σ -алгебр $\mathfrak{A}_1 \subset \mathfrak{A}_2 \subset \dots$, $T_n = \tau_1 + \dots + \tau_n$. Пусть, кроме того, задан момент остановки $\nu: \{\nu \leq n\} \in \mathfrak{A}_n$.

Лемма 5. Если $E(\tau_n | \mathfrak{A}_{n-1}) \leq a$, то

$$ET_\nu \leq aE\nu.$$

Доказательство. Мы можем считать, не ограничивая общности, что $E\nu < \infty$ (иначе неравенство тривиально). Доказательство по сути повторяет рассуждения в теореме 4.2. Имеем

$$ET_\nu = \sum_{k=1}^{\infty} E(T_k; \nu \geq k) = \sum_{k=1}^{\infty} E(\tau_k, \nu \geq k). \quad (41)$$

Изменение порядка суммирования здесь законно, т. к. слагаемые неотрицательны. Далее $\{\nu \leq k-1\} \in \mathfrak{A}_{k-1}$ и, следовательно, $\{\nu \geq k\} \in \mathfrak{A}_{k-1}$. Поэтому

$$E(\tau_k; \nu \geq k) = EI(\nu \geq k)E(\tau_k | \mathfrak{A}_{k-1}) \leq aP(\nu \geq k).$$

Сравнивая с (40), получим

$$ET_\nu \leq a \sum_{k=1}^{\infty} P(\nu \geq k) = aE\nu.$$

Доказательство леммы 4. Пусть цепь X стартует в точке $x \in V$. Рассмотрим последовательные моменты T_1, T_2, \dots попадания X в V , $T_0 = 0$. Положим $Y_0 = x$, $Y_k = X_{T_k}(x)$, $k = 1, 2, \dots$. Тогда в силу строго марковского свойства последовательность (Y_k, T_k) образует цепь Маркова. Положим $\mathfrak{A}_k = \sigma(T_1, \dots, T_k; Y_1, \dots, Y_k)$, $\tau_k = T_k - T_{k-1}$, $k = 1, 2, \dots$. Тогда $\nu = \min\{k : Y_k \in W\}$ образует момент остановки относительно $\{\mathfrak{A}_k\}$. Очевидно, что $E(\tau_k | \mathfrak{A}_{k-1}) \leq E$. Оценим $E\nu$. Для этого оценим

$$p_k \equiv P(\nu \geq km) \leq P\left(\bigcap_{j=1}^{T_{km}} \{X_j \notin W\}\right) = \\ = EI\left(\bigcap_{j=1}^{T_{(k-1)m}} \{X_j \notin W\}\right) E\left(I\left(\bigcap_{j=T_{(k-1)m}+1}^{T_{km}} \{X_j \notin W\}\right) \middle| \mathfrak{A}_{(k-1)m}\right).$$

Так как $\tau_j \geq 1$, то последний множитель по условию леммы и по свойству строгой марковости не превосходит

$$P\left(\bigcap_{j=1}^m \{X_j^{new}(X_{T_{(k-1)m}}) \notin W\}\right) \leq (1-q).$$

Таким образом, $p_k \leq (1-q)p_{k-1} \leq (1-q)^k$, $E\nu \leq m/q$, и в силу леммы 5 $ET_\nu \leq E_m/q$. Остается заметить, что $\tau_W(x) = T_\nu$. ■

Пример 10. *Случайное блуждание с отражением.* Пусть ξ_1, ξ_2, \dots — независимые одинаково распределенные случайные величины,

$$X_{n+1} = |X_n + \xi_{n+1}|, \quad n = 0, 1, \dots \quad (42)$$

Если ξ_k , а вместе с ними и X_k не решетчатые, то цепь X не имеет, вообще говоря, атома. Если, например, ξ_k имеют плотность распределения $f(t)$ относительно меры Лебега, то $P(X_k(x) = y) = 0$ при любых $x, y, k \geq 1$. Мы будем предполагать, что выполнено более широкое условие (А).

(А). *В разложении*

$$P(\xi_k < t) = p_a F_a(t) + p_c F_c(t)$$

на абсолютно непрерывную (F_a) и сингулярную (F_c) компоненты выполняется неравенство $p_a > 0$.

Следствие 6. Если выполнено условие (А), $a = E\xi_k < 0$, $E|\xi_k| < \infty$, то цепь Маркова, определенная в (42), удовлетворяет условиям теоремы 14 и, стало быть, эргодична в смысле сходимости по вариации.

Доказательство. Убедимся сначала, что цепь удовлетворяет условиям следствия 3. Так как в нашем случае $|X_1(x) - x| \leq |\xi_1|$, то первое из условий (40) выполнено. Далее

$$E\xi(x) = E|x + \xi_1| - x = E(\xi_1; \xi_1 \geq -x) - E(2x + \xi_1; \xi_1 < -x) \rightarrow E\xi_1$$

при $x \rightarrow \infty$, так как

$$xP(\xi_1 < -x) \leq E(|\xi_1|, |\xi_1| > x) \rightarrow 0.$$

Поэтому найдется N такое, что $E\xi(x) \leq a/2 < 0$ при $x \geq N$. Выполнение условий (I₀), (I) при $V = [0, N]$ доказано.

Проверим теперь выполнение условия (II) для множества $W = [0, h]$, при некотором h . Пусть $f(t)$ есть плотность распределения F_a в условии (А). Тогда можно считать, что существуют $f_0 > 0$ и отрезок $[t_1, t_2]$, $t_2 > t_1$ такие, что

$f(t) > f_0$ при $t \in [t_1, t_2]$. Ясно, что плотность величины $x + \xi_1$ будет больше f_0 на $[x + t_1, x + t_2]$. Положим $h = (t_2 - t_1)/2$. Тогда при $0 \leq x \leq h$ будет выполняться $[t_2 - h, t_2] \subset [x + t_1, x + t_2]$.

Пусть для определенности $t_2 > 0$. Сказанное будет означать тогда, что плотность $x + \xi_1$ будет больше f_0 на $[(t_2 - h)^+, t_2]$ при всех $x \leq h$ и, стало быть,

$$\inf_{x \leq h} P(X_1(x) \in B) \geq p_1 \int_B f_0(t) dt,$$

где

$$f_0(t) = \begin{cases} f_0 & \text{при } t \in [(t_2 - h)^+, t_2], \\ 0 & \text{в противном случае.} \end{cases}$$

Это и означает выполнение условия (II) на множестве $W = [0, h]$. Случай $t_2 \leq 0$ рассматривается аналогично.

Нам остается воспользоваться леммой 4, из которой будет следовать выполнение условия (I) для множества W . Условие леммы 4 очевидным образом выполнено (при достаточно большом m распределение $X_m(x)$, $x \leq N$ будет иметь на W положительную абсолютно непрерывную компоненту). По этой же причине цепь X не может быть периодичной. Таким образом, все условия теоремы 14 выполнены. ■

Пример 11. Осциллирующее случайное блуждание. Пусть даны две независимые последовательности ξ_1, ξ_2, \dots и η_1, η_2, \dots независимых и одинаково распределенных в каждой последовательности случайных величин. Положим

$$X_{n+1} = \begin{cases} X_n + \xi_{n+1}, & \text{если } X_n \geq 0; \\ X_n + \eta_{n+1}, & \text{если } X_n < 0. \end{cases} \quad (43)$$

Такое случайное блуждание называют *осциллирующим*. Очевидно, оно образует цепь Маркова в $\mathcal{X} = (-\infty, \infty)$.

Следствие 7. Если хотя одно из распределений ξ_k или η_k удовлетворяет условию (A), $-\infty < E\xi_k < 0$, $\infty > E\eta_k > 0$, то цепь (43) будет удовлетворять условиям теоремы 14 и, стало быть, будет эргодической.

Доказательство совершенно аналогично доказательству следствия 6. Надо лишь для проверки условия (I) взять $g(x) = |x|$ и $V = [-N, N]$. После чего воспользоваться леммой 4 при $W = [0, h]$, если условию (A) удовлетворяют ξ_k (и $W = [-h, 0)$, если (A) удовлетворяют η_k). ■

Отметим, что условие (A) в примерах 10, 11 может быть ослаблено до наличия абсолютно непрерывной компоненты у сумм $\sum_{j=1}^m \xi_j$ (или $\sum \eta_j$) при некотором m .

С другой стороны, если распределения этих сумм при всех m сингулярны, то сходимость распределений $P(x, n, \cdot)$ по вариации не может иметь место. Если, например, в примере 10 $P(\xi_k = -\sqrt{2}) = P(\xi_k = 1) = 1/2$, то $E\xi_k < 0$ и условие (I) будет выполнено, а условие (II) нет. Невозможна и сходимость $P(x, n, \cdot)$ по вариации к предельному распределению π . Действительно, из уравнения для инвариантного распределения π легко следует, что это распределение с необходимостью непрерывно. С другой стороны, скажем, распределения $P(0, n, \cdot)$ сосредоточены на счетном множестве \mathcal{N} чисел $|-k\sqrt{2} + l|$; $k, l = 1, 2, \dots$. Поэтому $P(0, n, \mathcal{N}) = 1$ при всех n , $\pi(\mathcal{N}) = 0$. Поэтому может иметь лишь слабая сходимость распределений $P(x, n, \cdot)$ к $\pi(\cdot)$. И хотя такая сходимость сомнений не вызывает, достаточно простое доказательство этого факта нам неизвестно.

Пример 12 (Продолжение примеров 6, 8). Пусть $\mathcal{X} = [0, 1]$, ξ_1, ξ_2, \dots независимы и одинаково распределены, $X_{n+1} = X_n + \xi_{n+1} \pmod{1}$ или, что то же, $X_{n+1} = \{X_n + \xi_{n+1}\}$, где $\{x\}$ означает дробную часть x . Здесь условие (I) очевидным образом выполнено при $V = \mathcal{X} = [0, 1]$. Если ξ_k удовлетворяют условию (A), то, так же, как в примере 10, будет выполнено условие (II) для множества $W = [0, h]$ при некотором $h > 0$, что вместе с леммой 4, как и раньше, приводит к выполнению условий теоремы 14. Инвариантное распределение π в этом примере будет равномерным на $[0, 1]$. Для простоты будем считать, что распределение ξ_k имеет плотность $f(t)$, при этом не ограничивая общности можно считать $\xi_k \in [0, 1]$, ($f(t) = 0$ при $t \notin [0, 1]$). Тогда плотность $p(x) = 1$ инвариантной меры π удовлетворяет уравнению для инвариантной меры:

$$p(x) = 1 = \int_0^x dy f(x - y) + \int_x^1 dy f(x - y + 1) = \int_0^1 f(y) dy.$$

Так как стационарное распределение единственно, то $\pi = U_{0,1}$. При этом по теореме 5 Приложения 3 наряду со сходимостью $P(x, n, \cdot)$ к $U_{0,1}$ по вариации будет иметь место сходимость плотностей $P(x, n, dt)/dt$ к 1 по мере Лебега.

Тот факт, что инвариантное распределение является равномерным, остается верным и для произвольных нерешетчатых распределений ξ_k . Однако, как уже отмечалось в примере 8, в общем случае (без выполнения условия (A)) возможна лишь слабая сходимость распределений $P(x, n, \cdot)$ к равномерному (см. [6, 10]).

Информация и энтропия*

§ 1. Определения, свойства информации и энтропии

Предположим, что производится некоторый эксперимент. При этом событие A может произойти с вероятностью 0,99. Тогда естественно считать, что появление события A не несет в себе большой информации. Появление же события \bar{A} несет в себе значительную информацию, так как оно нами не ожидалось. Пусть, например, событие A означает, что самолет приземлился благополучно. Тогда сообщение о том, что произошло событие \bar{A} (обычно оно маловероятно), будет нести для встречающих (ожидающих исхода эксперимента) весьма существенную информацию.

Принята следующая количественная мера информации. Пусть A и B — события некоторого вероятностного пространства $\langle \Omega, \mathfrak{F}, P \rangle$.

Определение 1. Количеством информации $I(A|B)$, которое заключено в событии (сообщении) B относительно A , называется число

$$I(A|B) = \log \frac{P(A|B)}{P(A)}.$$

Появление события $B = A$ можно интерпретировать как сообщение о том, что наступило A .

Определение 2. Число $I(A|A) = I(A)$ называется количеством информации $I(A)$, заключающейся в сообщении A :

$$I(A|A) = I(A) = -\log P(A).$$

Из этого определения видно, что чем больше вероятность события A , тем меньше $I(A)$. Как правило, в определении информации используется логарифм при основании 2, так что единицу информации несет, например, сообщение о том, что в семье родился мальчик (или девочка) (предполагается, что эти события равновероятны, а $-\log_2 p = 1$ при $p = 1/2$). В дальнейшем в этом параграфе везде запись $\log x$ следует понимать как $\log_2 x$.

Если события A и B независимы, то $I(A|B) = 0$. Это означает, что событие B не несет в себе никакой информации относительно A и наоборот. Следует заметить, что всегда $I(A|B) = I(B|A)$.

Легко видеть, что если события A и B независимы, то

$$I(AB) = I(A) + I(B). \quad (1)$$

* Понятия количества информации и энтропии были введены К. Э. Шенноном в 1948 г. Для некоторых частных ситуаций понятие количества информации рассматривалось и в более ранних работах (например, Р. Хартли, 1928 г.). Изложение § 2 этой главы основывается во многом на работе А. Я. Хинчина [24].

Рассмотрим пример. Пусть случайным образом на одну из клеток шахматной доски ставится фигура. В этом случае информация о том, что фигура находится в клетке с номером k (событие A), будет равна $I(A) = \log 64 = 6$. Пусть B_1 — событие, заключающееся в том, что фигура находится в i строке, а B_2 — что фигура находится в j столбце. Сообщение A можно передать, передав сначала B_1 , затем B_2 . Имеем

$$I(B_1) = \log 8 = 3 = I(B_2).$$

Поэтому

$$I(B_1) + I(B_2) = 6 = I(A),$$

так что при передаче сообщения A «по частям» требуется сообщить то же количество информации, равное 6, что и при одноразовой передаче. Можно привести и другие примеры, обнаруживающие естественность введенных числовых характеристик.

Пусть G есть некоторый эксперимент с исходами E_1, \dots, E_N , которые осуществляются с вероятностями p_1, \dots, p_N .

Информация, полученная нами в результате испытания G , есть случайная величина $J_G = J_G(\omega)$, принимающая значение $-\log p_j$, $j = 1, \dots, N$, на множестве E_j .

Таким образом, если в вероятностном пространстве $(\Omega, \mathfrak{F}, P)$, соответствующем эксперименту G , Ω совпадает с множеством (E_1, \dots, E_N) , то $J_G(\omega) = I(\omega)$.

Определение 3. Математическое ожидание этой информации $EJ_G = -\sum p_j \log p_j$ называется *энтропией* эксперимента. Мы будем обозначать ее символами

$$H_p = H(G) = -\sum_{j=1}^N p_j \log p_j,$$

где $p = (p_1, \dots, p_N)$. При $p_j = 0$ произведение $p_j \log p_j$ мы полагаем по непрерывности равным 0.

Энтропия эксперимента есть в некотором смысле мера его неопределенности. Пусть, например, эксперимент имеет два исхода A и B с вероятностями p и $q = 1 - p$. Энтропия такого эксперимента равна

$$H_p = -p \log p - (1 - p) \log(1 - p) = f(p).$$

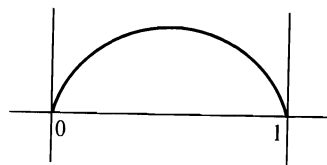


Рис. 13.

График этой функции представлен на рис. 13. Единственный максимум $f(p)$, равный $\log 2 = 1$, достигается в точке $p = 1/2$. Это схема с максимальной неопределенностью. Если p уменьшается, то неопределенность вместе с H_p тоже уменьшается, $H_p = 0$, если $p = (0, 1)$ или $(1, 0)$.

Эти свойства легко усмотреть и в общем случае.

Свойства энтропии.

1. $H(G) = 0$, если и только если существует такое j , $1 \leq j \leq N$, что

$$p_j = P(E_j) = 1.$$

2. $H(G)$ максимально, если $p_j = 1/N$ для всех j .

Доказательство. Так как вторая производная функции $\varphi(x) = x \log x$ на $[0, 1]$ положительна, то $\varphi(x)$ выпукла. Следовательно, для любых $q_i \geq 0$ таких, что

$\sum_{i=1}^N q_i = 1$, и любых $x_i \geq 0$ выполняется неравенство

$$\varphi \left(\sum_{i=1}^N q_i x_i \right) \leq \sum_{i=1}^N q_i \varphi(x_i).$$

Если взять $q_i = 1/N$, $x_i = p_i$, то получим

$$\left(\frac{1}{N} \sum p_i \right) \log \left(\frac{1}{N} \sum p_i \right) \leq \sum \frac{1}{N} p_i \log p_i.$$

Полагая $u = (\frac{1}{N}, \dots, \frac{1}{N})$, находим отсюда

$$-\log \frac{1}{N} = \log N = H u \geq - \sum p_i \log p_i = H p.$$

Заметим, что если энтропия $H(G)$ равна своему наибольшему значению $H(G) = \log N$, то с вероятностью 1 $J_G(\omega) = \log N$, т. е. информация J_G становится постоянной.

3. Пусть G_1 и G_2 — два независимых эксперимента. Запишем исходы и соответствующие вероятности исходов в этих экспериментах следующим образом:

$$G_1 = \begin{pmatrix} E_1, & \dots, & E_N \\ p_1, & \dots, & p_N \end{pmatrix}, \quad G_2 = \begin{pmatrix} A_1, & \dots, & A_M \\ q_1, & \dots, & q_M \end{pmatrix}.$$

Комбинируя исходы этих двух экспериментов, мы получим новый эксперимент:

$$G = G_1 \times G_2 = \begin{pmatrix} E_1 A_1, & E_1 A_2, & \dots, & E_N A_M \\ p_1 q_1, & p_1 q_2, & \dots, & p_N q_M \end{pmatrix}.$$

Информация J_G , полученная в результате этого эксперимента, есть случайная величина, принимающая значения $-\log p_i q_j$ с вероятностями $p_i q_j$, $i = 1, \dots, N$; $j = 1, \dots, M$. Но такое же распределение имеет, очевидно, сумма $J_{G_1} + J_{G_2}$ независимых величин, равных соответственно информации от экспериментов G_1 и G_2 . Таким образом, информация, полученная от последовательности независимых экспериментов, равна сумме информации от этих экспериментов. Так как при этом, очевидно,

$$E J_G = E J_{G_1} + E J_{G_2},$$

то мы получаем, что для независимых G_1 , G_2 энтропия $H(G)$ равна сумме

$$H(G) = H(G_1) + H(G_2)$$

энтропий экспериментов G_1 и G_2 .

4. Если эксперименты G_1 и G_2 зависимы, то эксперимент G будет выглядеть следующим образом:

$$G = \begin{pmatrix} E_1 A_1, & E_1 A_2, & \dots, & E_N A_M \\ q_{11}, & q_{12}, & \dots, & q_{NM} \end{pmatrix},$$

где $q_{ij} = p_i p_{ij}$, а p_{ij} есть условная вероятность события A_j при условии E_i , так что

$$\sum_{j=1}^M q_{ij} = p_i = P(E_i), \quad i = 1, \dots, N;$$

$$\sum_{i=1}^N q_{ij} = q_j = P(A_j), \quad j = 1, \dots, M.$$

В этом случае равенство $J_G = J_{G_1} + J_{G_2}$, вообще говоря, уже не будет иметь места. Введем случайную величину J_2^* , равную $-\log p_{ij}$ на множестве $E_i A_j$. Тогда, очевидно, $J_G = J_{G_1} + J_2^*$. Так как

$$P(A_j|E_i) = p_{ij},$$

то величину J_2^* при фиксированном i можно рассматривать как информацию от эксперимента G_2 при условии, что произошло событие E_i . Величину

$$E(J_2^*|E_i) = - \sum_{j=1}^M p_{ij} \log p_{ij}$$

мы будем называть *условной энтропией* $H(G_2|E_i)$ эксперимента G_2 при условии E_i , а величину

$$E J_2^* = - \sum_{i,j} q_{ij} \log p_{ij} = \sum_i p_i H(G_2|E_i)$$

условной энтропией $H(G_2|G_1)$ эксперимента G_2 относительно эксперимента G_1 . Во введенных обозначениях, очевидно,

$$H(G) = H(G_1) + H(G_2|G_1).$$

Мы докажем, что в этом равенстве всегда

$$H(G_2|G_1) \leq H(G_2),$$

т. е. для двух экспериментов G_1 и G_2 энтропия $H(G)$ всегда не превосходит сумму энтропий $H(G_1)$ и $H(G_2)$:

$$H(G) = H(G_1 \times G_2) \leq H(G_1) + H(G_2).$$

При этом равенство имеет место только в том случае, когда $q_{ij} = p_i q_j$, т. е. когда G_1 и G_2 независимы.

Для доказательства заметим сначала, что для двух любых распределений (u_1, \dots, u_n) и (v_1, \dots, v_n) справедливо неравенство

$$- \sum u_i \log u_i \leq - \sum u_i \log v_i, \quad (2)$$

причем равенство возможно лишь при $v_i = u_i$, $i = 1, \dots, n$. Это следует из выпуклости функции $\log x$, в силу которой при любых $a_i > 0$

$$\sum u_i \log a_i \leq \log \left(\sum u_i a_i \right)$$

(равенство возможно, если только $a_1 = a_2 = \dots = a_n$). Полагая $a_i = v_i/u_i$, мы получим (2).

Далее имеем

$$H(G_1) + H(G_2) = - \sum q_{ij} (\log p_i + \log q_j) = - \sum q_{ij} \log p_i q_j.$$

Так как $\{p_i q_j\}$, очевидно, образует распределение, то в силу (2)

$$- \sum q_{ij} \log p_i q_j \geq - \sum q_{ij} \log q_{ij} = H(G),$$

при этом равенство возможно лишь при $q_{ij} = p_i q_j$.

5. Как мы видели в п. 3, информация, полученная в результате эксперимента G_1^n , состоящего в n -кратном повторении независимых испытаний G_1 , равна

$$J_{G_1^n} = - \sum_{j=1}^N \nu_j \log p_j,$$

где ν_j — число появлений исхода E_j . По закону больших чисел $\nu_j/n \xrightarrow{p} p_j$ при $n \rightarrow \infty$ и

$$\frac{1}{n} J_{G_1^n} \xrightarrow{p} H(G_1) = H_p.$$

В заключение этого параграфа отметим, что рассмотренная мера количества информации, полученной от эксперимента, может быть *выведена* как единственно возможная (с точностью до постоянного множителя), если отправляться от нескольких простых требований, которые естественно предъявить к величине, количественно оценивающей информацию¹⁾.

Интересно отметить также связь введенных понятий с вероятностями больших отклонений. Как видно из теорем 5.3, 5.6, разность между «смещенной» энтропией $-\sum p_j^* \ln p_j$ и энтропией $-\sum p_j^* \ln p_j^*$ ($p_j^* = \nu_j/n$ есть относительные частоты появления E_j) является аналогом функции отклонений (см. § 8.8) в многомерном случае.

§ 2. Энтропия конечной цепи Маркова.

Теорема об асимптотическом поведении информации длинного сообщения, ее приложения

1. Пусть $\{X_k\}_{k=1}^\infty$ — стационарная конечная цепь Маркова, состоящая из одного класса существенных состояний без подклассов, и пусть E_1, \dots, E_N — состояния цепи. Стационарность цепи означает, что $P(X_1 = j) = \pi_j$ совпадают с финальными вероятностями. Тогда, очевидно,

$$P(X_2 = j) = \sum_k \pi_k p_{kj} = \pi_j, \quad P(X_3 = j) = \pi_j \quad \text{и т. д.}$$

Пусть G_k есть эксперимент, состоящий в выяснении значения X_k (т. е. состояния, в которое попала система на k -м шаге). Тогда, если $X_{k-1} = i$, то энтропия k -го шага равна

$$H(G_k | X_{k-1} = i) = - \sum_j p_{ij} \log p_{ij}.$$

Энтропия стационарной цепи Маркова есть, по определению, число

$$H = EH(G_k | X_{k-1}) = H(G_k | G_{k-1}) = - \sum_i \pi_i \sum_j p_{ij} \log p_{ij}.$$

Рассмотрим первые n шагов цепи Маркова X_1, \dots, X_n . Энтропия этого сложного эксперимента $G^{(n)} = G_1 \times \dots \times G_n$ в силу марковости равна

$$H(G^{(n)}) = H(G_1) + H(G_2 | G_1) + \dots + H(G_n | G_{n-1}) = - \sum \pi_j \log p_j + (n-1)H \sim nH$$

¹⁾ См., например, [21].

при $n \rightarrow \infty$. Если бы X_k были независимы, то, как мы видели, здесь было бы точное равенство.

2. Рассмотрим теперь цепочку (X_1, \dots, X_n) как некоторое сообщение (событие) C_n и обозначим, как и прежде, $I(C_n) = -\log P(C_n)$ количество информации, содержащейся в сообщении C_n . Значение $I(C_n)$ есть функция на пространстве элементарных событий, равная информации $J_{G^{(n)}}$, содержащейся в эксперименте $G^{(n)}$. Покажем сейчас, что, как и для независимых X_k , эта информация с вероятностью, близкой к 1, ведет себя асимптотически как nH . Так что H есть приближенное значение количества информации, приходящееся в среднем на одно испытание в последовательности $\{X_k\}_{k=1}^{\infty}$.

Теорема 1. При $n \rightarrow \infty$

$$\frac{I(C_n)}{n} = \frac{-\log P(C_n)}{n} \xrightarrow{p} H.$$

Утверждение теоремы означает, что все цепочки C_n при любом $\delta > 0$ можно разбить на 2 класса. Для первого класса $|I(C_n)/n - H| < \delta$. Сумма вероятностей элементов второго класса сходится к 0 при $n \rightarrow \infty$.

Доказательство. Построим по данной цепи новую цепь Маркова $\{Y_k\}_{k=1}^{\infty}$ следующим образом. Положим $Y_k = (X_k, X_{k+1})$. Состояниями этой цепи будут пары состояний (E_i, E_j) цепи $\{X_k\}$, для которых $p_{ij} > 0$. Переходные вероятности определяются, очевидно, следующим образом:

$$p_{(i,j)(k,l)} = \begin{cases} 0, & j \neq k, \\ p_{kl}, & j = k. \end{cases}$$

Заметим, что по индукции легко доказать соотношение

$$p_{(i,j)(k,l)}(n) = p_{jk}(n-1)p_{kl}. \quad (3)$$

Из определения $\{Y_k\}$ следует, что для этой цепи верна эргодическая теорема. Это видно и непосредственно из (3), при этом

$$\lim_{n \rightarrow \infty} p_{(i,j)(k,l)}(n) = \pi_k p_{kl}.$$

Нам потребуется теперь закон больших чисел для числа попаданий $m_{(k,l)}(n)$ цепи $\{Y_j\}_{j=1}^{\infty}$ в состояние (k, l) за время n . В силу этого закона (см. теорему 11.8)

$$\frac{m_{kl}(n)}{n} \xrightarrow{p} \pi_k p_{kl} \quad \text{при } n \rightarrow \infty.$$

Рассмотрим случайную величину $P(C_n)$

$$\begin{aligned} P(C_n) &= P(E_{X_1} E_{X_2} \cdots E_{X_n}) = P(E_{X_1}) P(E_{X_2} | E_{X_1}) \cdots P(E_{X_n} | E_{X_{n-1}}) = \\ &= p_{X_1} p_{X_1 X_2} \cdots p_{X_{n-1} X_n} = p_{X_1} \prod_{(k,l)} p_{kl}^{m_{(k,l)}(n-1)} \end{aligned}$$

Произведение здесь берется по всем парам (k, l) . Поэтому $(\pi_i = P(X_1 = i))$

$$\log P(C_n) = \log p_{X_1} + \sum_{k,l} m_{(k,l)}(n-1) \log p_{kl},$$

$$\frac{\log P(C_n)}{n} \xrightarrow{p} \sum_{k,l} \pi_k p_{kl} \log p_{kl} = -H.$$

3. Теорема 1 имеет важное следствие. Упорядочим все цепочки C_n длины n в порядке убывания их вероятностей и будем отбирать наиболее вероятные до тех пор, пока впервые суммарная вероятность отобранных цепочек не превзойдет некоторый заранее заданный уровень α , $0 < \alpha < 1$. Число отобранных цепочек обозначим $M_\alpha(n)$.

Теорема 2. Для каждого $0 < \alpha < 1$ существует один и тот же

$$\lim_{n \rightarrow \infty} \frac{\log M_\alpha(n)}{n} = H.$$

Доказательство. Пусть $\delta > 0$ — некоторое число, которое может быть произвольно малым. Цепочку C_n отнесем к категории K_1 , если ее вероятность $P(C_n) > 2^{-n(H-\delta)}$. Будем говорить, что C_n принадлежит к категории K_2 , если

$$2^{-n(H+\delta)} < P(C_n) \leq 2^{-n(H-\delta)}$$

и, наконец, C_n принадлежит к третьей категории K_3 , если

$$P(C_n) \leq 2^{-n(H+\delta)}.$$

Так как по теореме 1 $P(C_n \in K_1 \cup K_3) \rightarrow 0$ при $n \rightarrow \infty$, то в количестве $M_\alpha(n)$ (этим символом мы будем обозначать и само множество отобранных цепочек) войдут только цепочки из K_1 и K_2 . При этом последняя цепочка, вошедшая в $M_\alpha(n)$, обозначим ее $C_{\alpha,n}$, будет из K_2 . Это значит, что

$$M_\alpha(n) 2^{-n(H+\delta)} < \sum_{C_n \in M_\alpha(n)} P(C_n) < \alpha + P(C_{\alpha,n}) < \alpha + 2^{-n(H-\delta)}.$$

Отсюда следует, что

$$\frac{\log M_\alpha(n)}{n} < \frac{\log(\alpha + 2^{-n(H-\delta)})}{n} + H + \delta.$$

В силу произвольности δ

$$\limsup_{n \rightarrow \infty} \frac{\log M_\alpha(n)}{n} \leq H.$$

С другой стороны, цепочки из K_2 , принадлежащие $M_\alpha(n)$, имеют суммарную вероятность $\geq \alpha - P(K_1)$. Если $M_\alpha^{(2)}(n)$ — число указанных цепочек, то

$$M_\alpha^{(2)}(n) 2^{-n(H-\delta)} \geq \alpha - P(K_1).$$

Следовательно,

$$M_\alpha(n) 2^{-n(H-\delta)} \geq \alpha - P(K_1).$$

Так как $P(K_1) \rightarrow 0$ при $n \rightarrow \infty$, то при достаточно большом n

$$\frac{\log M_\alpha(n)}{n} \geq H - \delta + \frac{1}{n} \log \frac{\alpha}{2}.$$

Отсюда следует, что $\liminf_{n \rightarrow \infty} \frac{\log M_\alpha(n)}{n} \geq H$. ■

Можно получить теперь полезную интерпретацию этой теоремы. Пусть N — число состояний цепи. Предположим для простоты, что $N = 2^m$. Тогда число

все возможных сообщений длины n (цепочек C_n) будет равно $N^n = 2^{nm}$. Предположим далее, что эти сообщения передаются с помощью двоичного кода, так что на кодирование каждого состояния нами затрачивается m двоичных символов. Значит, при таком способе передачи, назовем его *непосредственным кодированием*, длина сообщений составит nm (в русском языке, например, для моделирования которого можно использовать цепи Маркова, можно считать $N = 32$, $m = 5$). Утверждение теоремы 2 означает, что при больших n из этих 2^{nm} цепочек с вероятностью $1 - \varepsilon$, $\varepsilon > 0$, будет передаваться лишь 2^{nH} сообщений. Вероятность появления остальных будет мала, если мы фиксируем малое ε . Отсюда легко установить существование другого — более экономного кода, с помощью которого на передачу сообщения с большой вероятностью потребуется меньшее число единиц. Действительно, отобранные 2^{nH} сообщений можно просто занумеровать с помощью того же двоичного кода и передавать их номер. На это потребуется, очевидно, nH двоичных единиц. Так как всегда $H \leq \log N = m$, то мы получим экономию в длительности сообщения в $m/H \geq 1$ раз.

Это есть частный случай так называемой *основной теоремы кодирования* для цепей Маркова: при больших n существует код, с помощью которого с высокой вероятностью исходное сообщение C_n может быть передано последовательностью сигналов в m/H раз более короткой, чем при непосредственном кодировании.

Приведенный выше способ кодирования, конечно, представляет собой скорее утрированный пример, чем эффективный рецепт укорачивания сообщения. По этому поводу следует отметить, что отыскание действительно эффективных способов кодирования представляет собой довольно сложную задачу. В азбуке Морзе, например, разумно более часто встречающимся буквам ставить в соответствие более короткие последовательности точек и тире. Но сокращение текста в m/H раз при этом, конечно, не достигается. Некоторые приемы сокращения использовались и в этой книге. Например, часто встречающиеся слова «характеристическая функция» были заменены символом «х. ф.». Можно было бы добиться больших результатов, применяя, скажем, стенографию. Конструкция кода с высоким коэффициентом сокращения, безусловно, будет весьма сложной. Полученные теоремы указывают верхнюю границу для результата, которого мы при этом можем достигнуть.

Так как для последовательности независимых равновероятных знаков $H = \sum \frac{1}{n} \log N = m$, то такой текст *несократим*. Поэтому некоторым критерием при построении различных кодов может служить близость «новых» сообщений (закодированных в новом алфавите) к последовательности равновероятных символов.

Однако при этом следует учитывать, что та «избыточность» текста, с которой мы «боремся», во многих случаях представляет собой явление полезное и нужное — ведь иначе было бы невозможным выявлять опечатки или восстанавливать пропуски так же легко, как мы восстанавливаем, например, букву «р» в слове «инфо·мация». Читатель, видимо, знает, как трудно читать сильно сокращенный и формализованный математический текст. Так что при работе с идеальным кодом (если бы такой удалось найти) ошибки при передаче информации недопустимы, поскольку невозможно восстановить пропущенный или искаженный символ в последовательности равновероятных знаков. В связи с этим возникает одна из основных в теории информации задач о кодировании, которое оставляло бы наименьшую «избыточность», достаточную для ликвидации помех при передаче.

Г л а в а 14

Мартингалы

§ 1. Определения, простейшие свойства, примеры

В гл. 12 мы рассмотрели последовательности зависимых случайных величин X_0, X_1, \dots , связанных в цепь Маркова. Зависимость характеризовалась при этом вероятностями переходов, которые определяли распределение X_{n+1} при условии, что X_n известно. Это позволило весьма полно изучить свойства цепей Маркова.

В этой главе будет рассмотрен другой тип последовательностей зависимых случайных величин, в котором характеристика зависимости происходит не через условное распределение X_{n+1} , а лишь через среднее значение X_{n+1} при условии, что задана «предыстория» X_0, \dots, X_n . Оказывается, это также позволяет получить многие весьма общие результаты об исследуемом объекте.

Пусть дано вероятностное пространство $(\Omega, \mathfrak{F}, P)$, и на нем задана последовательность случайных величин X_0, X_1, \dots . Пусть далее на $(\Omega, \mathfrak{F}, P)$ выделено семейство (или поток) вложенных σ -алгебр $\{\mathfrak{F}_n\}_{n \geq 0}$: $\mathfrak{F}_0 \subseteq \mathfrak{F}_1 \subseteq \dots \subseteq \mathfrak{F}_n \subseteq \mathfrak{F}$.

Определение 1. Последовательность пар $\{X_n, \mathfrak{F}_n; n \geq 0\}$, называемая *стохастической последовательностью*, является *мартингалом* (говорят также, что $\{X_n\}$ есть *мартингал относительно системы σ -алгебр \mathfrak{F}_n*), если для всех $n \geq 0$

$$1) E|X_n| < \infty, \quad (1)$$

$$2) X_n \text{ измеримы относительно } \mathfrak{F}_n,$$

$$3) E(X_m | \mathfrak{F}_n) = X_n \text{ при любом } m > n. \quad (2)$$

Последовательность $\{X_n, \mathfrak{F}_n; n \geq 0\}$ называется *субмартингалом* (*супермартингалом*), если выполнены условия 1)–3), но знак «=» в 3) заменен на « \geq » (« \leq »).

Будем говорить, что последовательность $\{X_n\}$ образует *мартингал* (*субмартингал*, *супермартингал*), если для $\mathfrak{F}_n = \sigma(X_0, \dots, X_n)$ пары $\{X_n, \mathfrak{F}_n\}$ образуют последовательность того же наименования. Субмартингалы и супермартингалы часто называют *полумартингалами*.

Ясно, что соотношение (2) в силу свойств у.м.о. эквивалентно тому, что $E(X_{n+1} | \mathfrak{F}_n) = X_n$, поскольку в этом случае

$$E(X_m | \mathfrak{F}_n) = E[E(X_m | \mathfrak{F}_{m-1}) | \mathfrak{F}_n] = E(X_{m-1} | \mathfrak{F}_n) = \dots = X_n.$$

Аналогично обстоит дело с полумартингалами.

Если $\{X_n\}$ — мартингал, то $E(X_{n+1} | \sigma(X_0, \dots, X_n)) = X_n$, и по свойству у.м.о.

$$E(X_{n+1} | \sigma(X_n)) = E[E(X_{n+1} | \sigma(X_0, \dots, X_n)) | \sigma(X_n)] = E(X_n | \sigma(X_n)) = X_n.$$

Следовательно, для мартингалов, как и для цепей Маркова,

$$E(X_{n+1} | \sigma(X_0, \dots, X_n)) = E(X_{n+1} | \sigma(X_n)).$$

Однако на этом сходство кончается, так как для мартингала это равенство не сохраняется для распределений, но зато требуется, чтобы $E(X_{n+1}|\sigma(X_n)) = X_n$.

Пример 1. Пусть ξ_n , $n \geq 0$, независимы. Тогда $X_n = \xi_0 + \dots + \xi_n$ образует мартингал (субмартингал, супермартингал), если $E\xi_n = 0$ ($E\xi_n \geq 0$, $E\xi_n \leq 0$). Очевидно, что X_n связаны также в цепь Маркова.

То же справедливо относительно $X_n = \prod_{k=0}^n \xi_k$ при $E\xi_n = 1$.

Пример 2. Пусть ξ_n , $n \geq 0$, независимы. Тогда $X_n = \sum_{k=1}^n \xi_{k-1}\xi_k$, $n \geq 1$,

$X_0 = \xi_0$, образует мартингал, если $E\xi_n = 0$, так как $E(X_{n+1}|\sigma(X_0, \dots, X_n)) = X_n + E(\xi_n \xi_{n+1}|\sigma(\xi_n)) = X_n$.

Очевидно, что здесь $\{X_n\}$ не является цепью Маркова. Пример последовательности, которая образует цепь Маркова, но не является мартингалом, можно получить, рассматривая, скажем, блуждание на отрезке с отражением от границ отрезка (пример 12.1).

Наряду с множеством $\{0, 1, \dots\}$ индексов для X_n мы будем использовать и другие, например, $\{-\infty < n < \infty\}$ или $\{n \leq -1\}$, а также наборы целых чисел, включающие в себя значения $\pm\infty$, скажем, $\{0 \leq n \leq \infty\}$. Эти множества будем обозначать одним символом \mathcal{N} , а мартингалы (полумартингалы) записывать в форме $\{X_n, \mathfrak{F}_n; n \in \mathcal{N}\}$. При этом под $\mathfrak{F}_{-\infty}$ будем понимать σ -алгебру $\bigcap_{n \in \mathcal{N}} \mathfrak{F}_n$, а под \mathfrak{F}_{∞} будем понимать σ -алгебру $\sigma\left(\bigcup_{n \in \mathcal{N}} \mathfrak{F}_n\right)$, порожденную $\bigcup_{n \in \mathcal{N}} \mathfrak{F}_n$, так что $\mathfrak{F}_{-\infty} \subset \mathfrak{F}_n \subset \mathfrak{F}_{\infty} \subset \mathfrak{F}$ при любом $n \in \mathcal{N}$.

Определение 2. Стохастическая последовательность $\{X_n, \mathfrak{F}_n; n \in \mathcal{N}\}$ называется **мартингалом** (субмартингалом, супермартингалом), если выполнены условия определения 1 при любых $n \in \mathcal{N}$, $m \in \mathcal{N}$, $m > n$.

Если дан мартингал $\{X_n, \mathfrak{F}_n; n \in \mathcal{N}\}$ и левая граница n_0 множества \mathcal{N} конечна (например, $\mathcal{N} = \{0, 1, \dots\}$), то мартингал $\{X_n, \mathfrak{F}_n\}$ всегда можно продолжить «на всю ось», положив $\mathfrak{F}_n = \mathfrak{F}_{n_0}$, $X_n = X_{n_0}$ при $n < n_0$. Так же обстоит дело и с правым концом. Поэтому, если задан мартингал (полумартингал) $\{X_n, \mathfrak{F}_n; n \in \mathcal{N}\}$, то можно всегда, не ограничивая общности, считать, что задан мартингал (полумартингал) $\{X_n, \mathfrak{F}_n; -\infty < n < \infty\}$.

Пример 3. Пусть $\{\mathfrak{F}_n; -\infty < n < \infty\}$ — заданная последовательность вложенных σ -алгебр и ξ есть заданная случайная величина на $(\Omega, \mathfrak{F}, P)$, $E|\xi| < \infty$. Тогда $\{X_n, \mathfrak{F}_n; -\infty \leq n \leq \infty\}$, где $X_n = E(\xi|\mathfrak{F}_n)$, образует мартингал.

Действительно, по свойству у. м. о. при любом $m \leq \infty$, $m > n$

$$E(X_m|\mathfrak{F}_n) = E[E(\xi|\mathfrak{F}_m)|\mathfrak{F}_n] = E(\xi|\mathfrak{F}_n) = X_n.$$

Определение 3. Мартингал в примере 3 будем называть **мартингалом**, порожденным случайной величиной ξ (и семейством $\{\mathfrak{F}_n\}$).

Определение 4. Множество \mathcal{N}_+ будем называть **замыканием \mathcal{N} справа**, если

- 1) $\mathcal{N}_+ = \mathcal{N}$, когда максимальный элемент \mathcal{N} конечен,
- 2) $\mathcal{N}_+ = \mathcal{N} \cup \{+\infty\}$, если \mathcal{N} не ограничено справа.

Если $\mathcal{N} = \mathcal{N}_+$, то будем говорить, что \mathcal{N} **замкнуто справа**. Будем говорить, что мартингал (полумартингал) $\{X_n, \mathfrak{F}_n; n \in \mathcal{N}\}$ **замкнут справа**, если \mathcal{N} замкнуто справа.

Лемма 1. Мартингал $\{X_n, \mathfrak{F}_n; n \in \mathcal{N}\}$ порожден некоторой случайной величиной тогда и только тогда, когда он замкнут справа.

Доказательство леммы очевидно. В одну сторону оно следует из примера 3. В другую — из равенства

$$E(X_N | \mathfrak{F}_n) = X_n, \quad N = \sup\{k; k \in \mathcal{N}\},$$

так что $\{X_n, \mathfrak{F}_n\}$ порожден X_N . ■

Приведем один интересный и более конкретный пример мартингала, порожденного случайной величиной.

Пример 4. Пусть ξ_1, ξ_2, \dots независимы, одинаково распределены, $E|\xi_1| < \infty$. Положим

$$S_n = \xi_1 + \dots + \xi_n, \quad X_{-n} = \frac{S_n}{n},$$

$$\mathfrak{F}_{-n} = \sigma(S_n, S_{n+1}, \dots) = \sigma(S_n, \xi_{n+1}, \dots).$$

Тогда $\mathfrak{F}_{-n} \subset \mathfrak{F}_{-n+1}$ и при любом k , $1 \leq k \leq n$,

$$E(\xi_k | \mathfrak{F}_{-n}) = E(\xi_1 | \mathfrak{F}_{-n}).$$

Последнее равенство справедливо в силу симметрии. Из него следует, что

$$S_n = E(S_n | \mathfrak{F}_{-n}) = \sum_{k=1}^n E(\xi_k | \mathfrak{F}_{-n}) = nE(\xi_1 | \mathfrak{F}_{-n}),$$

$$\frac{S_n}{n} = E(\xi_1 | \mathfrak{F}_{-n}).$$

Это означает, что $\{X_n, \mathfrak{F}_n; n \leq -1\}$ образует мартингал, порожденный ξ_1 .

Получим ряд вспомогательных утверждений, характеризующих простейшие свойства мартингалов и полумартингалов. При этом мы часто будем ограничиваться рассмотрением лишь субмартингалов, так как соответствующие свойства супермартингалов сразу получаются переходом к последовательности $Y_n = -X_n$, если $\{X_n\}$ — субмартингал.

Лемма 2. 1) Свойство $\{X_n, \mathfrak{F}_n; n \in \mathcal{N}\}$ быть мартингалом эквивалентно неизменности по $m \geq n$ функций множества (интеграла)

$$E(X_m; A) = E(X_n; A) \quad (3)$$

при любом $A \in \mathfrak{F}_n$. В частности, $E X_m = \text{const}$.

2) Свойство $\{X_n, \mathfrak{F}_n; n \in \mathcal{N}\}$ быть субмартингалом эквивалентно монотонному возрастанию по $m \geq n$ функций множества

$$E(X_m; A) \geq E(X_n; A) \quad (4)$$

при каждом $A \in \mathfrak{F}_n$. В частности, $E X_m \uparrow$.

Доказательство является очевидным следствием определений. Если выполнено (3), то, по определению у. м. о., $X_n = E(X_m | \mathfrak{F}_n)$ и наоборот. Пусть теперь выполнено (4). Обозначим $Y_n = E(X_m | \mathfrak{F}_n)$. Тогда из (4) следует, что $E(Y_n; A) \geq E(X_n; A)$, $E(Y_n - X_n; A) \geq 0$ при любом $A \in \mathfrak{F}_n$. Отсюда вытекает, что с вероятностью 1 $Y_n = E(X_m | \mathfrak{F}_n) \geq X_n$. Обратное утверждение получается столь же просто. ■

Лемма 3. Пусть $\{X_n, \mathfrak{F}_n; n \in \mathcal{N}\}$ — мартингал и $g(x)$ — выпуклая вниз функция с $E|g(X_n)| < \infty$. Тогда $\{g(X_n), \mathfrak{F}_n; n \in \mathcal{N}\}$ образует субмартингал.

Если к тому же $g(x)$ не убывает, то утверждение сохранится и в том случае, когда $\{X_n, \mathfrak{F}_n; n \in \mathcal{N}\}$ — субмартингал.

Доказательство. Обозначим $\xi_{n+1} = X_{n+1} - X_n$. Тогда

$$E(g(X_{n+1})|\mathfrak{F}_n) = E(g(X_n + \xi_{n+1})|\mathfrak{F}_n).$$

Воспользуемся свойствами у.м.о. и, в частности, неравенством Иенсена, примененным к функции $g_1(x) = g(X_n + x)$ (см. § 4.8). Мы получим

$$E(g_1(\xi_{n+1})|\mathfrak{F}_n) \geq g_1(E(\xi_{n+1}|\mathfrak{F}_n)) = g_1(0) = g(X_n).$$

Точно так же проверяется утверждение для субмартингалов. ■

Лемма 4. Пусть $\{X_n, \mathfrak{F}_n; n \in \mathcal{N}\}$ — замкнутый справа субмартингал, $X_n(a) = \max(X_n, a)$. Тогда $\{X_n(a), \mathfrak{F}_n; n \in \mathcal{N}\}$ при любом a — равномерно интегрируемый субмартингал. Если $\{X_n, \mathfrak{F}_n; n \in \mathcal{N}\}$ — замкнутый справа мартингал, то он равномерно интегрируем.

Доказательство. Пусть $N = \sup\{k; k \in \mathcal{N}\}$. Тогда по лемме 3 $\{X_n(a), \mathfrak{F}_n; n \in \mathcal{N}\}$ — субмартингал. Следовательно, при любом $c > 0$ ($X^+ = \max(0, X)$)

$$cP(X_n(a) > c) \leq E(X_n(a); X_n(a) > c) \leq E(X_N(a); X_n(a) > c) \leq EX_N^+(a),$$

$$P(X_n(a) > c) \leq \frac{E(X_N^+(a))}{c} \rightarrow 0$$

при $c \rightarrow \infty$ равномерно по n . Поэтому мы имеем требуемую равномерную интегрируемость

$$\sup_n E(X_n(a); X_n(a) > c) \leq \sup_n E(X_N(a); X_n(a) > c) \rightarrow 0,$$

поскольку $\sup_n P(X_n(a) > c) \rightarrow 0$ при $c \rightarrow \infty$ (см. лемму 3 в Приложении 3; наличие срезки на уровне a позволило избежать оценки «отрицательных хвостов»).

Если $\{X_n(a), \mathfrak{F}_n; n \in \mathcal{N}\}$ — мартингал, то его равномерная интегрируемость следует из первого утверждения, примененного к субмартингалу $\{|X_n|, \mathfrak{F}_n; n \in \mathcal{N}\}$. ■

В какой-то мере природу мартингалов поясняет также следующий пример.

Пример 5. Пусть ξ_1, ξ_2, \dots — произвольная последовательность случайных величин, $E|\xi_k| < \infty$, $\mathfrak{F}_n = \sigma(\xi_1, \dots, \xi_n)$ при $n \geq 1$, $\mathfrak{F}_0 = (\emptyset, \Omega)$ (тривиальная σ -алгебра),

$$S_n = \sum_{k=1}^n \xi_k, \quad Z_n = \sum_{k=1}^n E(\xi_k | \mathfrak{F}_{k-1}), \quad X_n = S_n - Z_n.$$

Тогда $\{X_n, \mathfrak{F}_n; n \geq 1\}$ образует мартингал. Это следует из того, что

$$E(S_{n+1} - Z_{n+1} | \mathfrak{F}_n) = E(X_n + \xi_{n+1} - E(\xi_{n+1} | \mathfrak{F}_n) | \mathfrak{F}_n) = X_n.$$

Другими словами, для произвольной последовательности $\{\xi_n\}$ последовательность S_n можно «подправить» так называемой «предсказуемой» последовательностью Z_n (значение Z_n определяется по S_1, \dots, S_{n-1}) так, что $S_n - Z_n$ будет мартингалом.

Лемма 5. Если $\{X_n, \mathfrak{F}_n; n \geq 0\}$ — мартингал и $g(x)$ — выпуклая вниз функция, $E|g(X_n)| < \infty$, $n \geq 0$, то последовательность $\{g(X_n), \mathfrak{F}_n; n \geq 0\}$ образует субмартингал.

Если $\{X_n, \mathfrak{F}_n\}$ — субмартингал и $g(x)$ — выпуклая вниз неубывающая функция, $E|g(X_n)| < \infty$, то $\{g(X_n), \mathfrak{F}_n\}$ также образует субмартингал.

Доказательство обоих утверждений немедленно следует из неравенства Иенсена:

$$E(g(X_{n+1}) | \mathfrak{F}_n) \geq g(E(X_{n+1} | \mathfrak{F}_n)) \geq g(E(X_n | \mathfrak{F}_n)).$$

Ясно, что функция $g(x) = |x|^p$ при $p \geq 1$ удовлетворяет условиям первой части леммы, а функция $g(x) = e^{\lambda x}$ при $\lambda > 0$ — второй части леммы.

§ 2. О сохранении свойства быть мартингалом при замене времени на случайное.

Тождество Вальда

В этом параграфе $\mathcal{N} = \{n \geq 0\}$. Напомним определение момента остановки.

Определение 5. Для последовательности $\{\mathfrak{F}_n; n \geq 0\}$ случайную величину ν будем называть *моментом остановки* (относительно $\{\mathfrak{F}_n\}$), если $\{\nu \leq n\} \in \mathfrak{F}_n$ при любом $n \geq 0$.

Очевидно, что постоянная $\nu \equiv t$ является моментом остановки. Если ν — момент остановки, то $\nu(m) = \min(\nu, m)$ при любом фиксированном m также будет моментом остановки, поскольку при $n \geq m$ выполняется $\nu(m) \leq m \leq n$, $\{\nu(m) \leq n\} = \Omega \in \mathfrak{F}_n$; при $n < m$ $\{\nu(m) \leq n\} = \{\nu \leq n\} \in \mathfrak{F}_n$.

Если ν — момент остановки, то $\{\nu = n\} = \{\nu \leq n\} - \{\nu \leq n-1\} \in \mathfrak{F}_n$, $\{\nu \geq n\} = \Omega - \{\nu \leq n-1\} \in \mathfrak{F}_{n-1}$. Наоборот, если $\{\nu = n\} \in \mathfrak{F}_n$, то $\{\nu \leq n\} \in \mathfrak{F}_n$ и, следовательно, ν — момент остановки.

Пусть дан мартингал $\{X_n, \mathfrak{F}_n; n \geq 0\}$. Типичным примером момента остановки является момент ν попадания X_n в заданное измеримое множество B :

$$\nu = \inf\{n \geq 0: X_n \in B\}$$

($\nu = \infty$, если все $X_n \notin B$). Действительно,

$$\{\nu = n\} = \{X_0 \notin B, \dots, X_{n-1} \notin B, X_n \in B\} \in \mathfrak{F}_n.$$

Если ν — собственный момент остановки ($\{\nu = \infty\} = \emptyset$), то X_ν — случайная величина, так как

$$X_\nu = \sum_{n=0}^{\infty} X_n I_{\{\nu=n\}}.$$

σ -алгеброй \mathfrak{F}_ν будем называть σ -алгебру, образованную множествами $A \in \mathfrak{F}$, для которых $A \cap \{\nu = n\} \in \mathfrak{F}_n$, $n = 0, 1, \dots$. Можно представлять себе, что \mathfrak{F}_ν порождена событиями $\{\nu \leq n\} \cap B_n$, $n = 0, 1, \dots$, где $B_n \in \mathfrak{F}_n$. Ясно, что ν и X_ν измеримы относительно \mathfrak{F}_ν . Если ν_1, ν_2 — два момента остановки, то $\{\nu_2 \geq \nu_1\} \in \mathfrak{F}_{\nu_1}$, $\{\nu_2 \geq \nu_1\} \in \mathfrak{F}_{\nu_2}$, так как $\{\nu_2 \geq \nu_1\} = \bigcup_n \{\nu_2 = n\} \{\nu_1 \leq n\}$.

Мы уже знаем, что если $\{X_n, \mathfrak{F}_n\}$ — мартингал, то значение $E X_n$ постоянно при всех n . Сохранится ли это свойство для $E X_\nu$, где ν — момент остановки? Из тождества Вальда мы знаем, что для примера 1 это так. В общем случае справедлива

Теорема 1 (Дуб). Пусть $\{X_n, \mathfrak{F}_n; n \geq 0\}$ — мартингал (субмартингал) и ν_1, ν_2 — моменты остановки такие, что

$$E|X_{\nu_i}| < \infty, \quad i = 1, 2, \quad (5')$$

$$\lim_{n \rightarrow \infty} E(|X_n|; \nu_2 \geq n) = 0. \quad (5'')$$

Тогда на множестве $\{\nu_2 \geq \nu_1\}$

$$E(X_{\nu_2} | \mathfrak{F}_{\nu_1}) = X_{\nu_1} \quad (\geq X_{\nu_1}). \quad (6)$$

Это и есть обобщение мартингального (субмартингального) свойства для случайного времени.

Следствие 1. Если $\nu_2 = \nu \geq 0$ — произвольный момент остановки, то, полагая $\nu_1 = n$ (это тоже момент остановки), получим, что на множестве $\nu \geq n$

$$E(X_\nu | \mathfrak{F}_n) = X_n, \quad EX_\nu = EX_0,$$

или, что то же, при любом $A \in \mathfrak{F}_n \cap \{\nu \geq n\}$

$$E(X_\nu; A) = E(X_n; A).$$

Для субмартингалов здесь надо вместо «=» писать « \geq ».

Доказательство теоремы 1. Чтобы доказать (6), достаточно убедиться, что для любого $A \in \mathfrak{F}_{\nu_1}$

$$E(X_{\nu_2}; A \cap \{\nu_2 \geq \nu_1\}) = E(X_{\nu_1}; A \cap \{\nu_2 \geq \nu_1\}). \quad (7)$$

Так как ν_i дискретны, то для этого, в свою очередь, достаточно установить (7) для множеств $A_n = A \cap \{\nu_1 = n\} \in \mathfrak{F}_n$, $n = 0, 1, \dots$, т. е. установить равенство

$$E(X_{\nu_2}; A_n \cap \{\nu_2 \geq n\}) = E(X_n; A_n \cap \{\nu_2 \geq n\}). \quad (8)$$

Тем самым доказательство сведено к случаю $\nu_1 = n$. Имеем

$$\begin{aligned} E(X_n; A_n \cap \{\nu_2 \geq n\}) &= E(X_n; A_n \cap \{\nu_2 = n\}) + \\ &+ E(X_n; A_n \cap \{\nu_2 \geq n+1\}) = E(X_{\nu_2}; A_n \cap \{\nu_2 = n\}) + \\ &+ E(X_{n+1}; A_n \cap \{\nu_2 \geq n+1\}). \end{aligned}$$

Здесь использован тот факт, что $\{\nu_2 \geq n+1\} \in \mathfrak{F}_n$, и свойство (3) мартингала.

Применяя полученное равенство $m - n$ раз, находим

$$E(X_{\nu_2}; A_n \cap \{n \leq \nu_2 < m\}) = E(X_n; A_n \cap \{\nu_2 \geq n\}) - E(X_m; A_n \cap \{\nu_2 \geq m\}). \quad (9)$$

Последнее выражение в силу (5'') сходится к нулю для некоторой последовательности $m \rightarrow \infty$.

Так как $A_{n,m} = A_n \cap \{n \leq \nu_2 < m\} \uparrow B_n = A_n \cap \{n \leq \nu_2\}$, то по свойству интегралов и в силу (9)

$$E(X_{\nu_2}; A_n \cap \{n \leq \nu_2\}) = \lim_{m \rightarrow \infty} E(X_{\nu_2}; A_{n,m}) = E(X_n; A_n \cap \{\nu_2 \geq n\}).$$

Мы доказали равенство (8), а вместе с ним и теорему 1 для мартингалов. Изменения, которые надо внести в доказательство для субмартингалов, состоят в замене в нескольких местах знака равенства на неравенство. ■

Условия теоремы 1 бывают выполнены далеко не всегда, даже в довольно простых ситуациях. Рассмотрим, например, безобидную игру (см. примеры 4.8, 4.11),

в которой $z + S_n$ есть капитал первого игрока после n партий в игре с бесконечно богатым противником и при условии, что игрок еще не разорился. Здесь $z > 0$, $S_n = \sum_{k=1}^n \xi_k$, $P(\xi_k = \pm 1) = 1/2$, $\eta(z) = \min\{k: S_k = -z\}$ есть, очевидно, марковский момент; последовательность $\{S_n; n \geq 0\}$, $S_0 = 0$ образует мартингал: $S_{\eta(z)} = -z$. Следовательно, $ES_{\eta(z)} = -z \neq ES_n = 0$, и равенство (8) при $\nu_1 = 0$, $\nu_2 = \eta(z)$, $z > 0$, $n > 0$ не имеет места. В рассматриваемом случае это означает, что не выполнено условие (5'') (это связано с тем фактом, что $E\eta(z) = \infty$).

Вообще говоря, условия (5) теоремы 1 могут быть весьма трудными для проверки. Поэтому полезными для приложений являются следующие утверждения.

Положим для краткости

$$\xi_n = X_n - X_{n-1}, \quad \xi_0 = X_0, \quad Y_n = \sum_{k=0}^n |\xi_k|, \quad n = 0, 1, \dots$$

Лемма 5. Для выполнения (5'), (5'') (при $\nu_i = \nu$) достаточно, чтобы

$$EY_\nu < \infty. \quad (10)$$

Доказательство почти очевидно, поскольку $|X_\nu| \leq Y_\nu$,

$$E(|X_n|; \nu > n) \leq E(Y_\nu; \nu > n).$$

Так как $P(\nu > n) \rightarrow 0$, $EY_\nu < \infty$, то остается воспользоваться свойством интегралов, в силу которого $E(\eta; A_n) \rightarrow 0$, если $E|\eta| < \infty$, $P(A_n) \rightarrow 0$. ■

Введем следующие обозначения:

$$a_n = E(|\xi_n| | \mathfrak{F}_{n-1}), \quad \sigma_n^2 = E(\xi_n^2 | \mathfrak{F}_{n-1}), \quad n = 0, 1, 2, \dots,$$

где \mathfrak{F}_{-1} можно считать тривиальной σ -алгеброй.

Теорема 2. Пусть $\{X_n; n \geq 0\}$ есть мартингал (субмартингал) и ν — момент остановки (относительно $\{\mathfrak{F}_n = \sigma(X_0, \dots, X_n)\}$).

1) Если

$$E\nu < \infty \quad (11)$$

и при всех $n \geq 0$ на множестве $\{\nu \geq n\} \in \mathfrak{F}_{n-1}$ выполняется

$$a_n \leq c = \text{const}, \quad (12)$$

то

$$E|X_\nu| < \infty, \quad EX_\nu = EX_0 \quad (\geq EX_0). \quad (13)$$

2) Если, кроме того, $E\sigma_n^2 = E\xi_n^2 < \infty$, то

$$EX_\nu^2 = E\left(\sum_{k=1}^{\nu} \sigma_k^2\right). \quad (14)$$

Доказательство. В силу теоремы 1, следствия 1 и леммы 5 для доказательства (13) достаточно проверить, что условия (11), (12) влекут за собой (10). Вполне аналогично доказательству теоремы 4.2 имеем

$$E|Y_\nu| = \sum_{n=0}^{\infty} \left(\sum_{k=0}^n E(|\xi_k|; \nu \geq k) \right) = \sum_{k=0}^{\infty} \sum_{n=k}^{\infty} E(|\xi_k|; \nu \geq k) = \sum_{k=0}^{\infty} E(|\xi_k|; \nu \geq k).$$

Здесь $\{\nu \geq k\} = \Omega - \{\nu \leq k-1\} \in \mathfrak{F}_{k-1}$. Поэтому по условию (12)

$$E(|\xi_k|; \nu \geq k) = E(E(|\xi_k| | \mathfrak{F}_{k-1}); \nu \geq k) \leq cP(\nu \geq k).$$

Это означает, что

$$EY_\nu \leq c \sum_{k=0}^{\infty} P(\nu \geq k) = cE\nu < \infty.$$

Докажем теперь (14). Положим $Z_n = X_n^2 - \sum_0^n \sigma_k^2$. Нетрудно видеть, что Z_n — мартингал, так как $E(X_{n+1}^2 - X_n^2 - \sigma_{n+1}^2 | \mathfrak{F}_n) = E(2X_n\xi_{n+1} + \xi_{n+1}^2 - \sigma_{n+1}^2 | \mathfrak{F}_n) = 0$. Ясно также, что $E|Z_n| < \infty$ и что $\nu(n) = \min(\nu, n)$ вместе с ν есть марковский момент. Для пары $\{Z_k\}$, $\nu(n)$ условия (5'), (5'') в силу леммы 5, очевидно, всегда выполнены. Поэтому по первой части теоремы

$$EZ_{\nu(n)} = 0, \quad EX_{\nu(n)}^2 = E \sum_{k=1}^{\nu(n)} \sigma_k^2. \quad (15)$$

Осталось убедиться, что

$$\lim_{n \rightarrow \infty} EX_{\nu(n)}^2 = EX_\nu^2, \quad \lim_{n \rightarrow \infty} E \sum_{k=1}^{\nu(n)} \sigma_k^2 = E \sum_{k=1}^{\nu} \sigma_k^2. \quad (16)$$

Второе из этих равенств следует из теоремы о монотонной сходимости ($\nu(n) \uparrow \nu$, $\sigma_k^2 \geq 0$). Первое также будет следствием теоремы о монотонной сходимости, поскольку $X_{\nu(n)}^2 \xrightarrow{\text{п. в.}} X_\nu^2$, $X_{\nu(n)}^2 \uparrow$. Чтобы убедиться в последнем, заметим, что $\{X_n^2, \mathfrak{F}_n; n \geq 0\}$ образует субмартингал и, следовательно, для любого $A \in \mathfrak{F}_n$

$$\begin{aligned} E(X_{\nu(n)}^2; A) &= E(X_\nu^2; A \cap \{\nu \leq n\}) + E(X_n^2; A \cap \{\nu > n\}) \leq \\ &\leq E(X_\nu^2; A \cap \{\nu \leq n\}) + E(E(X_{n+1}^2 | \mathfrak{F}_n); A \cap \{\nu > n\}) = \\ &= E(X_\nu^2; A \cap \{\nu < n+1\}) + E(X_{n+1}^2; A \cap \{\nu \geq n+1\}) = E(X_{\nu(n+1)}^2; A). \end{aligned}$$

Итак, из (15), (16) следует (14), и теорема полностью доказана. ■

Мы получили основное утверждение теоремы 2:

$$EX_\nu = EX_0 \quad (\geq EX_0) \quad (17)$$

для мартингалов (субмартингалов) в качестве следствия теоремы 1. Однако его можно получать и непосредственно, тем более что в основе его лежат весьма прозрачные соотношения, которые, помимо прочего, позволяют обобщить это утверждение на несобственные моменты остановки. Мы будем называть момент остановки ν *несобственным*, если $0 < P(\nu < \infty) = 1 - P(\nu = \infty) < 1$. Несобственные моменты остановки появляются, например, в случае, когда ξ_k независимы, одинаково распределены, $a = E\xi_k < 0$, $X_n = \sum_{k=1}^n \xi_k$,

$$\nu = \eta(x) = \min\{k \geq 1: X_k > x\}, \quad x \geq 0.$$

Здесь $\nu < \infty$ определена только на тех траекториях $\{X_k\}$, для которых $\sup_k X_k > x$. Если последнее неравенство не выполнено, то мы положим $\nu = \infty$. Очевидно, что $P(\nu = \infty) = P(\sup_k X_k \leq x) > 0$.

Итак, для произвольного (возможно, несобственного) момента остановки имеем

$$E(X_\nu; \nu < \infty) = \sum_{k=0}^{\infty} E(X_k; \nu = k) = \sum_{k=0}^{\infty} [E(X_k; \nu \geq k) - E(X_k; \nu \geq k+1)]. \quad (18)$$

Допустим теперь, что здесь законно изменение порядка суммирования. Тогда в силу того, что $\{\nu \geq k+1\} \subset \mathfrak{F}_k$, мы получим

$$\begin{aligned} E(X_\nu; \nu < \infty) &= EX_0 + \sum_{k=0}^{\infty} E(X_{k+1} - X_k; \nu \geq k+1) = \\ &= EX_0 + \sum_{k=0}^{\infty} EI(\nu \geq k+1)E(X_{k+1} - X_k | \mathfrak{F}_k). \end{aligned} \quad (19)$$

Так как для мартингалов (субмартингалов) последние множители $= 0$ (≥ 0), то мы получаем следующее

Теорема 2А. Если изменение порядка суммирования в (18), (19) законно, то для мартингалов (субмартингалов)

$$E(X_\nu; \nu < \infty) = EX_0 \quad (\geq EX_0). \quad (20)$$

Условия (11), (12) теоремы 2 есть не что иное, как условия, обеспечивающие абсолютную сходимость рядов в (18) (см. доказательство теоремы 2) и в (19), т. е. сумма абсолютных величин слагаемых в (19) оценивается значением

$$\sum_{k=1}^{\infty} E a_k I(\nu \geq k) \leq c E \nu < \infty,$$

где, как и прежде, $a_k = E(|\xi_k| | \mathfrak{F}_{k-1})$, $\xi_k = X_k - X_{k-1}$. Это и обеспечивает законность изменения порядка суммирования.

Доказывать (20) с помощью (18) можно и несколько иначе, при этом выделяется простое условие, обеспечивающее требуемую законность изменения порядка суммирования. Предварительно заметим, что тождество (20) подразумевает существование $E(X_\nu; \nu < \infty)$, т. е. конечность значений $E(X_\nu^\pm; \nu < \infty)$, где $x^\pm = \max(\pm x, 0)$.

Теорема 2В. 1. Пусть $\{X_n, \mathfrak{F}_n\}$ — мартингал. Тогда условие

$$\lim_{n \rightarrow \infty} E(X_n; \nu > n) = 0 \quad (21)$$

необходимо и достаточно для того, чтобы

$$\lim_{n \rightarrow \infty} E(X_\nu; \nu \leq n) = EX_0. \quad (22)$$

Условие (21) и конечность хотя бы одного из значений $E(X_\nu^\pm; \nu < \infty)$ необходимы и достаточны для выполнения (20).

2. Если $\{X_n, \mathfrak{F}_n\}$ супермартингал и

$$\liminf_{n \rightarrow \infty} E(X_n; \nu > n) \geq 0, \quad (23)$$

то

$$\limsup_{n \rightarrow \infty} E(X_\nu; \nu \leq n) \leq EX_0.$$

Если к тому же хотя бы одно из значений $E(X_\nu^\pm; \nu < \infty)$ конечно, то

$$E(X_\nu; \nu < \infty) \leq EX_0.$$

3. Первые два утверждения теоремы полностью сохраняются, если в условиях (21), (23) $E(X_n; \nu > n)$ заменить на $E(X_n; \nu \geq n)$.

Симметричные утверждения верны для субмартингалов.

Доказательство. Как мы уже отмечали, для мартингалов $E(\xi_k; \nu \geq k) = 0$. Поэтому в силу (21)

$$\begin{aligned} EX_0 &= \lim_{n \rightarrow \infty} \left[EX_0 + \sum_{k=1}^n E(\xi_k; \nu \geq k) - E(X_n; \nu \geq n+1) \right] = \\ &= \lim_{n \rightarrow \infty} \sum_{k=0}^n [E(X_k; \nu \geq k) - E(X_k; \nu \geq k+1)] = \\ &= \lim_{n \rightarrow \infty} \sum_{k=0}^n E(X_k; \nu = k) = \lim_{n \rightarrow \infty} E(X_\nu; \nu \leq n). \end{aligned}$$

Из этих же равенств следует необходимость условия (21).

Если хотя бы одно из значений $E(X_\nu^\pm; \nu < \infty)$ конечно, то в силу теоремы о монотонной сходимости

$$\begin{aligned} \lim_{n \rightarrow \infty} E(X_\nu; \nu \leq n) &= \lim_{n \rightarrow \infty} E(X_\nu^+; \nu \leq n) - \lim_{n \rightarrow \infty} E(X_\nu^-; \nu \leq n) = \\ &= E(X_j^+; \nu < \infty) - E(X_\nu^-; \nu < \infty) = E(X_\nu; \nu < \infty). \end{aligned}$$

Третье утверждение теоремы следует из того, что $\nu(n) = \min(\nu, n)$ удовлетворяет условиям п. 1 теоремы (или условиям теорем 1, 2А) и, стало быть, для мартингала $\{X_n\}$

$$EX_0 = EX_{\nu(n)} = E(X_\nu; \nu < n) + E(X_n; \nu \geq n),$$

так что (22) влечет за собой сходимость $E(X_n; \nu \geq n) \rightarrow 0$ и наоборот.

Для полумартингалов доказательство аналогично. Теорема доказана. ■

Тот факт, что утверждения (20) и (22), вообще говоря, не совпадают даже при выполнении (21) (т.е. не всегда $\lim_{n \rightarrow \infty} E(X_\nu; \nu \leq n) = E(X_\nu; \nu < \infty)$) можно проил-

люстрировать следующим примером. Пусть ξ_k независимы, $P(\xi_k = 3^k) = P(\xi_k = -3^k) = 1/2$, величина ν не зависит от $\{\xi_k\}$, $P(\nu = k) = 2^{-k}$, $k = 1, 2, \dots$. Тогда $X_0 = 0$, $X_k = X_{k-1} + \xi_k$ при $k \geq 1$ образует мартингал, $EX_n = 0$, $P(\nu < \infty) = 1$, $E(X_n; \nu > n) = EX_n \cdot P(\nu > n) = 0$ и условие (21) выполнено. Оно означает в силу (22), что $\lim_{n \rightarrow \infty} P(X_\nu; \nu \leq n) = 0$ (в этом можно убедиться и непосредственно). С другой стороны, значение $E(X_\nu; \nu < \infty) = EX_\nu$ не определено, так как $EX_\nu^+ = EX_\nu^- = \infty$. Действительно,

$$\begin{aligned} X_{k-1} &\geq -\frac{3^k - 3}{2}, \quad \{\xi_k = 3^k\} \subset \left\{ X_k \geq \frac{3^k + 3}{2} \right\}, \quad P\left(X_k \geq \frac{3^k + 3}{2}\right) \geq \frac{1}{2}, \\ EX_k^+ &\geq \frac{3^k + 3}{4}, \quad EX_\nu^+ = \sum_{k=1}^{\infty} 2^{-k} EX_k^+ \geq \sum_{k=1}^{\infty} 2^{-k-2} 3^k = \infty. \end{aligned}$$

В силу симметрии имеем также $EX_\nu^- = \infty$.

Следствие 2. 1. Если $\{X_n, \mathcal{F}_n\}$ — неотрицательный мартингал, то условие (21) необходимо и достаточно для выполнения тождества (20).

2. Если $\{X_n, \mathcal{F}_n\}$ — неотрицательный супермартингал и ν — произвольный момент остановки, то всегда

$$E(X_\nu; \nu < \infty) \leq EX_0. \quad (24)$$

Доказательство. Утверждение очевидным образом следует из теоремы 2В, так как $EX_\nu^- = 0$ и для неотрицательных супермартингалов условие (21) всегда выполнено. ■

Из теоремы 2 вытекает уже известное нам тождество Вальда (см. следствие 4.1), дополненное еще одним полезным утверждением.

Теорема 3 (тождество Вальда). Пусть ζ_1, ζ_2, \dots независимы, одинаково распределены, $S_n = \zeta_1 + \dots + \zeta_n$, $S_0 = 0$ и существует $E\zeta_1 = a$. Пусть, кроме того, ν есть момент остановки такой, что $E\nu < \infty$. Тогда

$$ES_\nu = aE\nu. \quad (25)$$

Если к тому же $\sigma^2 = D\zeta_k < \infty$, то

$$E[S_\nu - \nu a]^2 = \sigma^2 E\nu. \quad (26)$$

Доказательство. Очевидно, что $X_n = S_n - na$ образует мартингал и для него выполнены условия (11), (12). Поэтому $EX_\nu = EX_0 = 0$, что эквивалентно (25), и $EX_\nu^2 = E\nu\sigma^2$, что эквивалентно (26). ■

Пример 6. Иллюстрацией использования теоремы 3 могут быть примеры 4.11, 4.16, касающиеся безобидной игры, когда $P(\zeta_k = \pm 1) = 1/2$, $\nu = \min\{k: S_k = z_2 \text{ или } S_k = -z_1\}$ (z_1 и z_2 — капиталы соответственно 1-го и 2-го игроков). Рассмотрим теперь случай $p = P(\zeta_k = 1) \neq 1/2$. Последовательность $X_n = (q/p)^{S_n}$, $n \geq 0$, $q = 1 - p$, образует мартингал, так как $E(q/p)^{\zeta_k} = p(q/p) + q(p/q) = 1$. По теореме 3 (вероятности P_1, P_2 определены в примере 4.11)

$$EX_\nu = EX_0 = 1, \quad P_1 \left(\frac{q}{p}\right)^{z_2} + P_2 \left(\frac{q}{p}\right)^{z_1} = 1.$$

Отсюда и из равенства $P_1 + P_2 = 1$ получаем

$$P_1 = \frac{(q/p)^{z_1} - 1}{(q/p)^{z_1} - (q/p)^{z_2}}, \quad P_2 = 1 - P_1.$$

Используя снова тождество Вальда, получим также

$$E\nu = \frac{ES_\nu}{E\zeta_1} = \frac{P_1 z_2 - P_2 z_1}{p - q}.$$

Отметим, что эти равенства можно было бы получить и элементарными методами¹⁾, но это потребовало бы больше места.

В тех случаях, когда S_ν имеет достаточно простую природу, утверждения типа теорем 1, 2 позволяют получать (или оценивать) и само распределение случайной величины ν . В этих ситуациях весьма полезным является также следующее утверждение.

¹⁾ См., например, [22], т. 1.

Предположим, что мы находимся в условиях теоремы 3, однако вместо условий на моменты ζ_n мы будем предполагать, что ζ_n удовлетворяют условию Крамера (ср. с § 8.8):

$$\psi(\lambda) = Ee^{\lambda\zeta} < \infty$$

при некотором $\lambda \neq 0$.

Другими словами, если

$$\lambda_+ = \sup(\lambda: \psi(\lambda) < \infty) \geq 0, \quad \lambda_- = \inf(\lambda: \psi(\lambda) < \infty) \leq 0,$$

то $\lambda_+ - \lambda_- > 0$. Везде в дальнейшем мы будем рассматривать лишь значения $\lambda \in B = \{\lambda: \psi(\lambda) < \infty\} \subset [\lambda_-, \lambda_+]$, для которых $\psi(\lambda) < \infty$. Для таких λ определен положительный мартингал

$$X_n = \frac{e^{\lambda S_n}}{\psi^n(\lambda)}, \quad X_0 = 1,$$

так что $EX_n = 1$.

Теорема 4. Пусть ν — произвольный момент остановки. Тогда

$$E\left(\frac{e^{\lambda S_\nu}}{\psi(\lambda)^\nu}; \nu < \infty\right) \leq 1 \quad (27)$$

и при любых $s > 1$, $r > 1$, $1/r + 1/s = 1$

$$E(e^{\lambda S_\nu}; \nu < \infty) \leq \{E[\psi^{(r\nu)/s}(\lambda s); \nu < \infty]\}^{1/r}. \quad (28)$$

Для того, чтобы выполнялось

$$E\left(\frac{e^{\lambda S_\nu}}{\psi(\lambda)^\nu}; \nu < \infty\right) = 1, \quad (29)$$

необходимо и достаточно, чтобы

$$\lim_{n \rightarrow \infty} E\left(\frac{e^{\lambda S_n}}{\psi(\lambda)^n}; \nu > n\right) = 0. \quad (30)$$

Замечание 1. Тождество (29) носит название *фундаментального тождества Вальда*. Обычно оно доказывается в литературе для п. н. конечных ν : $P(\nu < \infty) = 1$, и в этом случае представляет собой обобщение на случайное время ν очевидного равенства $Ee^{\lambda S_n} = \psi^n(\lambda)$. Первоначально тождество (29) было установлено Вальдом для специального случая, когда ν есть время выхода последовательности $\{S_n\}$ из некоторого конечного интервала (см. следствие 3), и сопровождалось весьма ограниченными условиями. Позже эти условия были сняты (см., например, [22] т. 2). Ниже мы получим более общее утверждение для задачи о первом выходе траектории $\{S_n\}$ из области с криволинейными границами.

Замечание 2. Фундаментальное тождество Вальда показывает, что при всей общности определения момента остановки ν существует жесткая функциональная связь (29) на совместное распределение ν и S_ν (распределение ξ_k предполагается известным). В тех случаях, когда одну из этих величин удастся каким-либо образом «вычислить» или «устранить» (см. примеры 6–8), тождество Вальда превращается в явную формулу для преобразования Лапласа над распределением другой величины. Если ν и S_ν оказываются независимыми (это бывает редко), то (29) дает связь

$$Ee^{\lambda S_\nu} = [E\psi(\lambda)^{-\nu}]^{-1}$$

между преобразованиями Лапласа над распределениями ν и S_ν .

Доказательство теоремы 4. Как мы уже отмечали $\{X_n, \mathfrak{F}_n; n \geq 0\}$ при $X_n = e^{\lambda S_n} \psi^{-n}(\lambda)$, $\mathfrak{F}_n = \sigma(\zeta_1, \dots, \zeta_n)$ образует положительный мартингал, $X_0 = 1$, $E X_n = 1$. Из следствия 2 немедленно следует, что выполнено (27).

Неравенство (28) есть результат применения неравенства Гельдера и (27):

$$E(e^{\lambda S_\nu})^{1/s}; \nu < \infty) = E \left[\left(\frac{e^{\lambda S_\nu}}{\psi^\nu(\lambda)} \right)^{1/s} \psi^{\nu/s}(\lambda); \nu < \infty \right] \leq 1 \cdot [E(\psi^{(\nu r)/s}(\lambda); \nu < \infty)]^{1/r}.$$

Последнее утверждение теоремы, относящееся к тождеству (29), также следует из теоремы 2В. ■

Рассмотрим теперь несколько важных частных случаев. Заметим, что $\psi(\lambda)$ является выпуклой функцией ($\psi''(\lambda) > 0$), $\psi(0) = 1$ и, стало быть, существует единственная точка λ_0 , в которой $\psi(\lambda)$ достигает своего минимума, $\psi(\lambda_0) \leq 1$ (см. также § 8.8). Пусть далее дана последовательность $g(n)$ такая, что $g^+(n) = \max(0, g(n)) = o(n)$ при $n \rightarrow \infty$.

Следствие 3. Если на множестве $\{\nu > n\}$ выполняется $S_n \leq g(n)$, то при $\lambda \in (\lambda_0, \lambda_+] \cap B$ выполнено (29).

Очевидно, что случайная величина $\nu = \nu_g = \min\{k \geq 1: S_k > g(k)\}$ удовлетворяет условиям следствия 3. Для моментов остановки ν_g можно рассматривать также случай $g(n)/n \rightarrow c \geq 0$, который сводится к случаю $g(n) = o(n)$ введением случайных величин $\zeta_k^* = \zeta_k - c$, $S_k^* = \sum_{j=1}^k \zeta_j^*$, для которых $\nu_g = \min\{k \geq 1: S_k^* > g(k) - ck\}$.

Доказательство следствия 3. При $\lambda > \lambda_0$, $\lambda \in B$ имеем

$$\begin{aligned} E \left(\frac{e^{\lambda S_n}}{\psi^n(\lambda)}; \nu > n \right) &\leq \psi^{-n}(\lambda) E(e^{\lambda S_n}; S_n \leq g(n)) \leq \\ &\leq \psi^{-n}(\lambda) E(e^{(\lambda - \lambda_0) S_n} \cdot e^{\lambda_0 S_n}; S_n \leq g(n)) \leq \psi^{-n}(\lambda) e^{(\lambda - \lambda_0) g(n)} E(e^{\lambda_0 S_n}; S_n \leq g(n)) \leq \\ &\leq \psi^{-n}(\lambda) e^{(\lambda - \lambda_0) g^+(n)} E e^{\lambda_0 S_n} = \left(\frac{\psi(\lambda_0)}{\psi(\lambda)} \right)^n e^{(\lambda - \lambda_0) g^+(n)} \rightarrow 0 \end{aligned}$$

при $n \rightarrow \infty$, т. к. $(\lambda - \lambda_0) g^+(n) = o(n)$. Следствие доказано. ■

Вернемся теперь к теореме 4 для произвольных моментов остановки. Оказывается, что ее утверждение можно сделать более полным, опираясь на преобразование Крамера, введенное в § 8.8, и совершенно не используя при этом мартингальный подход.

Наряду с исходным распределением P последовательности $\{\zeta_k\}_{k=1}^\infty$ введем в рассмотрение семейство распределений P_λ этой последовательности, порожденное в $\langle \mathbb{R}^\infty, \mathfrak{B}^\infty \rangle$ (см. § 5.5) конечномерными распределениями

$$\begin{aligned} P_\lambda(\zeta_k \in dx_k) &= \frac{e^{\lambda x_k}}{\psi(\lambda)} P(\zeta_k \in dx_k), \\ P_\lambda(\zeta_k \in dx_1, \dots, \zeta_n \in dx_n) &= \prod_{k=1}^n P_\lambda(\zeta_k \in dx_k). \end{aligned}$$

Это есть преобразование Крамера над распределением P .

Теорема 4А. При любом $\lambda \in B$

$$E\left(\frac{e^{\lambda S_\nu}}{\psi^\nu(\lambda)}; \nu < \infty\right) = P_\lambda(\nu < \infty). \quad (31)$$

Доказательство. Так как $\{\nu = n\} \in \sigma(\zeta_1, \dots, \zeta_n)$, то в пространстве \mathbb{R}^n существует борелевское множество D_n такое, что

$$\{\nu = n\} = \{(\zeta_1, \dots, \zeta_n) \in D_n\}.$$

Далее,

$$E\left(\frac{e^{\lambda S_\nu}}{\psi^\nu(\lambda)}; \nu < \infty\right) = \sum_{n=0}^{\infty} E\left(\frac{e^{\lambda S_n}}{\psi^n(\lambda)}; \nu = n\right),$$

где

$$\begin{aligned} E\left(\frac{e^{\lambda S_n}}{\psi^n(\lambda)}; \nu = n\right) &= \int_{(x_1, \dots, x_n) \in D_n} \frac{e^{\lambda(x_1 + \dots + x_n)}}{\psi^n(\lambda)} P(\zeta_1 \in dx_1, \dots, \zeta_n \in dx_n) = \\ &= \int_{(x_1, \dots, x_n) \in D_n} P_\lambda(\zeta_1 \in dx_1, \dots, \zeta_n \in dx_n) = P_\lambda(\nu = n). \end{aligned}$$

Это доказывает теорему. ■

Рассмотрим теперь для данной функции $g(n)$ момент остановки $\nu = \nu_g = \min\{k: S_k > g(k)\}$ (ср. со следствием 3). Утверждение теоремы 4А можно дополнить в этом случае следующим образом. Через E_λ мы будем обозначать математическое ожидание относительно распределения P_λ .

Следствие 4. 1. Если $g^+(n) = \max(0, g(n)) = o(n)$, $\lambda \in (\lambda_0, \lambda_+) \cap B$, то в соотношении (31) $P_\lambda(\nu_g < \infty) = 1$.

2. Если $g(n) \geq 0$, $\lambda < \lambda_0$, то $P_\lambda(\nu_g < \infty) < 1$.

3. При $\lambda = \lambda_0$ распределение P_{λ_0} величины ν может быть как собственным ($P_{\lambda_0}(\nu_g < \infty) = 1$), так и несобственным ($P_{\lambda_0}(\nu_g < \infty) < 1$). Если $\lambda_0 \in (\lambda_-, \lambda_+)$, $g(n) < (1 - \varepsilon)\sigma\sqrt{2n \ln \ln n}$ при всех $n \geq n_0$, начиная с некоторого n_0 , $\sigma^2 = E_{\lambda_0}\zeta_1^2$, то $P_{\lambda_0}(\nu_g < \infty) = 1$.

Если же $\lambda \in (\lambda_-, \lambda_+)$, $g(n) \geq 0$, $g(n) \geq (1 + \varepsilon)\sigma\sqrt{2n \ln \ln n}$ при $n \geq n_0$, то $P_{\lambda_0}(\nu_g < \infty) < 1$ (тривиальный случай $\zeta_k \equiv 0$ исключается).

Доказательство. Так как

$$E_\lambda \zeta_k = \frac{\psi'(\lambda)}{\psi(\lambda)},$$

то $E_\lambda \zeta_k$ имеет тот же знак, что и $\lambda - \lambda_0$, $E_{\lambda_0} \zeta_k = 0$ ($\psi'(\lambda_0) = 0$, если $\lambda_0 \in (\lambda_-, \lambda_+)$). Поэтому первое утверждение следствия вытекает из закона больших чисел для сумм

$$X_n = \sum_{k=1}^n \zeta_k, \text{ поскольку } E_\lambda \zeta_k > 0,$$

$$P_\lambda(\nu = \infty) = P_\lambda(X_n < g(n) \text{ при всех } n) < P(X_n < g^+(n)) \rightarrow 0$$

при $n \rightarrow \infty$.

Второе утверждение является следствием усиленного закона больших чисел, т. к. $E_\lambda \zeta_k < 0$, $P_\lambda(\nu = \infty) = P(\sup_n X_n \leq 0) > 0$.

Последнее утверждение следствия вытекает из закона повторного логарифма, доказанного в § 19.2. ■

Условие $g(n) \geq 0$ п. 2 следствия, очевидно, можно ослабить до условия $g(n) = o(n)$, $P(\nu > n) > 0$ при любом $n > 0$. Аналогичное замечание относится к п. 3.

Утверждение, аналогичное следствию 4, справедливо и для времени $\nu_{g-,g+}$ первого прохождения хотя бы одной из двух границ $g_{\pm}(n) = o(n)$:

$$\nu_{g-,g+} = \min\{k \geq 1: S_k > g_+(k) \text{ или } S_k < g_-(k)\}.$$

Следствие 5. При $\lambda \in B \setminus \{\lambda_0\}$ выполняется $P_{\lambda}(\nu_{g-,g+} < 1) = 1$.

Если $\lambda = \lambda_0 \in (\lambda_-, \lambda_+)$, то P_{λ} — распределение ν — может быть как собственным, так и несобственным.

Если $g_{\pm}(n) \leq \pm(1 - \varepsilon)\sigma\sqrt{2 \ln \ln n}$ при $n \geq n_0$, то $P_{\lambda_0}(\nu_{g-,g+} < \infty) = 1$.

Если $g_{\pm}(n) \geq 0$ и, кроме того, $g_{\pm}(n) \geq \pm(1 + \varepsilon)\sigma\sqrt{2 \ln \ln n}$ при $n \geq n_0$, то $P_{\lambda_0}(\nu_{g-,g+} < \infty) < 1$.

Доказательство. Первое утверждение вытекает из следствия 4, примененного к последовательностям $\{\pm X_n\}$. Второе — из закона повторного логарифма в § 19.2. ■

Рассмотрим теперь ряд соотношений, вытекающих из следствий 3, 4 (из тождества (29)) для случайных величин $\nu = \nu_g$ и $\nu = \nu_{g-,g+}$.

Пусть $a < 0$, $\psi(\lambda_+) \leq 1$. Так как $\psi'(0) = a < 0$, $\psi(\lambda)$ выпукла, то в области $\lambda > 0$ уравнение $\psi(\lambda) = 1$ будет иметь единственный корень $\mu > 0$. Полагая в (29) $\lambda = \mu$, мы получим

Следствие 6. Если $a < 0$, $\psi(\lambda_+) \geq 1$, то для моментов остановки $\nu = \nu_g$ и $\nu = \nu_{g-,g+}$ справедливо равенство

$$E(e^{\mu S_{\nu}}; \nu < \infty) = 1.$$

Таким образом, в этом следствии мы «избавились» от «компоненты» $\psi^{\nu}(\lambda)$ в тождестве (29). «Избавиться» от компоненты $e^{\lambda S_{\nu}}$ можно только для блужданий специального вида. К ним относятся так называемые блуждания, непрерывные сверху (или снизу; см. § 11.8) и блуждания с экспоненциальным (или геометрическим) распределением $\zeta_k^+ = \max(0, \zeta_k)$ или $\zeta_k^- = -\min(0, \zeta_k)$. Проиллюстрируем это на примере.

Пример 7. Вернемся к задаче об игре на разорение, которая обсуждалась в примере 6. В условиях этой игры следствие 4 при $g_-(n) = -z_1$, $g_+(n) = z_2$ дает

$$e^{\lambda z_2} E(\psi(\lambda)^{-\nu}; S_{\nu} = z_2) + e^{-\lambda z_1} E(\psi(\lambda)^{-\nu}; S_{\nu} = -z_1) = 1.$$

В частности, при $z_1 = z_2 = z$ и $p = 1/2$ из соображений симметрии получаем

$$\begin{aligned} E(\psi(\lambda)^{-\nu}; S_{\nu} = z) &= \frac{1}{e^{\lambda z} + e^{-\lambda z}}, \\ E(\psi(\lambda)^{-\nu}) &= \frac{2}{e^{\lambda z} + e^{-\lambda z}}. \end{aligned} \quad (32)$$

Пусть $\lambda(s)$ — единственный в области $\lambda > 0$ корень уравнения $s\psi(\lambda) = 1$, $s \in (0, 1)$. Так как $\psi(\lambda) = 1/2(e^{\lambda} + e^{-\lambda})$, то, решая квадратное уравнение, мы находим

$$e^{\lambda(s)} = \frac{1 + \sqrt{1 - s^2}}{s}.$$

Тождество (32) дает

$$Es^\nu = 2(e^{\lambda(s)z} + e^{-\lambda(s)z}).$$

Мы получили явный вид производящей функции случайной величины ν , который позволяет путем разложений элементарных функций в ряд найти в явной форме вероятности $P(\nu = n)$, $n = 1, 2, \dots$.

Пример 8. Простые явные формулы можно получать из тождества Вальда и в задаче с одной границей, когда $\nu = \nu_g$, $g(n) = z$. При этом класс распределений ζ_k может быть более широким, чем в примере 7. Мы предположим, что выполнено одно из двух условий (ср. с § 11.8).

1. Блуждание является дискретным (целочисленным) и непрерывным сверху, т. е. ζ_k целочисленны, $P(\zeta_k = 1) > 0$, $P(\zeta_k \geq 2) = 0$.

2. Блуждание является экспоненциальным сверху, т. е.

$$P(\zeta_k > t) = ce^{-\alpha t} \quad (33)$$

либо при всех $t > 0$, либо при $t = 0, 1, 2, \dots$, если блуждание целочисленно (геометрическое распределение).

Величина ν_g будет собственной тогда и только тогда, когда $E\zeta_k = \psi'(0) \geq 0$ (см. гл. 9–11). Для блужданий, непрерывных сверху, тождество Вальда (29) дает $(g(n) = z > 0, S_\nu = z)$

$$e^{\lambda z} E\psi^{-\nu}(\lambda) = 1, \quad \lambda > \lambda_0. \quad (34)$$

Уравнение $\psi(\lambda) = s^{-1}$ (ср. с примером 7) при $s \leq 1$ имеет в области $\lambda > \lambda_0$ единственный корень $\lambda(s)$. Поэтому тождество (34) можно записать в виде

$$Es^\nu = e^{-z\lambda(s)}. \quad (35)$$

Из этого утверждения следует ряд результатов глав 8, 11. Из него можно найти многие свойства распределения ν .

Аналогичные соотношения можно получить для блужданий, экспоненциальных сверху. Пусть, например, (33) справедливо при всех $t > 0$. В этом случае условное распределение $P(S_\nu > t | \nu = n, S_{n-1} = x)$ совпадает с распределением

$$P(\zeta_n > z - x + t | \zeta_n > z - x) = e^{-\alpha t},$$

и, как видно, не зависит от n и x . Это означает, что ν и S_ν независимы, $S_\nu = z + \gamma$, $\gamma \in \Gamma_\alpha$;

$$E\psi(\lambda)^{-\nu} = [Ee^{(z+\gamma)\lambda}]^{-1} = e^{-\lambda z \frac{\alpha - \lambda}{\alpha}}, \quad \lambda_0 < \lambda < \alpha;$$

$$Es^\nu = e^{-z\lambda(s) \frac{\alpha - \lambda(s)}{\alpha}},$$

где $\lambda(s)$, как и прежде, единственное в области $\lambda > \lambda_0$ решение уравнения $\psi(\lambda) = s^{-1}$. Отсюда можно получать все те же результаты, что и из (35).

Если $P(\zeta_k > t) = c_1 e^{-\alpha t}$, $P(\zeta_k < -t) = c_2 e^{-\beta t}$, $t > 0$, то в задаче с двумя границами для $\nu = \nu_{g_-, g_+}$, $g_+(n) = z_2$, $g_-(n) = -z_1$ совершенно аналогично получаем из (29)

$$\frac{\alpha e^{\lambda z_2}}{\alpha - \lambda} E(\psi^{-\nu}(\lambda); S_\nu \geq z_2) + \frac{\beta e^{-\lambda z_1}}{\beta + \lambda} E(\psi^{-\nu}(\lambda); S_\nu \leq -z_1) = 1, \quad \lambda \in (-\beta, \alpha).$$

§ 3. Неравенства

Отметим прежде всего, что свойство $E X_n \leq 1$ последовательности $X_n = \frac{e^{\lambda S_n}}{\psi_0(\lambda)^n}$, образующей при подходящей функции $\psi_0(\lambda)$ супермартингал, сохраняется при замене n на момент остановки ν (аналог неравенства (27)) и в значительно более общей, чем в теореме 4, ситуации, когда ζ_k могут быть зависимы.

Пусть, как и прежде, \mathfrak{F}_n — последовательность вложенных σ -алгебр, ζ_n измеримы относительно \mathfrak{F}_n . Предположим, что п. в.

$$E(e^{\lambda \zeta_n} | \mathfrak{F}_{n-1}) \leq \psi_0(\lambda). \quad (36)$$

Это условие всегда выполнено, если п. в.

$$P(\zeta_n \geq x | \mathfrak{F}_{n-1}) \leq G(x), \quad \psi_0(\lambda) = - \int e^{\lambda x} dG(x) < \infty.$$

В этом случае последовательность $X_n = e^{\lambda S_n} \psi_0^{-n}(\lambda)$ образует супермартингал:

$$E(X_n | \mathfrak{F}_{n-1}) \leq X_{n-1}, \quad E X_n \leq 1.$$

Теорема 4В. Пусть выполнено (36), ν — момент остановки. Тогда справедливы неравенства (27), (28) при замене ψ на ψ_0 .

Доказательство теоремы почти дословно повторяет рассуждения в теореме 4. ■

Получим теперь неравенства для распределений

$$\bar{X}_n = \max_{k \leq n} X_k \quad \text{и} \quad X_n^* = \max_{k \leq n} |X_k|$$

для произвольного субмартингала $\{X_n\}$.

Теорема 5 (Дуб). Пусть $\{X_n, \mathfrak{F}_n; n \geq 0\}$ — неотрицательный субмартингал. Тогда при всех $x \geq 0$, $n \geq 0$

$$P(\bar{X}_n > x) \leq E \frac{X_n}{x}.$$

Доказательство. Пусть

$$\nu = \eta(x) = \min\{k \geq 0: X_k > x\}, \quad \nu(n) = \min(\nu, n).$$

Очевидно, что n и $\nu(n)$ есть моменты остановки, $\nu(n) \leq n$, и поэтому по теореме 1 (см. (6) при $\nu_2 = n$, $\nu_1 = \nu(n)$)

$$E X_n \geq E X_{\nu(n)}.$$

Замечая, что $\{\bar{X}_n > x\} = \{X_{\nu(n)} > x\}$, по неравенству Чебышева получим

$$P(\bar{X}_n > x) = P(X_{\nu(n)} > x) \leq \frac{E X_{\nu(n)}}{x} \leq \frac{E X_n}{x}. \quad \blacksquare$$

Из теоремы 5 вытекает

Теорема 6 (второе неравенство Колмогорова). Пусть $\{\mathfrak{F}_n; n \geq 0\}$ — мартингал, для которого существует $E X_n^2$. Тогда $\{X_n^2, \mathfrak{F}_n; n \geq 0\}$ — субмартингал, и из теоремы 5 получаем

$$P(X_n^* > x) \leq E \frac{X_n^2}{x^2}.$$

Первоначально это неравенство было получено Колмогоровым для сумм $X_n = \xi_1 + \dots + \xi_n$ независимых случайных величин ξ_n . Теорема 6 обобщает доказательство Колмогорова на случай субмартингалов и усиливает неравенство Чебышева.

Справедливо также следующее обобщение теоремы 6.

Теорема 6А. Если $\{X_n, \mathfrak{F}_n; n \geq 0\}$ — мартингал, $E|X_n|^p < \infty$, $p \geq 1$, то $\{|X_n|^p, \mathfrak{F}_n; n \geq 0\}$ образует неотрицательный субмартингал и при всех $x > 0$

$$P(X_n^* \geq x) \leq \frac{E|X_n|^p}{x^p}.$$

Если $\{X_n, \mathfrak{F}_n; n \geq 0\}$ — субмартингал, $Ee^{\lambda X_n} < \infty$, $\lambda > 0$, то $\{e^{\lambda X_n}, \mathfrak{F}_n\}$ также образует неотрицательный субмартингал,

$$P(\bar{X}_k \geq x) \leq e^{-\lambda x} Ee^{\lambda X_n}.$$

Теорема 6А, как и теорема 6, немедленно следует из леммы 5 и теоремы 5.

Если $X_n = S_n = \sum_{k=1}^n \zeta_k$, где ζ_k независимы, одинаково распределены и удовлетворяют условию Крамера: $\lambda_+ = \sup\{\lambda: \psi(\lambda) < \infty\} > 0$, то для $P(\bar{X}_n > x)$ в случае $a = E\zeta_k < 0$ можно с помощью фундаментального тождества Вальда получить более точные неравенства.

Напомним, что функция $\psi(\lambda) = Ee^{\lambda \zeta_k}$ в случае $a = \psi'(0) < 0$ убывает в окрестности точки $\lambda = 0$ и при $\psi(\lambda_+) \geq 1$ имеет в области $\lambda > 0$ единственный корень μ уравнения $\psi(\lambda) = 1$.

Пусть ζ — случайная величина с тем же распределением, что и ζ_k . Обозначим

$$\psi_+ = \sup_{t>0} E(e^{\mu(\zeta-t)} | \zeta > t), \quad \psi_- = \inf_{t>0} E(e^{\mu(\zeta-t)} | \zeta > t).$$

Если

$$P(\zeta > t) = ce^{-\alpha t} \quad \text{при } t > 0$$

(здесь в (33) с необходимостью $\alpha > \mu$), то

$$P(\zeta - t > v | \zeta > t) = \frac{P(\zeta > t + v)}{P(\zeta > t)} = e^{-\alpha v},$$

$$\psi_+ = \psi_- = \frac{\alpha}{\alpha - \mu}.$$

Аналогичное равенство будет иметь место для целочисленных ξ с геометрическим распределением.

Для других распределений будет выполняться $\psi_+ > \psi_-$.

При выполнении названных выше условий для распределения случайной величины $S = \sup_k S_k$ справедливо следующее утверждение.

Теорема 6В.

$$\psi_+^{-1} e^{-\mu x} \leq P(S > x) \leq \psi_-^{-1} e^{-\mu x}. \quad (37)$$

Из теоремы следует, что для экспоненциальных правых частей распределения ζ (см. (33)) неравенства (37) переходят в точное равенство

$$P(S > x) = \frac{\alpha - \mu}{\alpha} e^{-\mu x}.$$

(Такой же результат мы получили в § 11.5.) Это говорит о том, что неравенства (37) неулучшаемы. Так как $\bar{S}_n = \max_{k \leq n} S_k \leq S$, то из них следует, что при любом n

$$P(\bar{S}_n > x) \leq \psi_-^{-1} e^{-\mu x}.$$

Доказательство теоремы 6В. Положим $\nu = \infty$, если $S = \sup_{k \geq 0} S_k \leq x$ и $\nu = \eta(x) = \min\{k: S_k > x\}$ в противном случае. Пусть далее $\chi(x) = S_{\eta(x)} - x$ есть величина эксцесса уровня x . Имеем

$$\begin{aligned} P(\chi(x) > v; \nu < \infty) &= \sum_{k=1}^{\infty} \int_{-\infty}^x P(\bar{S}_{k-1} \leq x, S_{k-1} \in du, \zeta_k > x - u + v) = \\ &= \sum_{k=1}^{\infty} \int_{-\infty}^x P(\bar{S}_{k-1} \leq x, S_{k-1} \in du, \zeta_k > x - u) P(\zeta_k > x - u + v | \zeta_k > x - u), \\ E(e^{\mu \chi(x)}; \nu < \infty) &\leq \sum_{k=1}^{\infty} \int_{-\infty}^x P(\bar{S}_{k-1} \leq x, S_{k-1} \in du, \zeta_k > x - u) \psi_+ = \\ &= \psi_+ \sum_{k=1}^{\infty} P(\nu = k) = \psi_+ P(\nu < \infty). \end{aligned}$$

Аналогично получаем

$$E(e^{\mu \chi(x)}; \nu < \infty) \geq \psi_- P(\nu < \infty).$$

Далее, в силу следствия 5 имеем

$$1 = E(e^{\mu S_\nu}; \nu < \infty) = e^{\mu x} E(e^{\mu \chi(x)}; \nu < \infty) \leq e^{\mu x} \psi_+ P(\nu < \infty).$$

Так как $P(\nu < \infty) = P(S > x)$, то отсюда получаем правое неравенство теоремы 6В. Точно так же получается левое неравенство. ■

Замечание 3. Мы доказали теорему 6В с помощью фундаментального тождества Вальда. Но ее можно доказывать и непосредственно с помощью следующих соотношений

$$\psi^n(\lambda) = E e^{\lambda S_n} \geq \sum_{k=1}^n E(e^{(S_k + S_n - S_k)\lambda}; \nu = k) = \sum_{k=1}^n E(e^{(x + \chi(x))\lambda} e^{(S_n - S_k)\lambda}; \nu = k). \quad (38)$$

Здесь случайные величины $e^{\lambda \chi(x)} I(\nu = k)$ и $S_n - S_k$ независимы, при этом, так же, как и прежде

$$E(e^{\lambda \chi(x)}; \nu = k) \geq \psi_- P(\nu = k).$$

Поэтому при λ таких, что $\psi(\lambda) \leq 1$

$$\psi^n(\lambda) \geq e^{\lambda x} \psi_- \sum_{k=1}^n \psi^{n-k}(\lambda) P(\nu = k) \geq \psi_- e^{\lambda x} \psi^n(\lambda) P(\nu \leq n).$$

Отсюда получаем

$$P(\bar{S}_n > x) = P(\nu \leq n) \leq \psi_-^{-1} e^{-\lambda x}.$$

Так как правая часть от n не зависит, то это неравенство справедливо и для $P(S > x)$. Оценка снизу получается аналогично. Надо лишь в исходном неравенстве (ср. с (38))

$$\psi^n(\lambda) = \sum_{k=1}^n E(e^{\lambda S_n}; \nu = k) + E(e^{\lambda S_n}; \nu > n)$$

доказать соотношения $E(e^{\lambda S_n}; \nu > n) = o(1)$ при $\lambda = \mu$, что мы проделывали в § 2.

Вернемся теперь к произвольным субмартингалам X_n и докажем еще одно неравенство, которое понадобится для теорем сходимости в следующем параграфе. Оно касается числа пересечений последовательностью X_n некоторой полосы. Пусть $a < b$ — два заданных числа. Положим $\nu_0 = 0$,

$$\nu_1 = \min\{n > 0: X_n \leq a\}, \quad \nu_2 = \min\{n > \nu_1: X_n \geq b\},$$

$$\dots \dots \dots \nu_{2k-1} = \min\{n > \nu_{2k-2}: X_n \leq a\}, \quad \nu_{2k} = \min\{n > \nu_{2k-1}: X_n \geq b\}.$$

Будем считать, что $\nu_m = \infty$, если траектория $\{X_n\}$ при $n > \nu_{m-1}$ никогда не пересечет соответствующий уровень. В этих обозначениях можно определить число пересечений снизу вверх траекторией X_0, \dots, X_n полосы $[a, b]$ как случайную величину

$$\nu(a, b; n) = \begin{cases} \max\{k: \nu_{2k} \leq n\}, & \text{если } \nu_2 \leq n, \\ 0, & \text{если } \nu_2 > n. \end{cases}$$

Обозначим $(a)^+ = \max(0, a)$.

Теорема 7 (Дуб). Пусть $\{X_n, \mathfrak{F}_n; n \geq 0\}$ — субмартингал. Тогда при всех n

$$E\nu(a, b; n) \leq \frac{E(X_n - a)^+}{b - a}. \quad (39)$$

Ясно, что само по себе неравенство (39) предполагает заданным лишь субмартингал $\{X_k, \mathfrak{F}_k; 0 \leq k \leq n\}$.

Доказательство. $\nu(a, b; n)$ совпадает с числом пересечений полосы $[0, b - a]$ последовательностью $(X_n - a)^+$. Но $\{(X_n - a)^+, \mathfrak{F}_n; n \geq 0\}$ образует неотрицательный субмартингал (см. пример 4) и, следовательно, можно, не ограничивая общности, считать $a = 0$, $X_n \geq 0$ и доказывать, что

$$E\nu(0, b; n) \leq \frac{EX_n}{b}.$$

Пусть

$$\eta_j = \begin{cases} 1, & \text{если } \nu_k < j \leq \nu_{k+1} \text{ для какого-нибудь нечетного } k, \\ 0, & \text{если } \nu_k < j \leq \nu_{k+1} \text{ для какого-нибудь четного } k. \end{cases}$$

На рис. 14 $\nu_1 = 2$, $\nu_2 = 5$, $\nu_3 = 8$; $\eta_j = 0$ при $j \leq 2$, $\eta_j = 1$ при $3 \leq j \leq 5$ и т. д. Нетрудно видеть (мы используем здесь преобразование Абеля), что $(X_0 = 0, \eta_0 = 1)$

$$\eta_0 X_0 + \sum_{j=1}^n \eta_j (X_j - X_{j-1}) = \sum_{j=0}^{n-1} X_j (\eta_j - \eta_{j+1}) + \eta_n X_n \geq b\nu(0, b; n).$$

Кроме того (\mathcal{N}_1 означает множество нечетных чисел),

$$\{\eta_j = 1\} = \bigcup_{k \in \mathcal{N}_1} \{\nu_k < j \leq \nu_{k+1}\} = \bigcup_{k \in \mathcal{N}_1} [\{\nu_k \leq j - 1\} - \{\nu_{k+1} \leq j - 1\}] \in \mathfrak{F}_{j-1}.$$

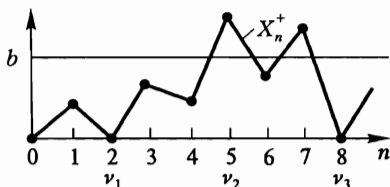


Рис. 14.

Поэтому и в силу того, что $E(X_j | \mathfrak{F}_{j-1}) - X_{j-1} \geq 0$, получим

$$\begin{aligned} bE\nu(0, b; n) &\leq E \sum_{j=1}^n \eta_j (X_j - X_{j-1}) = \sum_{j=1}^n E(X_j - X_{j-1}; \eta_j = 1) = \\ &= \sum_{j=1}^n E[E(X_j - X_{j-1} | \mathfrak{F}_{j-1}); \eta_j = 1] = \sum_{j=1}^n E[E(X_j | \mathfrak{F}_{j-1}) - X_{j-1}; \eta_j = 1] \leq \\ &\leq \sum_{j=1}^n E[E(X_j | \mathfrak{F}_{j-1}) - X_{j-1}] = \sum_{j=1}^n E(X_j - X_{j-1}) = EX_n. \end{aligned}$$

§ 4. Теоремы сходимости

Теорема 8 (теорема Дуба о сходимости мартингалов). Пусть $\{X_n, \mathfrak{F}_n; -\infty < n < \infty\}$ — субмартингал. Тогда

1) $X_{-\infty} = \lim_{n \rightarrow -\infty} X_n$ существует п.н., $EX_{-\infty}^+ < \infty$ и $\{X_n, \mathfrak{F}_n; -\infty \leq n < \infty\}$ образует субмартингал.

2) Если $\sup_n EX_n^+ < \infty$, то $X_{\infty} = \lim_{n \rightarrow \infty} X_n$ существует п.н., $EX_{\infty}^+ < \infty$. Если к тому же $\sup_n E|X_n| < \infty$, то $E|X_{\infty}| < \infty$.

3) Последовательность $\{X_n, \mathfrak{F}_n; -\infty \leq n \leq \infty\}$ образует субмартингал тогда и только тогда, когда последовательность $\{X_n^+\}$ равномерно интегрируема.

Доказательство. Так как

$$\{\overline{\lim} X_n > \underline{\lim} X_n\} = \bigcup_{\substack{a, b \\ \text{рациональные}}} \{\overline{\lim} X_n > b > a > \underline{\lim} X_n\}$$

(пределы берутся при $n \rightarrow \infty$), то предположение о расходимости с положительной вероятностью

$$P(\overline{\lim} X_n > \underline{\lim} X_n) > 0$$

означает, что найдутся рациональные $a < b$ такие, что

$$P(\overline{\lim} X_n > b > a > \underline{\lim} X_n) > 0. \quad (40)$$

Пусть $\nu(a, b; m)$ — число пересечений снизу вверх полосы $[a, b]$ последовательностью $Y_1 = X_{-m}, \dots, Y_m = X_{-1}$, $\nu(a, b) = \lim_{m \rightarrow \infty} \nu(a, b; m)$. Тогда (40) означает, что

$$P(\nu(a, b) = \infty) > 0 \quad (41)$$

По теореме 7 (примененной к последовательности Y_1, \dots, Y_m)

$$E\nu(a, b; m) \leq \frac{E(X_{-1} - a)^+}{b - a} \leq \frac{EX_{-1}^+ + |a|}{b - a}, \quad (42)$$

$$E\nu(a, b) \leq \frac{EX_{-1}^+ + |a|}{b - a}. \quad (43)$$

Но последнее неравенство противоречит (41) и доказывает, что $P(\overline{\lim} X_n = \underline{\lim} X_n) = 1$. Кроме того, по лемме Фату ($X_{-\infty}^+ = \underline{\lim} X_n^+$)

$$EX_{-\infty}^+ \leq \underline{\lim} EX_n^+ \leq EX_{-1}^+ < \infty. \quad (44)$$

Среднее неравенство здесь вытекает из того, что $\{X_n^+, \mathfrak{F}_n\}$ образует вместе с $\{X_n, \mathfrak{F}_n\}$ субмартингал (см. лемму 3) и, стало быть, $EX_n^+ \uparrow$.

Чтобы доказать, что $\{X_n, \mathfrak{F}_n; -\infty \leq n < \infty\}$ — субмартингал, нам надо по лемме 2 убедиться, что для любого $A \in \mathfrak{F}_{-\infty} \subset \mathfrak{F}$

$$E(X_{-\infty}; A) \leq E(X_n; A). \quad (45)$$

Положим $X_n(a) = \max(X_n, a)$. По лемме 4 $\{X_n(a), \mathfrak{F}_n; n \leq 0\}$ есть равномерно интегрируемый субмартингал. Поэтому при любом k , $-\infty < k < n$,

$$\begin{aligned} E(X_k(a); A) &\leq E(X_n(a); A), \\ E(X_{-\infty}(a); A) &= \lim_{k \rightarrow -\infty} E(X_k(a); A) \leq E(X_n(a); A). \end{aligned} \quad (46)$$

Устремляя $a \rightarrow -\infty$, по теореме о монотонной сходимости получим (45).

Второе утверждение теоремы доказывается точно так же. Надо лишь в (42), (43) правые части заменить на EX_n^+ и $\sup EX_n^+$ соответственно. Вместо (44) получим (здесь пределы берутся при $n \rightarrow \infty$)

$$EX_{\infty}^+ \leq \underline{\lim} EX_n^+ < \infty,$$

и если $\sup_n E|X_n| < \infty$, то

$$E|X_{\infty}| \leq \underline{\lim} E|X_n| < \infty.$$

Последнее утверждение теоремы доказывается точно так же, как первое (равномерная интегрируемость позволяет нам, наряду с (46), записать для любого $A \in \mathfrak{F}_n$

$$E(X_{\infty}(a); A) = \lim_{k \rightarrow \infty} E(X_k(a); A) \geq E(X_n(a); A).$$

В обратную сторону третье утверждение теоремы следует из леммы 4. ■

Получим теперь некоторые следствия теоремы 8.

До сих пор (см. § 4.8) мы при изучении сходимости у. м. о. имели дело с у. м. о. вида $E(X_n | \mathfrak{F})$. Теперь из теоремы 8 можно получить полезную теорему сходимости для у. м. о. иного вида.

Теорема 9 (Леви). Пусть на вероятностном пространстве $(\Omega, \mathfrak{F}, P)$ задано неубывающее семейство $\mathfrak{F}_1 \subseteq \mathfrak{F}_2 \subseteq \dots \subseteq \mathfrak{F}$ σ -алгебр и случайная величина ξ , $E|\xi| < \infty$. Пусть, как и прежде, $\mathfrak{F}_{\infty} = \sigma\left(\bigcup_n \mathfrak{F}_n\right)$ есть σ -алгебра, порожденная событиями из $\mathfrak{F}_1, \mathfrak{F}_2, \dots$. Тогда при $n \rightarrow \infty$

$$E(\xi | \mathfrak{F}_n) \xrightarrow{\text{п. н.}} E(\xi | \mathfrak{F}_{\infty}). \quad (47)$$

Доказательство. Положим $X_n = E(\xi|\mathfrak{F}_n)$. Мы уже знаем (см. пример 3), что $\{X_n, \mathfrak{F}_n; 1 \leq n \leq \infty\}$ есть мартингал и, следовательно, по теореме 8 п. в. существует $\lim_{n \rightarrow \infty} X_n = X_{(\infty)}$. Остается доказать, что $X_{(\infty)} = E(\xi|\mathfrak{F}_{\infty})$ (т.е. что $X_{(\infty)} = X_{\infty}$). Так как $\{X_n, \mathfrak{F}_n; 1 \leq n \leq \infty\}$ по лемме 4 является равномерно интегрируемым мартингалом, то

$$E(X_{(\infty)}; A) = \lim_{n \rightarrow \infty} E(X_n; A) = \lim_{n \rightarrow \infty} E(E(\xi|\mathfrak{F}_n); A) = E(\xi; A)$$

при $A \in \mathfrak{F}_k$ и при любом $k = 1, 2, \dots$. Это означает, что крайние члены последнего равенства, являющиеся конечными мерами, совпадают на алгебре $\bigcup_{n=1}^{\infty} \mathfrak{F}_n$. По теореме о единственности продолжения меры (Приложение 1) они будут совпадать при всех $A \in \sigma\left(\bigcup_{n=1}^{\infty} \mathfrak{F}_n\right) = \mathfrak{F}_{\infty}$. Следовательно, по определению у. м. о.

$$X_{(\infty)} = E(\xi|\mathfrak{F}_{\infty}) = X_{\infty}. \quad \blacksquare$$

Можно отметить также, что из равномерной интегрируемости $\{X_n, \mathfrak{F}_n; 1 \leq n \leq \infty\}$ следует, что в (47) можно $\xrightarrow{\text{п. н.}}$ заменить на $\xrightarrow{(1)}$.

Из теоремы 8 вытекает усиленный закон больших чисел. Действительно, обратимся к примеру 4. По теореме 8 существует предел $X_{-\infty} = \lim_{n \rightarrow -\infty} X_n = \lim_{n \rightarrow \infty} (S_n/n)$, который измерим относительно остаточной (тривиальной) σ -алгебры и, следовательно, с вероятностью 1 постоянен. Так как $EX_{-\infty} = E\xi_1$, то $(S_n/n) \xrightarrow{\text{п. н.}} E\xi_1$. \blacksquare

Можно получить также некоторые обобщения теорем о сходимости рядов в гл. 10 на случай зависимых слагаемых. Пусть $X_n = S_n = \sum_{k=1}^n \xi_k$, и X_n образуют субмартингал ($E(\xi_{n+1}|\mathfrak{F}_n) \geq 0$). Пусть, кроме того, $E|X_n| < c$ при всех n при некотором $c < \infty$. Тогда существует п. н. предел $S_{\infty} = \lim_{n \rightarrow \infty} S_n$. (Как и теорема 8, это есть некоторое обобщение теоремы о сходимости монотонной последовательности. Важнейшую роль при этом играет условие ограниченности $E|X_n|$.) В частности, если ξ_k независимы, $E\xi_k = 0$, и ξ_k имеют дисперсии σ_k^2 такие, что $\sum_{k=1}^{\infty} \sigma_k^2 < \sigma^2 < \infty$, то

$$E|X_n| \leq (EX_n^2)^{1/2} \leq \left(\sum_{k=1}^n \sigma_k^2\right)^{1/2} \leq \sigma < \infty,$$

и, следовательно, $S_n \xrightarrow{\text{п. н.}} S_{\infty}$. Мы получили в качестве следствия теорему Колмогорова о сходимости рядов.

Пример 8. Рассмотрим ветвящийся процесс $\{Z_n\}$ (см. § 7.7). Мы знаем, что Z_n допускает представление

$$Z_n = \zeta_1 + \dots + \zeta_{Z_{n-1}},$$

где ζ_k — одинаково распределенные целочисленные случайные величины, не зависящие между собой и от Z_{n-1} и представляющие собой числа потомков k -й частицы в $(n-1)$ -м поколении. Если принять $Z_0 = 1$ и обозначить $\mu = E\zeta_k$, то

$$E(Z_n|Z_{n-1}) = \mu Z_{n-1}, \quad EZ_n = \mu EZ_{n-1} = \mu^n.$$

Отсюда следует, что $X_n = Z_n/\mu^n$ образует мартингал, так как

$$E(X_n|X_{n-1}) = \mu^{1-n} Z_{n-1} = X_{n-1}.$$

Теорема 10. Последовательность $X_n = \mu^{-n} Z_n$ сходится почти наверное к собственной случайной величине X , $EX < \infty$. Х.ф. $\varphi(\lambda)$ случайной величины X удовлетворяет уравнению

$$\varphi(\mu\lambda) = f(\varphi(\lambda)),$$

где $f(v) = Ev^{\zeta_k}$.

Теорема 10 означает, что $\mu^{-n} Z_n$ имеет при $n \rightarrow \infty$ собственное предельное распределение.

Доказательство. Так как $X_n \geq 0$, $EX_n = 1$, то первое утверждение немедленно вытекает из теоремы 8.

Так как $E v^{Z_n}$ равно n -й итерации функции $f(v)$, то для х.ф. получим $(\varphi_\eta(\lambda) = E e^{i\lambda\eta})$

$$\varphi_{Z_n}(\lambda) = f(\varphi_{Z_{n-1}}(\lambda)),$$

$$\varphi_{X_n}(\lambda) = \varphi_{Z_n}(\mu^{-n}\lambda) = f(\varphi_{Z_{n-1}}(\mu^{-n}\lambda)) = f\left(\varphi_{X_{n-1}}\left(\frac{\lambda}{\mu}\right)\right).$$

Поскольку $X_n \Rightarrow X$ и функция f непрерывна, то отсюда получаем уравнение для х.ф. предельного распределения X :

$$\varphi(\lambda) = f\left(\varphi\left(\frac{\lambda}{\mu}\right)\right).$$

В § 7.7 было установлено, что в случае $\mu \leq 1$ процесс Z_n вырождается с вероятностью 1 и, следовательно, $P(X = 0) = 1$. Убедимся теперь, что в случае $\mu > 1$ распределение X не вырождено (не сосредоточено в нуле). Для этого достаточно доказать, что $\{X_n, 0 \leq n \leq \infty\}$ образует мартингал и что, следовательно, $EX = EX_n \neq 0$. По теореме 8 для этого достаточно убедиться, что последовательность X_n равномерно интегрируема. Чтобы упростить рассуждения, предположим, что $D\zeta_k = \sigma^2 < \infty$ и покажем, что в этом случае $EX_n^2 < c < \infty$ (это, конечно, влечет за собой требуемую равномерную интегрируемость X_n , см. § 6.1). Непосредственно проверяется, что справедливо тождество

$$Z_n^2 - \mu^{2n} = \sum_{k=1}^n [Z_k^2 - (\mu Z_{k-1})^2] \mu^{2n-2k}.$$

Так как

$$E[Z_k^2 - (\mu Z_{k-1})^2 | Z_{k-1}] = \sigma^2 Z_{k-1}$$

(надо вспомнить равенство $D\eta = E(\eta^2 - (E\eta)^2)$), то

$$DZ_n = E(Z_n^2 - \mu^{2n}) = \sum_{k=1}^n \mu^{2n-2k} \sigma^2 E Z_{k-1} = \mu^{2n} \sigma^2 \sum_{k=1}^n \mu^{-k-1} = \frac{\sigma^2 \mu^n (\mu^n - 1)}{\mu(\mu - 1)},$$

$$EX_n^2 = \mu^{-2n} E Z_n^2 = 1 + \frac{\sigma^2 (1 - \mu^{-n})}{\mu(\mu - 1)} \leq 1 + \frac{\sigma^2}{\mu(\mu - 1)}.$$

Мы доказали таким образом, что X — невырожденная случайная величина, $EX = 1$, $DX_n \rightarrow \sigma^2/(\mu(\mu - 1))$. Из последнего соотношения нетрудно получить, что $DX = \sigma^2/(\mu(\mu - 1))$. Установить это можно, доказав, например, фундаментальность X_n второго порядка и тем самым (см. теорему 3 § 6.1) сходимость $X_n \xrightarrow{(2)} X$.

§ 5. Ограниченность моментов стохастических последовательностей

Значительный интерес представляют условия, обеспечивающие ограниченность моментов стохастических последовательностей $\{X_n, \mathfrak{F}_n\}$, которые, в свою очередь, обеспечивают стохастическую ограниченность X_n , то есть свойство $\sup_n P(X_n > N) \rightarrow 0$ при $N \rightarrow \infty$.

Для простоты мы ограничимся рассмотрением неотрицательных последовательностей $X_n \geq 0$. Разумеется, мы имели бы более полное описание асимптотического поведения X_n при $n \rightarrow \infty$, если бы могли доказать сходимость распределений X_n к предельному, как это имело место для цепей Маркова или для субмартингалов в теореме 8. Однако, такая сходимость требует, чтобы последовательность X_n удовлетворяла более жестким условиям, чем те, которые будут использованы ниже (сравним также с главой 15. Отметим также, что ограниченность $E X_n$ входит в качестве основного условия сходимости субмартингалов).

Основными и весьма естественными элементами условий ограниченности, рассматриваемых ниже, будут отрицательность «сноса» $E(\xi_n | \mathfrak{F}_{n-1})$ в области $X_{n-1} > N$ при достаточно больших N (ср. со следствием 12.3; в противном случае траектория X_n может уходить на ∞) и ограниченность моментов $\xi_n = X_n - X_{n-1}$ соответствующих порядков.

Мы начнем с экспоненциальных моментов. Простейшими условиями, обеспечивающими ограниченность $\sup_n E e^{\lambda X_n}$ при некотором $\lambda > 0$ являются следующие: при всех $n \geq 1$ и некоторых $\lambda > 0$, $N < \infty$

$$E(e^{\lambda \xi_n} | \mathfrak{F}_{n-1}) I(X_{n-1} > N) \leq \varphi(\lambda) < 1, \quad (47)$$

$$E(e^{\lambda \xi_n} | \mathfrak{F}_{n-1}) I(X_{n-1} \leq N) \leq \psi(\lambda) < \infty. \quad (48)$$

Теорема 11. При выполнении (47), (48)

$$E(e^{\lambda X_n} | \mathfrak{F}_0) \leq \varphi(\lambda) e^{\lambda X_0} + \frac{\psi(\lambda) e^{\lambda N}}{1 - \varphi(\lambda)}. \quad (49)$$

Доказательство. Обозначим через A_n левую часть в (49). Тогда в силу (47), (48) получим

$$\begin{aligned} A_n &= E \{ E [e^{\lambda X_n} (I(X_{n-1} > N) + I(X_{n-1} \leq N)) | \mathfrak{F}_{n-1}] | \mathfrak{F}_0 \} \leq \\ &\leq E [e^{\lambda X_{n-1}} (\varphi(\lambda) I(X_{n-1} > N) + \psi(\lambda) I(X_{n-1} \leq N)) | \mathfrak{F}_0] \leq \varphi(\lambda) A_{n-1} + e^{\lambda N} \psi(\lambda). \end{aligned}$$

Отсюда непосредственно находим

$$A_n \leq A_0 \varphi^n(\lambda) + e^{\lambda N} \psi(\lambda) \sum_{k=0}^{n-1} \varphi^k(\lambda) \leq A_0 \varphi^n(\lambda) + \frac{e^{\lambda N} \psi(\lambda)}{1 - \varphi(\lambda)}.$$

Достаточными для выполнения (47), (48) являются условия:

$$E(\xi_n | \mathfrak{F}_{n-1}) \leq -\varepsilon < 0 \quad \text{на } \omega\text{-множестве } \{X_{n-1} > N\}, \quad (50)$$

$$E(e^{|\lambda \xi_n|} | \mathfrak{F}_{n-1}) \leq \psi_1(\lambda) < \infty \quad (51)$$

при некотором $\lambda > 0$. ■

Первое условие означает, что $Y_n = (X_n + \varepsilon n)I(X_{n-1} > N)$ образует супермартингал.

Докажем достаточность (50), (51). Воспользуемся неравенством

$$e^x \leq 1 + x + \frac{x^2}{2} e^{|x|},$$

которое следует из сравнения разложений в ряд правой и левой частей:

$$1 + x + \sum_{k=0}^{\infty} \frac{x^{k+2}}{(k+2)!} \leq 1 + x + \sum_{k=0}^{\infty} \frac{|x|^{k+2}}{2k!}.$$

Тогда на множестве $\{X_{n-1} > N\}$ будет выполняться

$$E(e^{\lambda \xi_n} | \mathfrak{F}_{n-1}) \leq 1 - \lambda \varepsilon + \frac{\lambda^2}{2} E(\xi_n^2 e^{\lambda |\xi_n|} | \mathfrak{F}_{n-1}).$$

Так как $x^2 < e^{\lambda x/2}$ при достаточно больших x , то с помощью неравенства Гельдера убеждаемся, что наряду с (51) будет выполняться также

$$E(\xi_n^2 e^{(\lambda |\xi_n|)/2} | \mathfrak{F}_{n-1}) \leq \psi_2(\lambda) < \infty.$$

Отсюда следует, что при достаточно малом λ на множестве $\{X_{n-1} > N\}$

$$E(e^{\lambda \xi_n} | \mathfrak{F}_{n-1}) \leq 1 - \lambda \varepsilon + \frac{\lambda^2}{2} \psi_2(\lambda) \equiv \varphi(\lambda) \leq 1 - \frac{\lambda \varepsilon}{2} < 1.$$

Следствие 6. Если в дополнение к условиям теоремы 11 распределение X_n сходится к предельному распределению: $P(X_n < t) \Rightarrow P(X < t)$, то

$$E e^{\lambda X} \leq \frac{e^{\lambda N} \psi(\lambda)}{1 - \varphi(\lambda)}.$$

Следствие вытекает из теоремы Фату—Лебега (см. также лемму 6.1):

$$E e^{\lambda X} \leq \liminf_{n \rightarrow \infty} E e^{\lambda X_n}.$$

Получим теперь оценки для обычных моментов. Обозначим $M^l(n) = E X_n^l$,

$$m(0) = 1, \quad m(1) = \sup_{n \geq 1} \sup_{\omega \in I(X_{n-1} > N)} E(\xi_n | \mathfrak{F}_{n-1}),$$

$$m(l) = \sup_{n \geq 1} \sup_{\omega} E(|\xi_n|^l | \mathfrak{F}_{n-1}), \quad l > 1.$$

Теорема 12. Предположим, что $E X_0^s < \infty$ при некотором $s > 1$ и существуют $N \geq 0$ и $\varepsilon > 0$ такие, что

$$m(1) \leq -\varepsilon, \tag{52}$$

$$m(s) < c < \infty. \tag{53}$$

Тогда

$$\liminf_{n \rightarrow \infty} M^{s-1}(n) < \infty. \tag{54}$$

Если, кроме того,

$$M^s(n+1) > M^s(n) - c_1 \tag{55}$$

при некотором $c_1 > 0$, то

$$\sup_n M^{s-1}(n) < \infty. \tag{56}$$

Следствие 7. Если выполнены условия (52), (53) и распределение X_n слабо сходится к некоторому предельному распределению $P(X_n < t) \Rightarrow P(X < t)$, то $EX^{s-1} < \infty$. Это утверждение вытекает из теоремы Фату—Лебега (см. также лемму 6.1)

$$EX^{s-1} \leq \lim_{n \rightarrow \infty} EX_n^{s-1}.$$

Утверждение следствия 7 неулучшаемо. В этом можно убедиться на примере последовательности $X_n = (X_{n-1} + \zeta_n)^+$, где ζ_k независимы и одинаково распределены. Предельное распределение X_n при $E\zeta_k < 0$ совпадает с распределением $S = \sup S_k$ (см. § 11.4). Из факторизационных тождеств можно извлечь, что ES^{s-1} конечно тогда и только тогда, когда $E(\zeta^+)^s < \infty$. Схема доказательства этого факта такова. Из теоремы 11.5 следует, что $ES^k = cE(\chi_+^k; \eta_+ < \infty)$, $c = \text{const} < \infty$. Из следствия 11.1 вытекает, что

$$1 - E(e^{i\lambda\chi_+}; \eta_+ < \infty) = (1 - Ee^{i\lambda\zeta}) \int_0^\infty e^{-i\lambda x} dH(x),$$

где $H(x)$ — функция восстановления случайной величины $\chi_-^0 \geq 0$. Так как $a_1 + b_1x \leq H(x) \leq a_2 + b_2x$ (см. теорему 9.4; a_i, b_i — постоянные), то путем интегрирования свертки $P(\chi_+ > x, \eta_+ < \infty) = \int_0^\infty P(\zeta > v + x) dH(v)$ по частям убеждаемся, что левая часть имеет тот же порядок малости при $x \rightarrow \infty$, как и $\int_0^\infty P(\zeta > v + x) dv$. Отсюда следует требуемое утверждение.

Вернемся к теореме 12. Отметим, что во всех наиболее часто встречающихся задачах $M^{s-1}(n)$ ведет себя «регулярно»: либо эта последовательность ограничена, либо $M^{s-1}(n) \rightarrow \infty$. Утверждение (54) означает, что при выполнении условий теоремы 12 вторая возможность исключается. Условие (55), обеспечивающее (56), также является весьма широким.

Доказательство теоремы 12. Пусть для простоты $s > 1$ — целое. Рассмотрим

$$E(X_n^s; X_{n-1} > N) = \int_N^\infty E((x + \xi_n)^s; X_{n-1} \in dx) = \sum_{l=0}^s C_s^l \int_N^\infty x^l E(\xi_n^{s-l}; X_{n-1} \in dx).$$

Мы только увеличим правую часть, если заменим ξ_n^{s-l} при $s-l \geq 2$ на $|\xi_n|^{s-l}$. Поэтому

$$E(X_n^s; X_{n-1} > N) \leq \sum_{l=0}^s C_s^l m(s-l) M_N^l(n-1),$$

где

$$M_N^l(n) = E(X_n^l; X_n > N).$$

Моменты $M^s(n) = EX_n^s$ удовлетворяют неравенствам

$$\begin{aligned} M^s(n) &\leq E[(N + |\xi_n|)^s; X_{n-1} \leq N] + \sum_{l=0}^s C_s^l m(s-l) M_N^l(n-1) \leq \\ &\leq 2^s [N^s + c] + \sum_{l=0}^s C_s^l m(s-l) M_N^l(n-1). \end{aligned} \quad (57)$$

Предположим теперь, что (54) неверно и $M^{s-1}(n) \rightarrow \infty$. Тогда тем более $M^s(n) \rightarrow \infty$, и существует подпоследовательность n' такая, что $M^s(n') > M^s(n'-1)$. Так как $M^l(n) \leq [M^{l+1}(n)]^{1/(l+1)}$, то из (52) и (57) мы получаем

$$M^s(n') \leq \text{const} + M^s(n'-1) + sM^{s-1}(n'-1)m(1) + o(M^{s-1}(n'-1)) \leq \\ \leq M^s(n'-1) - \frac{1}{2}s\varepsilon M^{s-1}(n'-1)$$

при достаточно больших n' . Это противоречит предположению $M^s(n) \rightarrow \infty$ и доказывает (54).

Если (56) неверно, то существует последовательность n' такая, что $M^{s-1}(n') \rightarrow \infty$, $M^s(n') > M^s(n'-1) - c_1$. Остается вновь воспользоваться приведенными выше рассуждениями.

Изменения в доказательстве для не целых $s > 1$ носят элементарный характер, и мы предоставляем их читателю. ■

Замечание 4. 1) Утверждения теорем 11, 12 останутся справедливыми, если выполнение неравенств (50) или $E(\xi_n + \varepsilon | \mathfrak{F}_{n-1})I(X_{n-1} > N) \leq 0$ требовать не при всех n , а лишь для $n \geq n_0$ при некотором $n_0 > 1$.

2) Как и в теореме 11, условие (52) означает, что последовательность $(X_n + \varepsilon n)I(X_{n-1} > N)$ образует супермартингал.

3) Условия теорем 11 и 12 можно ослабить, заменив их на «усредненные» условия. Рассмотрим, например, условие (47). Интегрируя его по множеству $X_{n-1} > x > N$, получим

$$E(e^{\lambda \xi_n}; X_{n-1} > x) \leq \varphi(\lambda)P(X_{n-1} > x)$$

или, что то же,

$$E(e^{\lambda \xi_n} | X_{n-1} > x) \leq \varphi(\lambda). \quad (58)$$

Обратное утверждение о том, что (58) при любом $x > N$ влечет (47), очевидно, не верно, так что условие (58) слабее условия (47). Аналогичное замечание справедливо относительно условия (50). Справедливы следующие обобщения теорем 11, 12.

Теорема 11А. Пусть при некоторых $\lambda > 0$, $N > 0$ и всех $x \geq N$ выполняется

$$E(e^{\lambda \xi_n} | X_{n-1} > x) \leq \varphi(\lambda) < 1$$

$$E(e^{\lambda \xi_n}; X_{n-1} \leq N) \leq \psi(\lambda) < \infty.$$

Тогда

$$Ee^{\lambda X_n} \leq \varphi^n(\lambda)Ee^{\lambda X(0)} + \frac{e^{\lambda N}\psi(\lambda)}{1 - \varphi(\lambda)}.$$

Обозначим

$$\overline{m}(1) = \sup_{n \geq 1} \sup_{x \geq N} E(\xi_n | X_{n-1} > x),$$

$$\overline{m}(l) = \sup_{n \geq 1} \sup_{x \geq N} E(|\xi_n|^l | X(n) > x), l > 1.$$

Теорема 12А. Пусть $E X_0^s < \infty$ и существуют $N \geq 0$, $\varepsilon > 0$ такие, что

$$\overline{m}(1) \leq -\varepsilon, \quad \overline{m}(s) < \infty$$

$$E(|\xi_n|^s; X_{n-1} \leq N) < c < \infty.$$

Тогда справедливо (54). Если, кроме того, выполнено (55), то справедливо (56).

Доказательство теорем 11А, 12А совершенно аналогично доказательству теорем 11, 12. Единственным дополнительным элементом в обоих случаях является интегрирование по частям. Мы проиллюстрируем это на доказательстве теоремы 11А. Рассмотрим

$$\begin{aligned}
 E(e^{\lambda X_n}; X_{n-1} > N) &= \int_N^{\infty} e^{\lambda x} E(e^{\lambda \xi_n}; X_{n-1} \in dx) = \\
 &= E(e^{\lambda(N+\xi_n)}; X_{n-1} > N) + \int_N^{\infty} \lambda e^{\lambda x} E(e^{\lambda \xi_n}; X_{n-1} > x) dx \leq \\
 &\leq E(e^{\lambda(N+\xi_n)}; X(n) > N) + \varphi(\lambda) \int_N^{\infty} \lambda e^{\lambda x} P(X_{n-1} > x) dx = \\
 &= e^{\lambda N} E(e^{\lambda \xi_n} - \varphi(\lambda); X_{n-1} > N) + \varphi(\lambda) \int_N^{\infty} e^{\lambda x} P(X_{n-1} \in dx) \leq \\
 &\leq \varphi(\lambda) E(e^{\lambda X_{n-1}}; X_{n-1} > N).
 \end{aligned}$$

Отсюда находим

$$\begin{aligned}
 \varphi_n(\lambda) &\equiv E e^{\lambda X_n} \leq E(e^{\lambda(X_{n-1}+\xi_n)}; X_{n-1} \leq N) + E(e^{\lambda X_n}; X_{n-1} > N) \leq \\
 &\leq e^{\lambda N} \psi(\lambda) + \varphi(\lambda) E(e^{\lambda X_{n-1}}; X_{n-1} > N) \leq \\
 &\leq e^{\lambda N} \psi(\lambda) - P(X_{n-1} \leq N) \varphi(\lambda) + \varphi(\lambda) \varphi_n(\lambda); \\
 \varphi_n(\lambda) &\leq \varphi^n(\lambda) \varphi_0(\lambda) + \frac{e^{\lambda N} \psi(\lambda)}{1 - \varphi(\lambda)}.
 \end{aligned}$$

■

К теоремам об ограниченности стохастических последовательностей могут быть отнесены также теорема 12.15 и следствие 12.3 о «положительной возвратности» таких последовательностей.

Стационарные (в узком смысле) последовательности

§ 1. Основные понятия

Пусть $\langle \Omega, \mathfrak{F}, P \rangle$ — исходное вероятностное пространство и $\xi = (\xi_0, \xi_1, \dots)$ — заданная на нем бесконечная последовательность случайных величин.

Определение 1. Последовательность ξ называется *стационарной* в узком смысле, если при любом k распределение вектора $(\xi_n, \dots, \xi_{n+k})$ от n , $n \geq 0$, не зависит.

Наряду с последовательностью ξ рассмотрим последовательность $(\xi_n, \xi_{n+1}, \dots)$. Так как конечномерные распределения этих последовательностей (т. е. распределения векторов $(\xi_m, \dots, \xi_{m+k})$) совпадают, то будут совпадать и распределения самих этих последовательностей (нужно воспользоваться теоремой о продолжении меры (см. Приложение 1) или теоремой Колмогорова (см. § 3.7)). Другими словами, для стационарной последовательности ξ при любом n и любом $B \in \mathfrak{B}^\infty$ (обозначения см. в § 3.7) выполняется

$$P(\xi \in B) = P((\xi_n, \xi_{n+1}, \dots) \in B).$$

Простейшим примером стационарной последовательности является последовательность независимых одинаково распределенных случайных величин $\zeta = (\zeta_0, \zeta_1, \dots)$. Ясно, что последовательность $\xi_k = \alpha_0 \zeta_k + \dots + \alpha_s \zeta_{k+s}$, $k = 0, 1, 2, \dots$, будет стационарной, но величины ξ_k уже не будут независимыми. То же справедливо и для последовательности

$$\xi_k = \sum_{j=0}^{\infty} \alpha_j \zeta_{k+j},$$

если $E|\zeta_k| < \infty$, $\sum |\alpha_j| < \infty$ или если $E\zeta_k = 0$, $D\zeta_k < \infty$, $\sum \alpha_j^2 < \infty$ (последние условия обеспечивают сходимость п. н. рассматриваемых рядов случайных величин; см. § 10.2). Аналогичным образом можно рассматривать «порожденные ζ » стационарные последовательности $\xi_k = g(\zeta_k, \zeta_{k+1}, \dots)$, где $g(x)$ — произвольный измеримый функционал $\mathbb{R}^\infty \rightarrow \mathbb{R}$.

Другой пример — стационарные цепи Маркова. Если $\{X_n\}$ — вещественнозначная цепь Маркова, имеющая инвариантную меру π и переходную вероятность $P(\cdot, \cdot)$, то цепь $\{X_n\}$ с $X_0 \in \pi$ будет образовывать стационарную последовательность, так как для нее распределение

$$P(X_n \in B_0, \dots, X_{n+k} \in B_k) = \int_{B_0} \pi(dx_0) \int_{B_1} P(x_0, dx_1) \cdots \int_{B_k} P(x_{k-1}, dx_k)$$

от n не зависит.

Произвольную стационарную последовательность $\xi = (\xi_0, \xi_1, \dots)$ всегда можно распространить до стационарной последовательности $\bar{\xi} = (\dots, \xi_{-1}, \xi_0, \xi_1, \dots)$, «заданной на всей оси».

Действительно, определим совместные распределения $(\xi_n, \dots, \xi_{n+k})$ при любых n , $-\infty < n < \infty$, и k , $k \geq 0$, как распределение (ξ_0, \dots, ξ_k) . Эти распределения, очевидно, будут согласованными (см. § 3.7), и по теореме Колмогорова будет существовать единственное распределение вероятностей на $\mathbb{R}_{-\infty}^{\infty} = \prod_{k=-\infty}^{\infty} \mathbb{R}_k$ с со-

ответствующей σ -алгеброй такое, что каждое конечномерное распределение есть проекция этого распределения на соответствующее подпространство. Остается в качестве случайного элемента $\bar{\xi}$ взять тождественное отображение $\mathbb{R}_{-\infty}^{\infty}$ в себя.

В некоторых последующих разделах нам будет удобно использовать стационарные последовательности, заданные на всей оси.

Пусть ξ — такая последовательность. Определим преобразование θ пространства $\mathbb{R}_{-\infty}^{\infty}$ в себя с помощью соотношений

$$(\theta x)_k = (x)_{k+1} = x_{k+1}, \quad (1)$$

где $(x)_k$ есть k -я координата вектора $x \in \mathbb{R}_{-\infty}^{\infty}$, $-\infty < k < \infty$. Очевидно, что преобразование θ обладает следующими свойствами.

1. Оно взаимоднозначно, θ^{-1} определяется соотношением $(\theta^{-1}x)_k = x_{k-1}$.
2. Последовательность $\theta\bar{\xi}$ снова является стационарной, и ее распределение совпадает с распределением $\bar{\xi}$:

$$P(\theta\bar{\xi} \in B) = P(\bar{\xi} \in B).$$

Последнее свойство преобразования θ естественно назвать свойством «сохранять меру».

Эти замечания объясняют в какой-то мере тот факт, что изучение свойств стационарных последовательностей шло исторически по пути изучения преобразований, сохраняющих меру. Исследования в этой области составляют значительный раздел современного анализа. В дальнейшем мы свяжем конструкцию стационарной последовательности с преобразованиями, сохраняющими меру, при этом эти преобразования удобнее будет считать «первичными».

Определение 2. Пусть $(\Omega, \mathfrak{F}, P)$ — исходное вероятностное пространство.

Преобразование T множества Ω в себя называется *сохраняющим меру*, если:

- 1) преобразование T измеримо, т.е. $T^{-1}A = \{\omega: T\omega \in A\} \in \mathfrak{F}$ для любого $A \in \mathfrak{F}$;
- 2) преобразование T сохраняет меру: $P(T^{-1}A) = P(A)$ для любого $A \in \mathfrak{F}$.

Пусть T — преобразование, сохраняющее меру, T^n — его n -я степень, $\xi = \xi(\omega)$ — некоторая случайная величина. Положим $U\xi(\omega) = \xi(T\omega)$, так что U есть преобразование над случайными величинами, $U^k\xi(\omega) = \xi(T^k\omega)$. Тогда последовательность

$$\xi = \{U^n\xi(\omega)\}_0^\infty = \{\xi(T^n\omega)\}_0^\infty \quad (2)$$

является стационарной последовательностью случайных величин.

Действительно, пусть $A = \{\omega: \xi \in B\}$, $B \in \mathfrak{B}^\infty$, $A_1 = \{\omega: \theta\xi \in B\}$. Имеем

$$\xi = (\xi(\omega), \xi(T\omega), \dots), \quad \theta\xi = (\xi(T\omega), \xi(T^2\omega), \dots).$$

Поэтому $\omega \in A_1$ тогда и только тогда, когда $T\omega \in A$, т.е. $A_1 = T^{-1}A$. Но $P(T^{-1}A) = P(A)$ и, стало быть, $P(A_1) = P(A)$, $P(A_n) = P(A)$ при любом n , где $A_n = \{\omega: \theta^n \xi \in B\}$.

Стационарные последовательности, определенные в (2), мы будем называть *последовательностями, порожденными преобразованием T* .

Для того, чтобы иметь возможность строить стационарные последовательности, определенные на всей оси, нам понадобятся преобразования, сохраняющие меру и действующие как в «положительном», так и в «отрицательном» направлениях.

Определение 2А. Преобразование T мы будем называть *сохраняющим меру в обоих направлениях*, если:

- 1) преобразование T взаимнооднозначно, область определения и область значений T совпадают со всем Ω ;
- 2) преобразования T и T^{-1} измеримы, т.е. $T^{-1}A = \{\omega: T\omega \in A\} \in \mathfrak{F}$, $TA = \{T\omega: \omega \in A\} \in \mathfrak{F}$ для любого $A \in \mathfrak{F}$;
- 3) преобразование T сохраняет меру: $P(T^{-1}A) = P(A)$ и, следовательно, $P(A) = P(TA)$ для любого $A \in \mathfrak{F}$.

Для таких преобразований мы можем, как и прежде, построить стационарные последовательности ξ , определенные на всей оси:

$$\xi = \{U^n \xi(\omega)\}_{-\infty}^{\infty} = \{\xi(T^n \omega)\}_{-\infty}^{\infty}.$$

Соображения, приведенные выше (перед определением 2), показывают, что такой подход «исчерпывает» все стационарные последовательности, заданные на $(\Omega, \mathfrak{F}, P)$, т.е. с любой стационарной последовательностью ξ можно связать сохраняющее меру преобразование T и случайную величину $\xi = \xi_0$ такие, что $\xi_k(\omega) = \xi_0(T^k \omega)$. При таком построении мы рассматриваем «выборочное вероятностное пространство» $(\mathbb{R}^{\infty}, \mathfrak{B}^{\infty}, P)$, для которого $\xi(\omega) = \omega$, $\theta = T$. Преобразование $\theta = T$ (т.е. преобразование (1)) мы будем называть *преобразованием поттраекторного сдвига*. Оно существует всегда и «порождает» любую стационарную последовательность.

Приведем теперь более простые примеры преобразований, сохраняющих меру (в обоих направлениях).

Пример 1. Пусть $\Omega = \{\omega_1, \dots, \omega_d\}$, $d \geq 2$, есть конечное множество, \mathfrak{F} — σ -алгебра всех его подмножеств, $T\omega_i = \omega_{i+1}$, $1 \leq i \leq d-1$, $T\omega_d = \omega_1$. Если $P(\omega_i) = 1/d$, то T и T^{-1} — сохраняющие меру преобразования.

Пример 2. Пусть $\Omega = [0, 1)$, \mathfrak{F} — σ -алгебра борелевских множеств, P — мера Лебега, s — фиксированное число. Тогда $T\omega = \omega + s \pmod{1}$ есть преобразование, сохраняющее меру в обоих направлениях.

В этих примерах пространство Ω является весьма бедным, что позволяет строить на них лишь стационарные последовательности с детерминированной или почти детерминированной зависимостью между элементами. Если выбрать в примере 1 $\xi(\omega)$ так, что $\xi(\omega_i)$ все различны, то значение $\xi_k(\omega) = \xi(T^k \omega)$ будет однозначно определять $T^k \omega$ и, стало быть, $T^{k+1} \omega$ и $\xi_{k+1}(\omega)$. То же можно сказать и о примере 2, если $\xi(\omega)$, $\omega \in [0, 1)$, будет монотонной функцией ω .

Как показывают рассуждения в начале параграфа, пространство $\Omega = \mathbb{R}^{\infty}$ является достаточно богатым, чтобы построить на нем любую стационарную последовательность.

Итак, мы видели, что понятие преобразования, сохраняющего меру, возникает естественным образом при изучении стационарных случайных процессов. Но не только в этом случае. Они возникают, например, и при изучении динамики некоторых физических систем. Действительно, все приведенные выше рассмотрения сохраняются, если вместо вероятности P рассматривать произвольную меру μ на (Ω, \mathfrak{F}) . Например, для $\Omega = \mathbb{R}^d$ значение $\mu(A)$, $A \in \mathfrak{F}$, может быть мерой Лебега (объемом) множества A . Свойство преобразования T сохранять меру будет означать, что любое множество A после воздействия преобразования T (которое соответствует, скажем, изменению состояния физической системы за единицу времени) будет сохранять свой объем. Это свойство является весьма естественным для «несжимаемых жидкостей». Многие законы, которые будут установлены ниже, будут в равной мере относиться и к таким физическим системам.

Вернемся к вероятностным моделям, т.е. к случаю, когда мера есть вероятность. В этих случаях оказывается, что для любого множества A , $P(A) > 0$, «траектория» $T^n\omega$, $n = 0, 1, \dots$, для п.в. $\omega \in A$ бесконечно много раз побывает в A (п.в. относительно распределения P).

Теорема 1 (Пуанкаре). Пусть T — сохраняющее меру преобразование, $A \in \mathfrak{F}$. Тогда для п.в. $\omega \in A$ соотношение $T^n\omega \in A$ выполнено для бесконечно многих $n \geq 1$.

Доказательство. Обозначим $N = \{\omega \in A: T^n\omega \notin A \text{ при всех } n \geq 1\}$. Так как $\{\omega: T^n\omega \in A\} \in \mathfrak{F}$, то нетрудно видеть, что $N \in \mathfrak{F}$. Очевидно, что $N \cap T^{-n}N = \emptyset$ при любом $n \geq 1$, $T^{-m}N \cap T^{-(m+n)}N = T^{-m}(N \cap T^{-n}N) = \emptyset$. Это значит, что мы имеем бесконечное число множеств $T^{-n}N$, $n = 0, 1, 2, \dots$, которые не пересекаются и имеют одинаковую вероятность. Это означает, очевидно, что $P(N) = 0$.

Таким образом, для каждого $\omega \in A \setminus N$ существует $n_1 = n_1(\omega)$ такое, что $T^{n_1}\omega \in A$.

Применим теперь это утверждение к сохраняющему меру преобразованию $T_k = T^k$, $k \geq 1$. Тогда для каждого $\omega \in A \setminus N_k$, $P(N_k) = 0$, найдется $n_k = n_k(\omega) \geq 1$ такое, что $(T^k)^{n_k}\omega \in A$. Так как $kn_k \geq k$, то теорема доказана. ■

Следствие 1. Пусть $\xi(\omega) \geq 0$, $A = \{\omega: \xi(\omega) > 0\}$. Тогда для п.в. $\omega \in A$

$$\sum_{n=0}^{\infty} \xi(T^n\omega) = \infty.$$

Доказательство. Положим $A_k = \{\omega: \xi(\omega) \geq 1/k\} \subset A$. Тогда в силу теоремы 1 рассматриваемый ряд расходится для п.в. $\omega \in A_k$. Остается заметить, что $A = \bigcup A_k$. ■

Замечание 1. Формально условие $P(A) > 0$ в теореме 1 и следствии 1 не требуется. Однако при его отсутствии эти утверждения становятся малосодержательными, так как множество $A \setminus N$ в доказательстве теоремы 1 может оказаться пустым. Пусть, например, в условиях примера 2 множество A одноточечно $A = \{\omega\}$, $\omega \in [0, 1)$. Тогда если s иррационально, то $T^k\omega$ ни при каких $k \geq 1$ не попадают в A . Действительно, если допустить противное, то мы придем к существованию целых k и m таких, что $\omega + sk - m = \omega$, $s = m/k$, что противоречит иррациональности s .

§ 2. Свойства эргодичности (метрической транзитивности), перемешивания и слабой зависимости

Определение 3. Множество $A \in \mathfrak{F}$ называется *инвариантным* (относительно сохраняющего меру преобразования T), если $T^{-1}A = A$. Множество $A \in \mathfrak{F}$ называется *почти инвариантным*, если множества A и $T^{-1}A$ отличаются друг от друга на множество вероятности 0: $P(A \oplus T^{-1}A) = 0$, где $A \oplus B$ — симметрическая разность $A \oplus B = A\bar{B} \cup \bar{A}B$.

Ясно, что класс всех инвариантных (почти инвариантных) множеств образует σ -алгебру, которую мы обозначим через \mathfrak{I} (\mathfrak{I}^*).

Лемма 1. Если A является почти инвариантным множеством, то найдется инвариантное множество B такое, что $P(A \oplus B) = 0$.

Доказательство. Положим $B = \limsup T^{-n}A$ ($\limsup A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k$ есть множество точек, принадлежащих бесконечно многим A_k). Тогда $T^{-1}B = \limsup T^{-(n+1)}A = B$, т.е. $B \in \mathfrak{I}$. Нетрудно видеть, что $A \oplus B \subset \bigcup_{k=0}^{\infty} (T^{-k}A \oplus T^{-(k+1)}A)$. Так как $P(T^{-k}A \oplus T^{-(k+1)}A) = P(A \oplus T^{-1}A) = 0$, то $P(A \oplus B) = 0$. ■

Определение 4. Сохраняющее меру преобразование T называется *эргодическим* (или *метрически транзитивным*), если каждое инвариантное множество имеет вероятность 0 или 1.

Стационарная последовательность $\{\xi_k\}$, ассоциированная с таким T (породившая T или порожденная T), также называется эргодической (или метрически транзитивной).

Лемма 2. Преобразование T эргодично тогда и только тогда, когда каждое почти инвариантное множество имеет вероятность 0 или 1.

Доказательство. Пусть T эргодично и $A \in \mathfrak{I}^*$. Тогда по лемме 1 найдется инвариантное множество B такое, что $P(A \oplus B) = 0$. Так как $P(B)$ равно нулю или единице, то $P(A)$ равно нулю или единице. Обратное утверждение очевидно. ■

Определение 5. Случайная величина $\zeta = \zeta(\omega)$ называется *инвариантной* (почти инвариантной), если $\zeta(\omega) = \zeta(T\omega)$ для всех $\omega \in \Omega$ (для почти всех $\omega \in \Omega$).

Теорема 2. Пусть T — сохраняющее меру преобразование. Следующие три условия эквивалентны:

- 1) T эргодично;
- 2) каждая почти инвариантная случайная величина п. н. постоянна;
- 3) каждая инвариантная случайная величина п. н. постоянна.

Доказательство. 1) \Rightarrow 2). Пусть T эргодично и ξ почти инвариантна, т.е. $\xi(\omega) = \xi(T\omega)$ п. н. Тогда для любого $v \in \mathbb{R}$ $A_v = \{\omega: \xi(\omega) \leq v\} \in \mathfrak{I}^*$ и по лемме 2 $P(A_v)$ равно нулю или единице. Пусть $V = \sup\{v: P(A_v) = 0\}$. Так как $A_v \uparrow \Omega$ при $v \uparrow \infty$ и $A_v \downarrow \emptyset$ при $v \downarrow -\infty$, то $|V| < \infty$ и

$$P(\xi(\omega) < V) = P\left(\bigcup_{n=1}^{\infty} \left\{\xi(\omega) < V - \frac{1}{n}\right\}\right) = 0.$$

Аналогично $P(\xi(\omega) > V) = 0$. Поэтому $P(\xi(\omega) = V) = 1$.

2) \Rightarrow 3). Очевидно.

3) \Rightarrow 1). Пусть $A \in \mathcal{I}$. Тогда I_A — инвариантная случайная величина, и так как она постоянна, то $I_A = 0$ или $I_A = 1$ п. н. Отсюда следует, что $P(A)$ равно нулю или единице. ■

Утверждение теоремы остается, очевидно, в силе, если в 3) рассматривать лишь ограниченные случайные величины. Кроме того, если ξ инвариантна, то срезка $\xi_{(N)} = \min(\xi, N)$ тоже инвариантна.

Вернемся к примерам 1, 2. В примере 1 $\Omega = (\omega_1, \dots, \omega_d)$, $T\omega_i = \omega_{i+1} \pmod{d}$, $P(\omega_i) = 1/d$. Такое преобразование T , очевидно, метрически транзитивно.

В примере 2 $\Omega = [0, 1)$, $T\omega = \omega + s \pmod{1}$, P — мера Лебега. Покажем, что T эргодично тогда и только тогда, когда s иррационально.

Рассмотрим случайную величину $\xi = \xi(\omega)$ с интегрируемым квадратом $E\xi^2(\omega) < \infty$. Тогда в силу равенства Парсеваля ряд Фурье для нее $\xi(\omega) = \sum a_n e^{2\pi i n \omega}$ обладает свойством $\sum |c_n|^2 < \infty$. Пусть s иррационально, а ξ инвариантна. Тогда

$$\begin{aligned} a_n &= E\xi(\omega)e^{-2\pi i n \omega} = E\xi(T\omega)e^{-2\pi i n T\omega} = \\ &= e^{-2\pi i n s} E\xi(T\omega)e^{-2\pi i n \omega} = e^{-2\pi i n s} E\xi(\omega)e^{-2\pi i n \omega} = e^{-2\pi i n s} a_n. \end{aligned}$$

Для иррациональных s это равенство возможно лишь при $a_n = 0$, $n \geq 1$, $\xi(\omega) = a_0 = \text{const}$. Это означает в силу теоремы 2, что T эргодично.

Пусть теперь $s = m/n$ рационально (m и n целые). Тогда множество

$$A = \bigcup_{k=0}^{n-1} \left\{ \omega: \frac{2k}{2n} \leq \omega < \frac{2k+1}{2n} \right\}$$

будет инвариантным, $P(A) = 1/2$. Это означает, что T не эргодично.

Определение 6. Сохраняющее меру преобразование T обладает свойством *перемешивания*, если для любых $A_1, A_2 \in \mathcal{F}$ при $n \rightarrow \infty$

$$P(A_1 \cap T^{-n}A_2) \rightarrow P(A_1)P(A_2). \quad (3)$$

Рассмотрим теперь стационарную последовательность $\xi = (\xi_0, \xi_1, \dots)$, порожденную T : $\xi_k(\omega) = \xi_0(T^k \omega)$.

Определение 7. Стационарная последовательность ξ обладает свойством *слабой зависимости*, если ξ_k и ξ_{k+n} асимптотически независимы при $n \rightarrow \infty$, т. е. для любых $B_1, B_2 \in \mathcal{B}$

$$P(\xi_k \in B_1, \xi_{k+n} \in B_2) \rightarrow P(\xi_0 \in B_1)P(\xi_0 \in B_2). \quad (4)$$

Теорема 3. Любая стационарная последовательность ξ , порожденная T , обладает свойством слабой зависимости тогда и только тогда, когда T обладает свойством перемешивания.

Доказательство. Пусть T обладает свойством перемешивания. Положим $A_i = \xi_0^{-1}(B_i)$, $i = 1, 2$, и примем в (4) $k = 0$. Тогда

$$P(\xi_0 \in B, \xi_n \in B_2) = P(A_1 \cap T^{-n}A_2) \rightarrow P(A_1)P(A_2).$$

Пусть теперь любая последовательность, порожденная T , обладает свойством слабой зависимости. Для заданных $A_1, A_2 \in \mathfrak{F}$ определим случайную величину

$$\xi(\omega) = \begin{cases} 0, & \text{если } \omega \notin A_1 \cup A_2; \\ 1, & \text{если } \omega \in A_1 \bar{A}_2; \\ 2, & \text{если } \omega \in A_1 A_2; \\ 3, & \text{если } \omega \in \bar{A}_1 A_2; \end{cases}$$

и положим $\xi_k(\omega) = \xi(T^k \omega)$. Тогда

$$P(A_1 \cap T^{-n} A_2) = P(0 < \xi_0 < 3, \xi_n > 2) \rightarrow P(0 < \xi_0 < 3)P(\xi_0 > 2) = P(A_1)P(A_2).$$

Пусть $\{X_n\}$ — стационарная вещественнозначная цепь Маркова с инвариантным распределением π , удовлетворяющая условиям эргодической теоремы, т. е. такая, что при каждом $B \in \mathfrak{B}$, $x \in \mathbb{R}$

$$P(X_n \in B | X_0 = x) \rightarrow \pi(B).$$

Тогда $\{X_n\}$ обладает свойством слабой зависимости и, следовательно, в силу теоремы 3 соответствующее преобразование T обладает свойством перемешивания. Действительно,

$$P(X_0 \in B_1, X_n \in B_2) = EI(X_0 \in B_1)P(X_n \in B_2 | X_0),$$

где последний множитель при каждом X_0 сходится к $\pi(B_2)$. Поэтому рассматриваемая вероятность сходится к $\pi(B_2)\pi(B_1)$.

Дальнейшая характеристика свойства перемешивания будет дана в теоремах 4, 5.

Введем теперь в рассмотрение несколько более широкие понятия, чем те, что даны в определениях 6, 7.

Определение 8. Преобразование T обладает свойством *перемешивания в среднем*, если при $n \rightarrow \infty$

$$\frac{1}{n} \sum_{k=1}^n P(A_1 \cap T^{-k} A_2) \rightarrow P(A_1)P(A_2). \quad (5)$$

Стационарная последовательность ξ обладает свойством *слабой зависимости в среднем*, если

$$\frac{1}{n} \sum_{k=1}^n P(\xi_0 \in B_1, \xi_k \in B_2) \rightarrow P(\xi_0 \in B_1)P(\xi_0 \in B_2). \quad (6)$$

Теорема 3А. Любая стационарная последовательность ξ , порожденная T , обладает свойством слабой зависимости в среднем тогда и только тогда, когда T обладает свойством перемешивания в среднем.

Доказательство совершенно аналогично доказательству теоремы 3, и мы предоставляем его читателю.

Если $\{X_n\}$ — периодическая с периодом d вещественнозначная стационарная цепь Маркова такая, что каждая из вложенных подцепей $\{X_{i+nd}\}_{n=0}^{\infty}$, $i = 0, \dots, d-1$, удовлетворяет условиям эргодичности с инвариантными распределениями $\pi^{(i)}$ на

непересекающихся множествах $\mathcal{X}_0, \dots, \mathcal{X}_{d-1}$, то «общее» инвариантное распределение π будет равно $\frac{1}{d} \sum_{i=0}^{d-1} \pi^{(i)}$, а цепь $\{X_n\}$ будет обладать свойством слабой зависимости в среднем. В то же время очевидно, что она не будет обладать при $d > 1$ свойством слабой зависимости.

Теорема 4. *Сохраняющее меру преобразование T эргодично тогда и только тогда, когда оно обладает свойством перемешивания в среднем.*

Доказательство. Пусть T обладает свойством перемешивания в среднем, и пусть $A_1 \in \mathfrak{F}$, $A_2 \in \mathcal{I}$. Тогда $A_2 = T^{-k} A_2$ и, стало быть, $P(A_1 \cap T^{-k} A_2) = P(A_1 A_2)$ при всех $k \geq 1$. Поэтому (5) означает, что $P(A_1 A_2) = P(A_1)P(A_2)$. При $A_1 = A_2$ получаем $P(A_2) = P^2(A_2)$ и, следовательно, $P(A_2)$ равно нулю или единице.

Доказательство обратного утверждения мы отложим до следующего параграфа. ■

Приведем еще одно важное свойство эргодических преобразований.

Теорема 5. *Сохраняющее меру преобразование T эргодично тогда и только тогда, когда для любого $A \in \mathfrak{F}$, $P(A) > 0$, выполняется*

$$P\left(\bigcup_{n=0}^{\infty} T^{-n} A\right) = 1. \quad (7)$$

Свойство (7) означает, что множества $T^{-n} A$, $n = 0, 1, \dots$, «заметают» все пространство Ω , что хорошо ассоциируется с термином «перемешивание».

Доказательство. Пусть T эргодично. Обозначим $B = \bigcup_{n=0}^{\infty} T^{-n} A$. Тогда $T^{-1} B \subset B$. Так как T сохраняет меру, то, кроме того, $P(T^{-1} B) = P(B)$. Отсюда следует, что $T^{-1} B = B$ с точностью до множества меры нуль и, стало быть, B почти инвариантно. Так как T эргодично, то $P(B)$ равно нулю или единице. Но $P(B) \geq P(A) > 0$ и, следовательно, $P(B) = 1$.

Обратно, если T не эргодично, то существует инвариантное множество A , $0 < P(A) < 1$, и, стало быть, для него $T^{-n} A = A$,

$$P(B) = P(A) < 1. \quad \blacksquare$$

Замечание 2. В § 1, 2 мы предполагали явно или неявно (главным образом для простоты изложения), что координаты ξ_k стационарной последовательности ξ вещественнозначны. Однако на самом деле это нигде, по существу, не использовалось, и мы могли бы так же, как при изучении цепей Маркова, считать фазовое пространство \mathcal{X} , в котором принимают значение ξ_k , произвольным измеримым пространством. Такие конструкции могут понадобиться в следующей главе. В следующем параграфе вещественнозначность (или векторнозначность) ξ_k будет использоваться по существу.

§ 3. Эргодическая теорема

В главе 10 мы доказали для последовательности ξ_1, ξ_2, \dots независимых одинаково распределенных случайных величин усиленный закон больших чисел:

$$\frac{S_n}{n} \xrightarrow[p. н.]{\text{}} E\xi_1,$$

где $S_n = \sum_{k=0}^{n-1} \xi_k$. Теперь такое же утверждение мы докажем при значительно более широких предположениях — для произвольной стационарной эргодической последовательности, т. е. последовательности, обладающей свойством слабой зависимости в среднем.

Итак, пусть $\{\xi_k\}$ — произвольная стационарная в узком смысле последовательность, T — ассоциированное с ней сохраняющее меру преобразование, \mathcal{I} — σ -алгебра инвариантных множеств.

Теорема 6 (Биркгоф, Хинчин). Если $E|\xi_0| < \infty$, то

$$\frac{1}{n} \sum_{k=0}^{n-1} \xi_k \xrightarrow{\text{п. н.}} E(\xi_0 | \mathcal{I}). \quad (8)$$

Если последовательность $\{\xi_k\}$ (или преобразование T) эргодична, то

$$\frac{1}{n} \sum_{k=0}^{n-1} \xi_k \xrightarrow{\text{п. н.}} E\xi_0. \quad (9)$$

В дальнейших рассуждениях, связанных с доказательством, мы будем пользоваться представлением $\xi_k = \xi(T^k\omega)$ при $\xi = \xi_0$. Нам понадобится

Лемма 3. Пусть

$$S_n(\omega) = \sum_{k=0}^{n-1} \xi(T^k\omega),$$

$$M_k(\omega) = \max\{0, S_1(\omega), \dots, S_k(\omega)\}.$$

Тогда при выполнении условий теоремы 6

$$E[\xi(\omega)I_{\{M_n > 0\}}(\omega)] \geq 0$$

при любом $n \geq 1$.

Доказательство. При всех $k \leq n$ имеем $S_k(T\omega) \leq M_n(T\omega)$ и, значит, $\xi(\omega) + M_n(T\omega) \geq \xi(\omega) + S_k(T\omega) = S_{k+1}(\omega)$. Так как $\xi(\omega) \geq S_1(\omega) - M_n(T\omega)$, то

$$\xi(\omega) \geq \max\{S_1(\omega), \dots, S_n(\omega)\} - M_n(T\omega).$$

Далее, так как $\{M_n(\omega) > 0\} = \{\max(S_1(\omega), \dots, S_n(\omega)) > 0\}$, то

$$E[\xi(\omega)I_{\{M_n > 0\}}(\omega)] \geq E(\max(S_1(\omega), \dots, S_n(\omega)) - M_n(T\omega)) I_{\{M_n > 0\}}(\omega) \geq$$

$$\geq E(M_n(\omega) - M_n(T\omega)) I_{\{M_n > 0\}}(\omega) \geq E(M_n(\omega) - M_n(T\omega)) = 0.$$

Доказательство теоремы 6. Утверждение (9) является очевидным следствием (8), так как для эргодического T σ -алгебра \mathcal{I} тривиальна и $E(\xi | \mathcal{I}) = E\xi$ п. н. Поэтому нам достаточно доказать (8).

Не ограничивая общности, можно считать, что $E(\xi | \mathcal{I}) = 0$, так как всегда можно вместо ξ рассматривать $\xi - E(\xi | \mathcal{I})$.

Пусть $\bar{S} = \limsup \frac{S_n}{n}$, $\underline{S} = \liminf \frac{S_n}{n}$. Для доказательства надо установить, что

$$0 \leq \underline{S} \leq \bar{S} \leq 0 \quad \text{п. н.} \quad (10)$$

Так $\bar{S}(\omega) = \bar{S}(T\omega)$, то \bar{S} инвариантна и, следовательно, инвариантным является также множество $A_\varepsilon = \{\bar{S}(\omega) > \varepsilon\}$ при любом $\varepsilon > 0$. Введем в рассмотрение величины

$$\begin{aligned}\xi^*(\omega) &= (\xi(\omega) - \varepsilon)I_{A_\varepsilon}(\omega), \\ S_k^*(\omega) &= \xi^*(\omega) + \dots + \xi^*(T^{k-1}\omega), \\ M_k^*(\omega) &= \max(0, S_1^*, \dots, S_k^*).\end{aligned}$$

Тогда в силу леммы 3 для любого $n \geq 1$

$$E\xi^*I_{\{M_n^* > 0\}} \geq 0.$$

Но при $n \rightarrow \infty$

$$\begin{aligned}\{M_n^* > 0\} &= \left\{ \max_{1 \leq k \leq n} S_k^* > 0 \right\} \uparrow \left\{ \sup_{k \geq 1} S_k^* > 0 \right\} = \\ &= \left\{ \sup_{k \geq 1} \frac{S_k^*}{k} > 0 \right\} = \left\{ \sup_{k \geq 1} \frac{S_k}{k} > \varepsilon \right\} \cap A_\varepsilon = A_\varepsilon.\end{aligned}$$

Последнее равенство следует из того, что

$$A_\varepsilon = \{\bar{S} > \varepsilon\} \subset \left\{ \sup_{k \geq 1} \frac{S_k}{k} > \varepsilon \right\}.$$

Далее, $E|\xi^*| \leq E|\xi| + \varepsilon$. Поэтому по теореме о мажорируемой сходимости

$$0 \leq E\xi^*I_{\{M_n^* > 0\}} \rightarrow E\xi^*I_{A_\varepsilon}.$$

Следовательно,

$$0 \leq E\xi^*I_{A_\varepsilon} = E(\xi - \varepsilon)I_{A_\varepsilon} = E\xi I_{A_\varepsilon} - \varepsilon P(A_\varepsilon) = EI_{A_\varepsilon} E(\xi | \mathcal{J}) - \varepsilon P(A_\varepsilon) = -\varepsilon P(A_\varepsilon).$$

Отсюда вытекает, что $P(A_\varepsilon) = 0$ для любого $\varepsilon > 0$, и, стало быть, $P(\bar{S} \leq 0) = 1$.

Аналогичным образом, рассматривая вместо ξ величины $-\xi$, мы получим

$$\overline{\lim} \left(-\frac{S_n}{n} \right) = -\underline{\lim} \frac{S_n}{n} = -\underline{S},$$

$P(-\underline{S} \leq 0) = 1$, $P(\underline{S} \geq 0) = 1$. Требуемые неравенства (10), а вместе с ними и теорема, доказаны. ■

Мы можем теперь закончить

Доказательство теоремы 4. Нам осталось доказать, что эргодичность T влечет за собой свойство перемешивания в среднем. Действительно, пусть T эргодично, $A_1, A_2 \in \mathfrak{F}$. Тогда в силу теоремы 6 $\zeta_n = \frac{1}{n} \sum_{k=1}^n I(T^{-k}A_2) \xrightarrow[\text{п. н.}]{} P(A_2)$, $I(A_1)\zeta_n \xrightarrow[\text{п. н.}]{} I(A_1)P(A_2)$. Так как $\zeta_n I(A_1)$ ограничены, то имеет место также сходимость $E\zeta_n I(A_1) \rightarrow P(A_2) \cdot P(A_1)$. Поэтому

$$\frac{1}{n} \sum_{k=1}^n P(A_1 \cap T^{-k}A_2) = EI(A_1)\zeta_n \rightarrow P(A_1)P(A_2).$$

Покажем теперь, что в (8), (9) имеет место также сходимость в среднем. ■

Теорема 7. При выполнении условий теоремы 6 наряду с (8), (9) выполняется соответственно

$$E \left| \frac{1}{n} \sum_{k=0}^{n-1} \xi_k - E(\xi_0 | \mathcal{I}) \right| \rightarrow 0, \quad (11)$$

$$E \left| \frac{1}{n} \sum_{k=0}^{n-1} \xi_k - E\xi_0 \right| \rightarrow 0 \quad (12)$$

при $n \rightarrow \infty$.

Доказательство. Мы воспользуемся опять представлением $\xi_k = \xi(T^k \omega)$. Для всякого $\varepsilon > 0$ всегда можно найти ограниченную случайную величину $\zeta(\omega)$, $|\zeta| \leq N$, такую, что $E|\xi - \zeta| \leq \varepsilon$. В качестве ζ можно взять, например, срезку ξ на уровне N . Тогда

$$\begin{aligned} E \left| \frac{1}{n} \sum_{k=0}^{n-1} \xi(T^k \omega) - E(\xi | \mathcal{I}) \right| &\leq E \left| \frac{1}{n} \sum_{k=0}^{n-1} \xi(T^k \omega) - \zeta(T^k \omega) \right| + \\ &+ E \left| \frac{1}{n} \sum_{k=0}^{n-1} \zeta(T^k \omega) - E(\zeta | \mathcal{I}) \right| + E \left| E(\xi | \mathcal{I}) - E(\zeta | \mathcal{I}) \right|. \end{aligned} \quad (13)$$

Так как $|\zeta| \leq N$, то по теореме о мажорируемой сходимости и в силу (8) второе слагаемое в правой части (13) сходится к нулю. Первое и последнее не превышают ε . Так как ε произвольно, то это доказывает (11). Сходимость (12) следует из (11) (см. доказательство теоремы 6). ■

Следствие 1. Если $\{\xi_k\}$ — стационарная метрически транзитивная последовательность, $a = E\xi_k < 0$, то $S(\omega) = \sup_{k \geq 0} S_k(\omega)$ есть собственная случайная величина.

Доказательство очевидно, так как при $0 < \varepsilon < -a$ выполняется $S_k < (a + \varepsilon)k < 0$ для всех $k \geq n(\omega) < \infty$. ■

Необычным в теореме 6 по сравнению с усиленным законом больших чисел в главе 10 является то, что предел $\frac{1}{n} \sum_{k=0}^{n-1} \xi_k$ может оказаться случайной величиной. Пусть, например, в условиях примера 1 мы положим $T\omega_k = \omega_{k+2}$ и будем считать, что $d = 2l$ четно. Тогда преобразование T не будет эргодичным, так как множество $A = \{\omega_1, \omega_3, \dots, \omega_{d-1}\}$ будет инвариантным, $P(A) = 1/2$. С другой стороны, очевидно, что для любой функции $\xi(\omega)$ сумма $\frac{1}{n} \sum_{k=0}^{n-1} \xi(T^k \omega)$ будет сходиться с вероятностью $1/2$ к $\frac{2}{d} \sum_{j=0}^{l-1} \xi(\omega_{2j+1})$ (если $\omega = \omega_i$ и i примет нечетное значение) и с вероятностью $1/2$ к $\frac{2}{d} \sum_{j=1}^l \xi(\omega_{2j})$ (если $\omega = \omega_i$ и i примет четное значение). Это предельное распределение и есть распределение $E(\xi | \mathcal{I})$.

Глава 16

Стохастически рекурсивные последовательности

§ 1. Основные понятия

Рассмотрим два измеримых фазовых пространства $\langle \mathcal{X}, \mathfrak{B}_\mathcal{X} \rangle$ и $\langle \mathcal{Y}, \mathfrak{B}_\mathcal{Y} \rangle$, и пусть $\{\xi_n\}$ — последовательность случайных элементов со значениями в \mathcal{Y} . Если $\langle \Omega, \mathfrak{F}, \mathbf{P} \rangle$ — основное вероятностное пространство, то $\{\omega: \xi_k \in B\} \in \mathfrak{F}$ для любого $B \in \mathfrak{B}_\mathcal{Y}$. Пусть, кроме того, на измеримом пространстве $\langle \mathcal{X} \times \mathcal{Y}, \mathfrak{B}_\mathcal{X} \times \mathfrak{B}_\mathcal{Y} \rangle$, где $\mathfrak{B}_\mathcal{X} \times \mathfrak{B}_\mathcal{Y}$ означает σ -алгебру, порожденную множествами $A \times B$, $A \in \mathfrak{B}_\mathcal{X}$, $B \in \mathfrak{B}_\mathcal{Y}$, задана измеримая функция $f: \mathcal{X} \times \mathcal{Y} \rightarrow \mathcal{X}$.

Для простоты под \mathcal{X} и \mathcal{Y} можно понимать вещественную прямую \mathbb{R} , а под $\mathfrak{B}_\mathcal{X}$, $\mathfrak{B}_\mathcal{Y}$ — σ -алгебры борелевских множеств.

Определение 1. Последовательность $\{X_n\}$, $n = 0, 1, \dots$, назовем *стохастически рекурсивной последовательностью* (с. р. п.), управляемой последовательностью $\{\xi_n\}$, если X_n удовлетворяют соотношениям

$$X_{n+1} = f(X_n, \xi_n) \quad (1)$$

при всех $n \geq 0$. Для простоты мы будем считать начальное значение X_0 не зависящим от $\{\xi_n\}$.

Распределение последовательности $\{X_n, \xi_n\}$ на $(\langle \mathcal{X} \times \mathcal{Y} \rangle^\infty, (\mathfrak{B}_\mathcal{X} \times \mathfrak{B}_\mathcal{Y})^\infty)$ строится очевидным образом по конечномерным распределениям аналогично тому, как по переходной функции $P(x, B) = \mathbf{P}(X_1(x) \in B)$ строится на $\langle \mathcal{X}^\infty, \mathfrak{B}_\mathcal{X}^\infty \rangle$ распределение цепи Маркова X . Конечномерные распределения $\{(X_0, \xi_0), \dots, (X_k, \xi_k)\}$ определяются соотношениями

$$\begin{aligned} \mathbf{P}(X_l \in A_l, \xi_l \in B_l; l = 0, \dots, k) = \\ = \int_{B_0} \dots \int_{B_k} \mathbf{P}(\xi_l \in dy_l, l = 0, \dots, k) \prod_{l=1}^k \mathbf{I}(f_l(X_0, y_0, \dots, y_l) \in A_l), \end{aligned}$$

где $f_1(x, y_0) = f(x, y_0)$, $f_l(x, y_0, \dots, y_l) = f(f_{l-1}(x, y_0, \dots, y_{l-1}), y_l)$.

Последовательность $\{\xi_n\}$, не ограничивая общности, можно считать заданной при всех $-\infty < n < \infty$ (как мы отмечали в § 15.1, для стационарной последовательности необходимое продолжение для $n < 0$ всегда можно осуществить с помощью теоремы Колмогорова).

Стохастически рекурсивная последовательность есть объект более общий, чем цепь Маркова. Ясно, что если ξ_k независимы, то X_n образует цепь Маркова. Справедливо и более сильное утверждение: при широких предположениях относительно пространства $\langle \mathcal{X}, \mathfrak{B}_\mathcal{X} \rangle$ для любой цепи Маркова $\{X_n\}$ можно построить функцию f

и последовательность независимых одинаково распределенных случайных величин $\{\xi_n\}$ такие, что будет выполнено (1). Проще всего это понять в случае, когда \mathcal{X} и \mathcal{Y} совпадают с вещественной прямой \mathbb{R} . Пусть $P(x, B)$, $B \in \mathfrak{B}$ есть переходная функция цепи $\{X_n\}$ и $F_x(t) = P(x, (-\infty, t))$ есть функция распределения $X_1(x)$ ($X_0 = x$). Тогда если $F_x^{-1}(t)$ — обратная (по t) функция к $F_x(t)$, $\alpha \in U_{0,1}$, то, как было установлено ранее (см., например, § 6.2), случайная величина $F_x^{-1}(\alpha)$ имеет функцию распределения $F_x(t)$. Следовательно, если $\{\alpha_n\}$ — последовательность независимых равномерно распределенных на $[0, 1]$ случайных величин, то последовательность $X_{n+1} = F_{X_n}^{-1}(\alpha_n)$ будет иметь то же распределение, что и исходная цепь $\{X_n\}$. Таким образом, рассматриваемая цепь Маркова является с.р.п. с функцией $f(x, y) = F_x^{-1}(y)$ и управляющей последовательностью $\{\alpha_n\}$, $\alpha_n \in U_{0,1}$.

Для более общих фазовых пространств \mathcal{X} аналогичная конструкция возможна, если σ -алгебра \mathfrak{X} является счетнопорожденной (т.е. порожденной счетным набором множеств из \mathcal{X} ; это всегда имеет место для борелевских σ -алгебр в $\mathcal{X} = \mathbb{R}^d$, $d \geq 1$; см. [26]).

Последовательность $f(\cdot, \xi_n)$ можно рассматривать как последовательность случайных преобразований пространства \mathcal{X} в себя. Основной вопрос, который нас будет интересовать в дальнейшем, будет опять, как и в главе 12, связан с существованием предельного распределения X_n при $n \rightarrow \infty$.

В последующих разделах мы рассмотрим три основных подхода к этой задаче.

§ 2. Эргодичность при наличии обновляющих событий. Условия ограниченности

Введем σ -алгебры $\mathfrak{F}_{l,n}^\xi = \sigma\{\xi_k; l \leq k \leq n\}$, $\mathfrak{F}_n^\xi = \sigma\{\xi_k; k \leq n\} = \mathfrak{F}_{-\infty,n}^\xi$, $\mathfrak{F}^\xi = \sigma\{\xi_k; -\infty < k < \infty\} = \mathfrak{F}_{-\infty,\infty}^\xi$.

Определение 2. Событие $A \in \mathfrak{F}_{n+m}^\xi$, $m \geq 0$, назовем *обновляющим* для с.р.п. $\{X_n\}$ на отрезке $[n, n+m]$, если существует такая измеримая функция $g: \mathcal{Y}^{m+1} \rightarrow \mathcal{X}$, что на множестве A (т.е. при $\omega \in A$)

$$X_{n+m+1} = g(\xi_n, \dots, \xi_{n+m}). \quad (2)$$

Очевидно, что при $\omega \in A$ соотношения вида $X_{n+m+k+1} = g_k(\xi_n, \dots, \xi_{n+m+k})$ будут справедливы при всех $k \geq 0$, где функция g_k зависит лишь от своих аргументов и определяется событием A .

Последовательность событий $\{A_n\}$, $A_n \in \mathfrak{F}_{n+m}^\xi$, где целое неотрицательное число m фиксировано, называется *обновляющей* для с.р.п. $\{X_n\}$, если существует такое целое число $n_0 \geq 0$, что для $n \geq n_0$ при $\omega \in A_n$ выполнено соотношение (2), причем функция g одна и та же для всех n .

Основной интерес для нас будут представлять «положительные» обновляющие события, т.е. события, имеющие положительную вероятность $P(A_n) > 0$.

Простейшим примером обновляющего события является попадание X_n в некоторую фиксированную точку x_0 : $A_n = \{X_n = x_0\}$, при этом $m = 0$. Однако такое событие может иметь нулевую вероятность. Ниже мы рассмотрим еще один, более содержательный пример.

Смысл введения обновляющих событий состоит в следующем. После попадания траектории $\{X_k, \xi_k\}$, $k \leq n+m$, в обновляющее множество $A \in \mathfrak{F}_{n+m}^\xi$

дальнейшее течение процесса не зависит от значений $\{X_k\}$, $k \leq n + m$, а зависит лишь от ξ_k, ξ_{n+1}, \dots . Это не дает полной «регенерации» процесса, с которой мы имели дело в главе 12 при изучении цепей Маркова (прежде всего потому, что теперь ξ_k , вообще говоря, зависимы), но все же позволит нам установить эргодичность последовательности X_n (примерно в том же смысле, что и в главе 12).

Отметим, что, вообще говоря, событие A и, следовательно, функция g могут зависеть от начального значения X_0 .

В дальнейшем будет предполагаться, что последовательность $\{\xi_n\}$ стационарна. Символом U будет обозначаться сохраняющее меру преобразование сдвига \mathcal{F}^ξ -измеримых случайных величин, порожденное $\{\xi_n\}$, так что $U\xi_n = \xi_{n+1}$, и символом T — преобразование сдвига множеств (событий) из σ -алгебры \mathcal{F}^ξ : $\xi_{n+1}(\omega) = \xi_n(T\omega)$. Соответственно, символами U^n и T^n , $n \geq 0$, будут обозначаться степени (итерации) этих преобразований (так что $U^1 = U$, $T^1 = T$, U^0 и T^0 суть тождественные преобразования), а символами U^{-n} , T^{-n} — преобразования, обратные к U^n и T^n соответственно.

Последовательность событий $\{A_k\}$ назовем *стационарной*, если $A_k = T^k A_0$ при всех k .

Пример 1. Рассмотрим вещественнозначную последовательность

$$X_{n+1} = (X_n + \xi_n)^+, \quad X_0 = \text{const} \geq 0, \quad (3)$$

где $x^+ = \max(0, x)$, $\{\xi_n\}$ — стационарная метрически транзитивная последовательность.

Используя те же рассуждения, что и в § 11.4, находим, что уравнение (3) имеет явное решение

$$X_n = S_{n-1} - \min(-X_0, S_0, \dots, S_{n-1}), \quad (4)$$

где $S_n = \sum_{k=0}^n \xi_k$. Из этого представления видно, что событие

$$A_n^0 = \{\omega: S_{n-1} \leq S_{n-2}, \dots, S_{n-1} \leq S_0, S_{n-1} \leq -X_0\}$$

является обновляющим, так как оно влечет за собой $X_n = 0$, $X_{n+1} = \xi_n^+$, так что в этом случае $m = 0$, $g(y) = y^+$, $n_0 = 0$. Событие

$$A_n^0 = \{\xi_{n-1} \leq 0, \xi_{n-1} + \xi_{n-2} \leq 0, \dots, \xi_{n-1} + \dots + \xi_1 \leq 0; \xi_{n-1} + \dots + \xi_0 \leq -X_0\}$$

можно несколько сузить, рассмотрев вместо него при $n \geq n_0$ событие

$$\begin{aligned} A_n &= \{\xi_{n-1} \leq 0, \dots, \xi_{n-1} + \dots + \xi_{n-n_0} \leq 0; \xi_{n-1} + \dots + \xi_{n-n_0-1} \leq \\ &\leq -X_0, \xi_{n-1} + \dots + \xi_{n-n_0-2} \leq -X_0, \dots\} = \\ &= \{S_{n-1}^0 \leq 0, \dots, S_{n-1}^{n_0} \leq 0; S_{n-1}^{n_0+1} \leq -X_0, S_{n-1}^{n_0+2} \leq -X_0, \dots\}, \end{aligned}$$

где

$$S_n^k = \xi_n + \dots + \xi_{n-k}.$$

Очевидно, что $A_n = T^n A_0$, так что A_n при $n \geq n_0$ образуют стационарную последовательность обновляющих событий:

$$X_{n+1} = \xi_n^+ \quad \text{при} \quad n \geq n_0$$

на множестве A_n .

Если предположить, что $E\xi_1 < 0$, последовательность $\{\xi_n\}$ метрически транзитивна и, следовательно, $\xi_{-1} + \dots + \xi_{-n} \rightarrow -\infty$ п. н. при $n \rightarrow \infty$, то найдется такой номер $n_0 = n_0(X_0)$, что $P(A_n) > 0$ при $n \geq n_0$.

С другой стороны, если определить события B_n , число m и функцию $g: \mathbb{R}^{m+1} \rightarrow \mathbb{R}$ равенствами

$$m = n_0, \quad B_n = T^m A_n, \quad g(y_0, \dots, y_m) \equiv y_m^+,$$

то события $B_n \in \mathfrak{F}_{n+m}^\xi$ будут обновляющими для $\{X_n\}$ на отрезке $[n, n+m]$ при всех $n \geq 0$, так что в этом случае n_0 можно считать равным нулю.

Аналогичное соображение можно использовать и при произвольных обновляющих событиях. Поэтому в дальнейшем будем без ограничения общности предполагать, что число n_0 , фигурирующее в определении последовательности обновляющих событий, равно нулю.

Конструкции нестационарных обновляющих событий могут быть проще. В примере 1 таковыми являются события $A_n = \{X_{n+1} = 0\} \in \mathfrak{F}_{0,n}^\xi$ при $m = 0$, $g \equiv 0$. Можно рассматривать также вложенные в A_{n+m} обновляющие события

$$A_n^{(m)} = \{X_n \leq b; S_n^m \leq \min(-b, S_n^0, \dots, S_n^{m-1})\} \in \mathfrak{F}_{0,n+m}^\xi$$

при некотором $m \geq 1$, где $S_n^k = \xi_n + \dots + \xi_{n+k}$. События $A_n^{(m)}$ также соответствуют функции $g \equiv 0$.

Теорема 1. Пусть $\{\xi_n\}$ — произвольная стационарная последовательность, и для с. р. н. $\{X_n\}$ существует такая последовательность обновляющих событий $\{A_n\}$, что

$$P\left(\bigcup_{j=1}^n A_j T^{-s} A_{j+s}\right) \rightarrow 1 \quad \text{при } n \rightarrow \infty \quad (5)$$

равномерно по $s \geq 1$. Тогда на одном с $\{X_n\}$ вероятностном пространстве можно определить стационарную последовательность $\{X^n \equiv U^n X^0\}$, удовлетворяющую уравнениям $X^{n+1} = f(X^n, \xi_n)$ и такую, что

$$P\{X_k = X^k \text{ при всех } k \geq n\} \rightarrow 1 \quad \text{при } n \rightarrow \infty. \quad (6)$$

Если последовательность $\{\xi_n\}$ метрически транзитивна, а события A_n стационарны, то соотношения $P(A_0) > 0$ и $P\left(\bigcup_{n=0}^\infty A_n\right) = 1$ эквивалентны и влекут за собой (5), (6).

Отметим, что если так же, как и в главе 12, ввести меру $\pi(B) = P(X^0 \in B)$, то из сходимости (6) следует сходимость по вариации:

$$\sup_{B \in \mathfrak{B}_X} |P(X_n \in B) - \pi(B)| \rightarrow 0 \quad \text{при } n \rightarrow \infty.$$

Доказательство теоремы 1. Покажем сначала, что из сходимости (5) следует, что

$$P\left(\bigcap_{k=0}^\infty \{X_{n+k} \neq U^{-s} X_{n+k+s}\}\right) \rightarrow 0 \quad \text{при } n \rightarrow \infty \quad (7)$$

равномерно по $s \geq 0$. Для фиксированного $s \geq 1$ рассмотрим последовательность $X_j^s = U^{-s} X_{s+j}$. Она определена при $j \geq -s$, $X_{-s}^s = X_0$, $X_{-s+1}^s = f(X_{-s}^s, \xi_{-s}) =$

$f(X_0, \xi_{-s})$ и т.д. Ясно, что событие $\{X_j = X_j^s \text{ при каком-нибудь } j \in [0, n]\}$ влечет за собой событие $\{X_{n+k} = X_{n+k}^s \text{ при всех } k \geq 0\}$. Покажем, что

$$P\left(\bigcup_{j=1}^n \{X_j = X_j^s\}\right) \rightarrow 1 \quad \text{при } n \rightarrow \infty.$$

Положим для простоты $m = 0$. Тогда для того, чтобы произошло событие $X_{j+1} = X_{j+1}^s$ достаточно, чтобы одновременно произошли события A_j и $T^{-s}A_{j+s}$. Другими словами,

$$\bigcup_{j=0}^{n-1} A_j T^{-s} A_{j+s} \subseteq \bigcup_{j=1}^n \{X_j = X_j^s\} \subset \bigcup_{k=0}^{\infty} \{X_{n+k} = X_{n+k}^s\}.$$

Поэтому из (5) следует (7) и сходимость

$$P(X_k^n \neq X_k^{n+s}) \rightarrow 0 \quad \text{при } n \rightarrow \infty$$

равномерно по $k \geq 0$ и $s \geq 0$. Если ввести метрику $\rho: \rho(x, y) = 1$ при $x \neq y$, $\rho(x, x) = 0$, то сказанное означает, что для любого $\delta > 0$ найдется такое N , что

$$P(\rho(X_k^n, X_k^{n+s}) > \delta) = P(\rho(X_k^n, X_k^{n+s}) \neq 0) < \delta$$

при $n \geq N$ для любых $k \geq 0$, $s \geq 0$, т.е. последовательность X_k^n фундаментальна относительно сходимости по вероятности при каждом k . Так как любое пространство X является полным относительно введенной метрики, то (см. лемму 4.2) существует такая случайная величина X^k , что $X_k^n \xrightarrow{p} X^k$ при $n \rightarrow \infty$. В силу специфики метрики ρ это означает, что

$$P(X_k^n \neq X^k) \rightarrow 0 \quad \text{при } n \rightarrow \infty. \quad (8)$$

Последовательность X^k является стационарной. Действительно, при $n \rightarrow \infty$

$$P(X^{k+1} \neq UX^k) = P(X_{k+1}^n \neq UX_k^n) + o(1) = P(X_{k+1}^n \neq X_{k+1}^{n-1}) + o(1) = o(1).$$

Так как вероятность $P(X^{k+1} \neq UX^k)$ от n не зависит, то $X^{k+1} = UX^k$ п.н.

Далее, $X_{n+k+1} = f(X_{n+k}, \xi_{n+k})$ и, следовательно,

$$X_{k+1}^n = U^{-n} f(X_{n+k}, \xi_{n+k}) = f(X_k^n, \xi_k). \quad (9)$$

Здесь правая и левая части сходятся по вероятности к X^{k+1} и $f(X^k, \xi_k)$ соответственно. Это означает, что

$$X^{k+1} = f(X^k, \xi_k).$$

Для доказательства сходимости (6) остается заметить, что в силу (9) значения X_k^n и X^k , совпав при каком-нибудь k , при больших значениях k уже не различаются. Поэтому наряду с (8) справедливо соотношение

$$P\left(\bigcup_{k \geq 0} \{X_k^n \neq X^k\}\right) = P\left(\bigcup_{k \geq 0} \{X_{k+n} \neq X^{k+n}\}\right) \rightarrow 0 \quad \text{при } n \rightarrow \infty,$$

что эквивалентно (6).

Последнее утверждение теоремы вытекает из теоремы 15.5. ■

Замечание 1. Оказывается, что условие (5) является и необходимым для сходимости (6) (см. [6]). Более подробно о сходимости с.р.п. и их обобщений, а также о связи условий (5) с условиями (I), (II) главы 12 см. в [6].

В примере 1 последовательность X^k легко найти в явном виде:

$$X^k = \sup_{j \geq -1} S_{k-1}^j, \quad S_n^{-1} = 0. \quad (10)$$

Эти случайные величины в силу следствия 15.1 будут собственными. Нетрудно видеть также, что при $X_0 = 0$ (см. (4))

$$U^{-n} X_{n+k} \uparrow X^k. \quad (11)$$

Рассмотрим теперь условия ограниченности с.р.п. в пространствах $\mathcal{X} = [0, \infty)$ и $\mathcal{X} = (-\infty, \infty)$. Утверждения об ограниченности мы будем формулировать в форме существования стационарных мажорант, т. е. стационарных последовательностей M_n таких, что

$$X_n \leq M_n \quad \text{при всех } n.$$

Это позволит использовать эти результаты для построения стационарных обновляющих событий.

Мы будем строить мажоранты для более широкого, чем с.р.п., класса случайных последовательностей, удовлетворяющих неравенствам

$$X_{n+1} \leq (X_n + h(X_n, \xi_n))^+, \quad (12)$$

где измеримая функция h в свою очередь будет оцениваться достаточно простыми функциями от X_n и ξ_n . Последовательность $\{\xi_n\}$, как и прежде, мы будем предполагать заданной на всей оси.

Теорема 2. Предположим, что существуют число $N > 0$ и измеримая функция g_1 , обладающая свойством $Eg_1(\xi_n) < 0$, такие, что

$$h(x, y) \leq \begin{cases} g_1(y) & \text{при } x > N, \\ g_1(y) + N - x & \text{при } x \leq N. \end{cases} \quad (13)$$

Тогда если $X_0 \leq M < \infty$, то стационарная последовательность

$$M_n = \max(M, N) + \sup_{j \geq -1} S_{n-1}^j, \quad (14)$$

где $S_n^{-1} = 0$, $S_k^j = g_1(\xi_k) + \dots + g_1(\xi_{k-j})$, образует мажоранту для последовательности X_n , удовлетворяющей (12).

Доказательство. Положим для краткости $g_1(\xi_i) = \zeta_i$ и обозначим $\max(M, N) = Z$, $Z_n = X_n - Z$. Тогда Z_n будут удовлетворять соотношениям

$$\begin{aligned} Z_{n+1} &\leq (Z_n + Z + \zeta_n)^+ - Z \leq (Z_n + \zeta_n)^+ \quad \text{при } Z_n > N - Z, \\ Z_{n+1} &\leq (N + \zeta_n)^+ - Z \leq \zeta_n^+ \quad \text{при } Z_n \leq N - Z. \end{aligned}$$

Рассмотрим теперь последовательность $\{Y_n\}$, определенную соотношениями $Y_0 = 0$,

$$Y_{n+1} = (Y_n + \zeta_n)^+.$$

Предположим, что $Z_n \leq Y_n$. Тогда если $Z_n > N - Z$, то $Z_{n+1} \leq (Z_n + \zeta_n)^+ \leq (Y_n + \zeta_n)^+ = Y_{n+1}$. Если $Z_n \leq N - Z$, то $Z_{n+1} \leq \zeta_n^+ \leq (Y_n + \zeta_n)^+ = Y_{n+1}$. Так как

$Z_0 \leq 0 = Y_0$, то очевидно, что $Z_n \leq Y_n$ при всех n . Но решение уравнения для Y_n нам известно и в силу (10), (11)

$$X_n - Z \leq \sup_{j \geq -1} S_{n-1}^j.$$

Теорема доказана. ■

Теорема 2А. *Предположим, что существуют число $N > 0$ и измеримые функции g_1, g_2 , обладающие свойствами*

$$Eg_1(\xi_n) < 0, \quad Eg_2(\xi_n) < \infty \quad (15)$$

и такие, что

$$h(x, y) \leq \begin{cases} g_1(y) & \text{при } x > N, \\ g_1(y) + g_2(y) & \text{при } x \leq N. \end{cases} \quad (16)$$

Тогда если $X_0 \leq M < \infty$, то выполнены условия теоремы 2 (возможно, при других N и g_1) и для X_n существует стационарная мажоранта вида (13).

Доказательство. Обозначим $g = -Eg_1(\xi_n) > 0$ и найдем число $L > 0$ такое, что $E(g_2(\xi_n); g_2(\xi_n) > L) \leq g/2$. Введем функцию $g_1^*(y) = g_1(y) + g_2(y)I(g_2(y) > L)$. Тогда $Eg_1^*(\xi_n) \leq -g/2 < 0$ и

$$\begin{aligned} h(x, y) &\leq g_1(y) + g_2(y)I(x \leq N) \leq \\ &\leq g_1^*(y) + g_2(y)I(x \leq N) - g_2(y)I(g_2(y) > L) \leq \\ &\leq g_1^*(y) + LI(x \leq N) \leq g_1^*(y) + (L + N - x)I(x \leq N) \leq \\ &\leq g_1^*(y) + (L + N - x)I(x \leq L + N). \end{aligned}$$

Это означает, что имеют место неравенства (12), в которых надо вместо N взять $N^* = N + L$. Теорема доказана. ■

Еще раз отметим, что в теоремах 2, 2А не предполагалось, что $\{X_n\}$ есть с. р. п.

Читатель может заметить сходство условий теорем 2, 2А с условиями ограниченности в § 14.5, и в теореме 12.5 и следствии 12.3.

Форма утверждений теорем 2, 2А позволяет строить стационарные обновляющие события для весьма широкого класса неотрицательных ($\mathcal{X} = [0, \infty)$) с. р. п., имеющих, скажем, «положительный атом» в точке нуль. Такие последовательности удобно записывать в виде

$$X_{n+1} = (X_n + h(X_n, \xi_n))^+. \quad (17)$$

Пример 2. Пусть с. р. п. (см. (1)) описывается уравнением (17) и удовлетворяет условиям (13) или (16), где функция h достаточно «регулярна», чтобы $C_{n,T} = \bigcap_{t \leq T} \{h(t, \xi_n) \leq t\}$ при любом T было событием (достаточно, например, чтобы h имела не более чем счетное число точек разрыва по t ; тогда несчетное пересечение $\bigcap_{t \leq T}$ в определении $C_{n,T}$ можно записать в виде счетного пересечения). Пусть, кроме того, существует $L > 0$ такое, что

$$P(M_n < L, C_{n,L}) > 0, \quad (18)$$

(M_n определено в (14)). Тогда событие $A_n = \{M_n < L\}C_{n,L}$, очевидно, является положительным стационарным обновляющим событием с функцией $g(y) = (h(0, y))^+$,

$m = 0$. Стало быть, с. р. п., удовлетворяющая (17), удовлетворяет условиям теоремы 1 и является эргодической в смысле утверждения (6).

Может оказаться, что за один шаг из точки $t \leq L$ невозможно достичь точки нуля, но это может оказаться возможным за $m > 1$ шагов. Если B есть множество таких $(\xi_n, \dots, \xi_{n+m})$, которые осуществляют такой переход, и $P(M_n < L, B) > 0$, то события $A_n = \{M_n < L\}B$ также будут положительными стационарными обновляющимися событиями.

§ 3. Условия эргодичности, связанные с монотонностью f

Рассмотрим теперь условия эргодичности с. р. п., связанные с аналитическими свойствами функции f в (1). Как уже отмечалось, последовательность $f(x, \xi_k)$, $k = 1, 2, \dots$, можно рассматривать как последовательность случайных преобразований пространств \mathcal{X} в себя. Соотношение (1) показывает, что X_{n+1} есть результат $(n+1)$ -кратного применения к начальному значению $X_0 = x \in \mathcal{X}$ этих случайных преобразований $f(\cdot, \xi_k)$, $k = 0, 1, \dots, n$. Если обозначить через ξ_n^{n+k} вектор $\xi_n^{n+k} = (\xi_n, \dots, \xi_{n+k})$ и через f_k — k -ю итерацию функции f , то представление (1) дает возможность при $X_0 = x$ записать

$$X_{n+1} = X_{n+1}(x) = f_{n+1}(x, \xi_n^n),$$

где $f_1(x, y_1) = f(x, y_1)$, $f_2(x, y_1, y_2) = f(f(x, y_1), y_2)$ и т. д., так что справедливы «прямые» и «обратные» уравнения

$$f_{n+1}(x, \xi_n^n) = f(f_n(x, \xi_0^{n-1}), \xi_n) = f_n(f(x, \xi_0), \xi_1^n). \quad (19)$$

В этом параграфе будут изучаться с. р. п., для которых функция f в представлении (1) обладает свойством монотонности по первому аргументу. Для этого мы должны предполагать, что в пространстве \mathcal{X} определено отношение частичного порядка « \geq ». В пространстве $\mathcal{X} = \mathbb{R}^d$ векторов $x = (x(1), \dots, x(d))$ (или в его подпространствах) отношение порядка « \geq » можно вводить естественным образом, полагая $x_1(\cdot) \geq x_2(\cdot)$, если при всех k выполняется $x_1(k) \geq x_2(k)$.

Будем предполагать, далее, что для каждой неубывающей последовательности $x_1 \leq x_2 \leq \dots \leq x_n \leq \dots$ существует предел $x \in \mathcal{X}$, т. е. такой наименьший элемент $x \in \mathcal{X}$, для которого $x_k \leq x$ при всех k . Мы будем писать в этом случае $x_k \uparrow x$ или $\lim_{k \rightarrow \infty} x_k = x$. В $\mathcal{X} = \mathbb{R}^d$ такая сходимость будет означать обычную сходимость. При этом пространство \mathbb{R}^d необходимо пополнить точками с бесконечными координатами.

Теорема 3 (Лойнес). Пусть преобразование $f = f(x, y)$ и пространство \mathcal{X} удовлетворяют следующим условиям:

- 1) существует такое $x_0 \in \mathcal{X}$, что $f(x_0, y) \geq x_0$ при всех $y \in \mathcal{Y}$;
- 2) функция f монотонна по первому аргументу: $f(x_1, y) \geq f(x_2, y)$, если $x_1 \geq x_2$;
- 3) функция f непрерывна по первому аргументу относительно введенной сходимости: $f(x_n, y) \uparrow f(x, y)$, если $x_n \uparrow x$.

Тогда существует такая стационарная случайная последовательность $\{X^n\}$, удовлетворяющая уравнению (1): $X^{n+1} = UX^n = f(X^n, \xi_n)$, что

$$U^{-n}X_{n+s}(x_0) \uparrow X^s \quad \text{при } n \rightarrow \infty, \quad (20)$$

где сходимость имеет место для всех элементарных исходов.

Так как распределения X_n и $U^{-n}X_n$ совпадают, то в случае, когда сходимость случайных величин $\eta_n \uparrow \eta$ означает сближение (в подходящем смысле) их распределений (как это имеет место в $\mathcal{X} = \mathbb{R}^d$), из теоремы 1 следует также сходимость при $n \rightarrow \infty$ распределений X_n к распределению X^0 .

Замечание 1. Существенным недостатком этой теоремы является то, что она справедлива лишь для одного начального значения $X_0 = x_0$. Этот недостаток исчезает, если точка x_0 достижима с вероятностью 1 из любого $x \in \mathcal{X}$, а ξ_k независимы. В этом случае x_0 , скорее всего, будет положительным атомом, и применима также теорема 12.12 для цепей Маркова.

Предельная последовательность X^s в (20) может быть и «несобственной» (в пространствах $\mathcal{X} = \mathbb{R}^d$ она может принимать и бесконечные значения). Последовательность $\{X^s\}$ будет собственной, если с. р. п. $\{X_n\}$ удовлетворяет, например, условиям теоремы 2 или теорем § 14.5.

Доказательство теоремы 3. Обозначим

$$v_s^{-k} = f_{k+s}(x_0, \xi_{-k}^{s-1}) = U^{-k}f_{k+s}(x_0, \xi_0^{s+k-1}) = U^{-k}X_{k+s}(x_0).$$

Здесь индекс $-k$ указывает номер элемента управляющей последовательности $\{\xi_n\}_{n=-\infty}^{\infty}$, начиная с которого элементы этой последовательности используются для построения с. р. п. Индекс s указывает «момент времени», в который мы наблюдаем состояние с. р. п. Из «обратного» уравнения в (19) получаем

$$v_s^{-k-1} = f_{k+s}(f(x_0, \xi_{-k-1}), \xi_{-k}^{s-1}) \geq f_{k+s}(x_0, \xi_{-k}^{s-1}) = v_s^{-k}.$$

Это значит, что последовательность v_s^{-k} возрастает с ростом k и, стало быть, существует такая случайная величина $X^s \in \mathcal{X}$, что $v_s^{-k} = U^{-k}X(x_0, k+s) \uparrow X^s$ при $k \rightarrow \infty$. Далее, v_s^{-k} есть функция от ξ_{-k}^{s-1} . Стало быть, X^s есть функция от $\xi_{-\infty}^{s-1}$: $X^s = G(\xi_{-\infty}^{s-1})$. Поэтому $UX^s = UG(\xi_{-\infty}^{s-1}) = G(\xi_{-\infty}^s) = X^{s+1}$, что означает стационарность $\{X^s\}$. Используя «прямое» уравнение в (19), получаем, что

$$v_s^{-k-1} = f(f_{k+s}(x_0, \xi_{-k-1}^{s-2}), \xi_{-k-1}^{s-1}) = f(v_{s-1}^{-k-1}, \xi_{-k-1}^{s-1}).$$

Переход к пределу при $k \rightarrow \infty$ в силу непрерывности f дает $X^s = f(X^{s-1}, \xi_{s-1})$. Теорема доказана. ■

Пример 1, очевидно, удовлетворяет всем условиям теоремы 3 при $\mathcal{X} = [0, \infty)$, $x_0 = 0$, $f(x, y) = (x + y)^+$.

§ 4. Условия эргодичности для сжимающих в среднем преобразований, удовлетворяющих условию Липшица

В этом параграфе мы будем предполагать, что \mathcal{X} есть полное сепарабельное метрическое пространство с метрикой ρ . Рассмотрим следующие условия на итерации $f_k(x, \xi_0^{k-1}) = X_k(x)$.

(В) (ограниченность). Для некоторого $x_0 \in \mathcal{X}$ и любого $\delta > 0$ найдется такое $N = N_\delta$, что для всех $n \geq 1$

$$P(\rho(x_0, X_n(x_0)) > N) = P(\rho(x_0, f_n(x_0, \xi_0^{n-1})) > N) < \delta.$$

Очевидно, что с. р. п., удовлетворяющие условиям теорем 2, 2А или теорем § 14.5, будут удовлетворять (В).

(C) (сжатие в среднем). Функция f непрерывна по первому аргументу и, кроме того, выполнено следующее условие:

Существуют числа $m \geq 1$, $\beta > 0$ и измеримая функция $c: \mathbb{R}^m \rightarrow \mathbb{R}$ такие, что при всех x_1, x_2

$$\rho(f_m(x_1, \xi_0^{m-1}), f_m(x_2, \xi_0^{m-1})) \leq c(\xi_0^{m-1})\rho(x_1, x_2),$$

$$m^{-1}E \ln c(\xi_0^{m-1}) \leq -\beta < 0.$$

Отметим, что условия ограниченности и сжатия ((B) и (C)), вообще говоря, между собой не связаны. Пусть, например, $\mathcal{X} = \mathbb{R}$, $\rho(x, y) = |x - y|$, $f(x, y) = bx + y$,

$$X_{n+1} = bX_n + \xi_n, \quad X_0 \geq 0, \quad \xi_n \geq 0.$$

Тогда при $0 < b < 1$ условие (C) выполнено очевидным образом, так как $|f(x_1, y) - f(x_2, y)| = b|x_1 - x_2|$. В то же время условие (B) будет выполнено тогда и только тогда, когда $E \ln \xi_0 < \infty$. Действительно, если $E \ln \xi_0 = \infty$, то п. н. бесконечно много раз наступает событие $\ln \xi_k > -2k \ln b$. Но X_{n+1} распределено так же, как

$$b^{n+1}X_0 + \sum_{k=0}^n b^k \xi_k = b^{n+1}X_0 + \sum_{k=0}^n \exp\{k \ln b + \ln \xi_k\},$$

где сумма в правой части имеет неограниченно растущее с ростом n число слагаемых, больших $\exp\{-k \ln b\}$. Это означает, что $X(n+1) \xrightarrow{p} \infty$ при $n \rightarrow \infty$.

Аналогично рассматривается возможность $E \ln \xi_k < \infty$. Тот факт, что из (B), вообще говоря, не следует (C), очевиден.

Будем считать, как прежде, что «управляющая» стационарная последовательность $\{\xi_n\}_{n=-\infty}^{\infty}$ задана на всей оси, и будем обозначать через U соответствующий сохраняющий распределения оператор сдвига.

Сходимость по вероятности и п. н. последовательности \mathcal{X} -значных случайных величин $\eta_n \in \mathcal{X}$: $\eta_n \xrightarrow{p} \eta$, $\eta_n \xrightarrow{\text{п. н.}} \eta$ определяется естественным образом с помощью соотношений: $P(\rho(\eta_n, \eta) > \delta) \rightarrow 0$ при $n \rightarrow \infty$ и $P(\rho(\eta_k, \eta) > \delta \text{ хотя бы для одного } k \geq n) \rightarrow 0$ при $n \rightarrow \infty$ для любого $\delta > 0$ соответственно.

Теорема 4. Пусть выполнены условия (B), (C). Тогда существует такая стационарная последовательность $\{X^n\}$, удовлетворяющая (1): $X^{n+1} = UX^n = f(X^n, \xi_n)$, что для любого фиксированного x

$$U^{-n}X_{n+s}(x) \xrightarrow{\text{п. н.}} X^s \quad \text{при } n \rightarrow \infty. \quad (21)$$

Эта сходимость равномерна по x из любого ограниченного множества $C \subset \mathcal{X}$.

Из теоремы 2 следует слабая сходимость при $n \rightarrow \infty$ распределений $X_n(x)$ к распределению X^0 . Условие (B), очевидно, необходимо для эргодичности. Как показывает пример обобщенной авторегрессии, рассмотренный ниже, условие (C) в ряде случаев также оказывается необходимым.

Обозначим $Y_n = U^{-n}X_n(x_0)$. Нам потребуется

Лемма 1. Пусть выполнены условия (B), (C) и стационарная последовательность $\{c(\xi_{km+m-1}^{km+m-1})\}_{k=-\infty}^{\infty}$ эргодична. Тогда для любого $\delta > 0$ найдется такое n_δ , что при всех $k \geq 0$

$$P(\rho(Y_{n+k}, Y_n) < \delta \quad \text{при всех } n \geq n_\delta) \geq 1 - \delta. \quad (22)$$

Для эргодичности $\{c(\xi_{km}^{km+m-1})\}$ достаточно, чтобы преобразование T^m было метрически транзитивным.

Утверждение леммы означает, что равномерно по k с вероятностью 1 расстояние $\rho(Y_{n+k}, Y_n)$ стремится к нулю при $n \rightarrow \infty$. Соотношение (22) можно записать также в виде $P(A_\delta) \leq \delta$, где

$$A_\delta = \bigcup_{n \geq n_\delta} \{\rho(Y_{n+k}, Y_n) \geq \delta\}.$$

Доказательство. В силу (B) существует такое $N = N_\delta$, что при всех $k \geq 1$

$$P(\rho(x_0, X_k(x_0)) > N) \leq \frac{\delta}{4}.$$

Поэтому $P(A_\delta) \leq \delta/3 + P(A_\delta; \rho(x_0, \theta_{n,k}) \leq N)$. Случайная величина $\theta_{n,k} = U^{-n-k} \times X_k(x_0)$ имеет то же распределение, что и $X_k(x_0)$. Далее, в силу (C)

$$\begin{aligned} \rho(Y_{n+k}, Y_n) &\leq \rho(f_{n+k}(x_0, \xi_{-n-k}^{-1}), f_n(x_0, \xi_{-n}^{-1})) \leq \\ &\leq (\xi_{-n}^{-1}) \rho(f_{n+k-m}(x_0, \xi_{-n-k}^{-m-1}), f_{n-m}(x_0, \xi_{-n}^{-m-1})) = \\ &= c(\xi_{-m}^{-1}) \rho(U^{-n-k} X_{n+k-m}(x_0), U^{-n} X_{n-m}(x_0)). \end{aligned} \quad (23)$$

Обозначим через B_s множество чисел n , допускающих представление вида $n = lm + s$, $l = 0, 1, 2, \dots$, $0 \leq s < m$, и положим $\lambda_j = \ln c(\xi_{-jm}^{-(j-1)m-1})$, $j = 1, 2, \dots$. Тогда для $n \in B_s$ из (23) и подобных ему соотношений получаем

$$\rho(Y_{n+k}, Y_n) \leq \exp\left\{\sum_{j=1}^l \lambda_j\right\} \rho(U^{-n-k} X_{k+s}(x_0), U^{-n} X_s(x_0)), \quad (24)$$

где последнее расстояние (обозначим его через ρ) можно оценить сверху:

$$\rho \leq \rho(x_0, U^{-n-k} X_{k+s}(x_0)) + \rho(x_0, U^{-n} X_s(x_0)).$$

Случайные величины $U^{-n} X_j(x_0)$ имеют то же распределение, что и $X_j(x_0)$. В силу условия (B) найдется $N = N_\delta$ такое, что при всех $j \geq 1$

$$P(\rho(x_0, X_j(x_0)) > N) \leq \frac{\delta}{4m}.$$

Следовательно, при всех n , k , s будет выполняться $P(\rho > 2N) < \delta/2m$. На дополнительном множестве $\{\rho \leq 2N\}$ правая часть (23) не превосходит $2N \exp\left\{\sum_{j=1}^l \lambda_j\right\}$.

Так как $E\lambda_j \leq -m\beta < 0$ и последовательность $\{\lambda_j\}$ метрически транзитивна, то в силу эргодической теоремы 15.6 $\sum_{j=1}^l \lambda_j < -m\beta l/2$ при всех $l \geq l(\omega)$, где $l(\omega)$ — собственная случайная величина. Выберем l_1 и l_2 так, чтобы выполнялись неравенства

$$-\frac{m\beta l_1}{2} < \ln \delta - \ln 2N, \quad P(l(\omega) > l_2) < \frac{\delta}{2}.$$

Тогда, полагая $l_\delta = \max(l_1, l_2)$, $n_\delta = ml_\delta$, $A_\delta^s = \bigcup_{n \geq n_\delta, n \in B_s} \{\rho(Y_{n+k}, Y_n) \geq \delta\}$, получим

$$P(A_\delta^s) \leq P(\rho > 2N) + P(A_\delta^s; \rho \leq N) \leq \frac{\delta}{2m} + P\left(\bigcup_{l \geq l_\delta} \left\{2N \exp\left\{-\sum_{j=0}^l \lambda_j\right\} \geq \delta\right\}\right).$$

Но пересечение события под знаком последней вероятности с событием $\{l_\delta \geq l(\omega)\}$ пусто. Поэтому событие под знаком вероятности вложено в событие $\{l(\omega) > l_\delta\}$ и

$$P(A_\delta^s) \leq \frac{\delta}{m}, \quad P(A_\delta) \leq \sum_{s=0}^{m-1} P(A_\delta^s) \leq \delta.$$

Лемма доказана. ■

Лемма 2 (полнота \mathcal{X} относительно сходимости по вероятности). Пусть \mathcal{X} — полное метрическое пространство. Если последовательность \mathcal{X} -значных случайных элементов η_n такова, что для любого $\delta > 0$

$$P_n \equiv \sup_{k \geq 0} P(\rho(\eta_{n+k}, \eta_n) > \delta) \rightarrow 0$$

при $n \rightarrow \infty$, то существует случайный элемент $\eta \in \mathcal{X}$ такой, что $\eta_n \xrightarrow{P} \eta$ ($P(\rho(\eta_n, \eta) > \delta) \rightarrow 0$ при $n \rightarrow \infty$).

Доказательство. Для заданных ε и δ выберем n_k , $k = 0, 1, \dots$, так, чтобы выполнялось

$$\sup_s P(\rho(\eta_{n_k+s}, \eta_{n_k}) > 2^{-k}\delta) < \varepsilon 2^{-k},$$

и положим для краткости $\eta_{n_k} = \zeta_k$. Рассмотрим множество $D = \bigcap_{k=0}^{\infty} D_k$, $D_k = \{\omega: \rho(\zeta_{k+1}, \zeta_k) \leq 2^{-k}\delta\}$. Имеем $P(D) > 1 - 2\varepsilon$, и для любого $\omega \in D$ выполняется $\rho(\zeta_{k+s}(\omega), \zeta_k(\omega)) < \delta 2^{k-1}$ при всех $s \geq 1$. Следовательно, последовательность $\zeta_k(\omega)$ фундаментальна в \mathcal{X} , и существует $\eta = \eta(\omega) \in \mathcal{X}$ такое, что $\zeta_k(\omega) \rightarrow \eta(\omega)$. Так как ε произвольно, то сказанное означает, что $\zeta_k \xrightarrow{\text{п. н.}} \eta$ при $k \rightarrow \infty$,

$$P(\rho(\zeta_0, \eta) > 2\delta) \leq P\left(\bigcup_{k=0}^{\infty} \rho(\zeta_{k+1}, \zeta_k) > 2^{-k}\delta\right) \leq \sum_{k=0}^{\infty} P(\rho(\zeta_{k+1}, \zeta_k) > 2^{-k}\delta) \leq 2\varepsilon.$$

Поэтому для любого $n \geq n_0$

$$P(\rho(\eta_n, \eta) > 3\delta) \leq P(\rho(\eta_n, \eta_{n_0}) > \delta) + P(\rho(\zeta_0, \eta)) \leq 3\varepsilon.$$

Так как ε и δ произвольны, то лемма доказана. ■

Доказательство теоремы 4. Из леммы 1 следует, что

$$\sup_k P(\rho(Y_{n+k}, Y_n) > \delta) \rightarrow 0 \text{ при } n \rightarrow \infty.$$

Это означает, что последовательность Y_n фундаментальна относительно сходимости по вероятности, и в силу леммы 2 существует такая случайная величина X^0 , что

$$Y_n \xrightarrow{P} X^0, \tag{25}$$

$$U^{-n} X_{n+s}(x_0) = U^s (U^{-n-s} X_{n+s}(x_0)) = U^s Y_{n+s} \rightarrow U^s X^0 \equiv X^s.$$

В силу непрерывности f

$$U^{-n}X_{n+s+1}(x_0) = U^{-n}f(X_{n+s}(x_0), \xi_{n+s}) = f(U^{-n}X_{n+s}(x_0), \xi_s) \xrightarrow{p} f(X^s, \xi_s) = X^{s+1}.$$

Мы доказали требуемую сходимость для фиксированного начального условия x_0 . При произвольном $x \in C_N = \{z: \rho(x_0, z) \leq N\}$ имеем

$$\rho(U^{-n}X_n(x), X^0) \leq \rho(U^{-n}X_n(x), U^{-n}X_n(x_0)) + \rho(U^{-n}X_n(x_0), X^0), \quad (26)$$

где первое слагаемое в правой части равномерно по $x \in C_N$ сходится по вероятности к нулю. Для $n = lm$ это следует из неравенства (см. условие (C))

$$\rho(U^{-n}X_n(x), U^{-n}X_n(x_0)) \leq N \exp\left\{\sum_{j=1}^l \lambda_j\right\} \quad (27)$$

и приведенных выше рассуждений. Для $n = lm + s$, $m > s > 0$, справедливы аналогичные соотношения. Отсюда и из (25), (26) следует, что $U^{-n}X_{n+s}(x) \xrightarrow{p} X^s = U^s X^0$ равномерно по $x \in C_N$. Это доказывает утверждение теоремы относительно сходимости по вероятности.

Докажем теперь сходимость с вероятностью 1. Для этого следует повторить рассуждения в доказательстве леммы 1, но оценивать не $\rho(Y_{n+k}, Y_n)$, а $\rho(X^0, U^{-n}X_n(x))$. Считая для простоты, что $s = 0$ (n кратно m), получим (аналогично (24)), что для любого x

$$\rho(X^0, U^{-n}X_n(x)) \leq \rho(x, U^{-n}X^0) \exp\left\{\sum_{j=1}^l \lambda_j\right\}. \quad (28)$$

Остальные рассуждения леммы 1 остаются без изменений. Они означают, что для любого $\delta > 0$ при достаточно больших n_δ

$$P\left(\bigcup_{n \geq n_\delta} \{\rho(X^0, U^{-n}X_n(x)) > \delta\}\right) < \delta.$$

Теорема доказана. ■

Пример 3. Обобщенная авторегрессия. Пусть $\mathcal{X} = \mathbb{R}$. Процесс обобщенной авторегрессии определим с помощью соотношений

$$X_{n+1} = G(\zeta_n F(X_n) + \eta_n), \quad (29)$$

где F, G — функции $\mathbb{R} \rightarrow \mathbb{R}$, $\xi_n = (\zeta_n, \eta_n)$ есть управляющая последовательность, так что $\{X_n\}$ есть с. р. п. с функцией $f(x, y) = G(y_1 F(x) + y_2)$, $y = (y_1, y_2) \in \mathbb{R}^2 = \mathcal{Y}$.

Если функции F и G не убывают и непрерывны слева, $G(x) \geq 0$ при всех $x \in \mathbb{R}$, элементы ζ_n неотрицательны, то процесс (29) удовлетворяет условиям теоремы 3 и, следовательно, $U^{-n+s}X_n(0) \uparrow X^s$ с вероятностью 1 (при $n \rightarrow \infty$). Чтобы установить сходимость к собственной стационарной последовательности X^s , надо доказать равномерную по n ограниченность по вероятности $X_n(0)$ (см. ниже).

Выясним теперь, при каких условиях (29) будет удовлетворять условиям теоремы 4. Предположим, что функции F и G удовлетворяют условию Липшица

$$|G(x_1) - G(x_2)| \leq c_G |x_1 - x_2|, \quad |F(x_1) - F(x_2)| \leq c_F |x_1 - x_2|.$$

Тогда

$$|f(x_1, \xi_0) - f(x_2, \xi_0)| \leq c_G |\zeta_0 (F(x_1) - F(x_2))| \leq c_F c_G |\zeta_0| |x_1 - x_2|. \quad (30)$$

Теорема 5. При сделанных предположениях последовательность (29) будет удовлетворять условию (C) при $\rho(x_1, x_2) = |x_1 - x_2|$, если

$$\ln c_G c_F + E \ln |\zeta_0| < 0. \quad (31)$$

Последовательность (29) будет удовлетворять условию (B), если выполнено (31) и, кроме того,

$$E(\ln |\eta_0|)^+ < \infty. \quad (32)$$

При выполнении (31), (32) последовательность (29) помимо этого имеет также стационарную мажоранту, т. е. существует такая стационарная последовательность M_n (зависящая от X_0), что $|X_n| \leq M_n$ при всех n .

Доказательство теоремы 5. Выполнение (C) в силу (30) очевидно. Докажем (B). Для этого мы построим стационарную мажоранту для $|X_n|$. Это можно сделать, пользуясь результатами теорем 2, 2A. В нашем случае проще действовать непосредственно, пользуясь неравенствами

$$|G(x)| \leq |G(0)| + c_G |x|, \quad |F(x)| \leq |F(0)| + c_F |x|,$$

где мы будем считать для простоты, что $G(0)$ и $F(0)$ конечны. Тогда

$$\begin{aligned} |X_{n+1}| &\leq |G(0)| + c_G |\zeta_n| |F(X_n)| + c_G |\eta_n| \leq \\ &\leq |G(0)| + c_G c_F |\zeta_n| |X_n| + c_G |\zeta_n| |F(0)| + c_G |\eta_n| = \beta_n |X(n)| + \gamma_n, \end{aligned}$$

где

$$\begin{aligned} \beta_n &= c_G c_F |\zeta_n| \geq 0, \\ \gamma_n &= |G(0)| + c_G |\zeta_n| |F(0)| + c_G |\eta_n|, \\ E \ln \beta_n &< 0, \quad E(\ln \gamma_n)^+ < \infty. \end{aligned}$$

Отсюда получаем ($x = X_0$)

$$\begin{aligned} |X_{n+1}| &\leq |x| \prod_{j=0}^n \beta_j + \sum_{l=0}^{n-1} \left(\prod_{j=n-l}^n \beta_j \right) \gamma_{n-l-1} + \gamma_n, \\ U^{-n} |X_{n+1}| &\leq |x| \prod_{j=-n}^0 \beta_j + \sum_{l=0}^{\infty} \left(\prod_{j=-l}^0 \beta_j \right) \gamma_{-l-1} + \gamma_0. \end{aligned} \quad (33)$$

Обозначим

$$\alpha_i = \ln \beta_j, \quad S_l = \sum_{j=-l}^0 \alpha_j.$$

В силу усиленного закона больших чисел значения $S_l - al$, где $2a = E\alpha_j < 0$, лишь конечное число раз превосходят нуль. Поэтому для всех l , за исключением этих значений,

$$\prod_{j=-l}^0 \beta_j < e^{al}.$$

С другой стороны, γ_{-l-1} лишь конечное число раз превышают уровень l . Это означает, что ряд в (33) (мы обозначим его через R) сходится с вероятностью 1. Кроме того,

$$S = \sup_{k \geq 0} S_k \geq S_n$$

является собственной случайной величиной. Мы получаем в результате, что для всех n

$$U^{-n}|X_{n+1}| \leq |x|e^S + R + \gamma_0,$$

где все слагаемые в правой части суть собственные случайные величины. Требуемая мажоранта

$$M_n = U^{n-1}(|x|e^S + R + \gamma_0)$$

построена. Отсюда следует выполнение (В). Теорема доказана. \blacksquare

Утверждение теоремы 5 можно перенести и на многомерный случай $\mathcal{X} = \mathbb{R}^d$, $d > 1$ (см. [6]).

Отметим, что условия (31), (32) являются в известном смысле необходимыми не только для сходимости $U^{-n+s}X_n(x) \rightarrow X^s$, но и для одной лишь ограниченности $X_n(x)$ (или X^0). Это обстоятельство лучше всего проиллюстрировать в случае, когда $F(x) = G(x) = x$. В этом случае $U^{-n}X_{n+s+1}(x)$ и X^{s+1} допускают явные представления

$$U^{-n}X_{n+s+1}(x) = x \prod_{j=-n}^s \zeta_j + \sum_{l=0}^{n+s} \prod_{j=s-l}^s \zeta_j \eta_{s-l-1} + \eta_s,$$

$$X^{s+1} = \sum_{l=0}^{\infty} \prod_{j=s-l}^s \zeta_j \eta_{s-l-1} + \eta_s.$$

Предположим, что $E \ln \zeta \geq 0$, $\eta \equiv 1$, и положим $s = 0$, $z_j = \ln \zeta_j$, $Z_l = \sum_{j=-l}^0 z_j$. Тогда $X^1 = 1 + \sum_{l=0}^{\infty} e^{Z_l}$, $\sum_{l=0}^{\infty} I(Z_l \geq 0) = \infty$ с вероятностью 1 и, следовательно, $X^1 = \infty$, $X_n \rightarrow \infty$ с вероятностью 1.

Если $E[\ln \eta]^+ = \infty$, $\zeta = b < 1$, то $X^1 = \eta_0 + b \sum_{l=0}^{\infty} \exp\{y_{-l-1} + l \ln b\}$, где $y_j = \ln \eta_j$; число появлений событий $\{y_{-l-1} > -l \ln b\}$ бесконечно с вероятностью 1. Это означает, что $X^1 = \infty$, $X_n \rightarrow \infty$ с вероятностью 1.

Случайные процессы с непрерывным временем

§ 1. Общие определения

Определение 1. Случайным процессом¹⁾ называется семейство случайных величин $\xi(t) = \xi(t, \omega)$, заданных на одном вероятностном пространстве $\langle \Omega, \mathfrak{F}, P \rangle$ и зависящих от параметра t , принимающего значения из некоторого множества T .

Обозначать случайный процесс будем символами $\{\xi(t), t \in T\}$.

Последовательности независимых случайных величин ξ_1, ξ_2, \dots , рассматривавшиеся в предыдущих параграфах, являются случайными процессами, для которых $T = \{1, 2, \dots\}$. То же самое справедливо и относительно сумм S_1, S_2, \dots слагаемых ξ_1, ξ_2, \dots . Случайными процессами являются также последовательности $\{X_n, n = 0, 1, \dots\}$ величин, связанных в цепь Маркова, мартингалы $\{X_n, n \in \mathcal{N}\}$, стационарные и стохастические рекурсивные последовательности, описанные в предыдущих главах. Такие процессы, у которых множество T можно отождествить со всей или с частью последовательности $\{\dots, -1, 0, 1, \dots\}$, обычно называются процессами с *дискретным временем* или *случайными последовательностями*.

Если T совпадает с некоторым числовым интервалом $T = [a, b]$ (это может быть также вся прямая $-\infty < t < \infty$ или полупрямая $t \geq 0$), то совокупность $\{\xi(t), t \in T\}$ называют процессом с *непрерывным временем*.

Простым примером таких процессов могут служить процессы восстановления $\{\eta(t), t \geq 0\}$, описанные в гл. 9.

В этой главе мы будем рассматривать лишь процессы с *непрерывным временем*. Интерпретация параметра t как времени не обязательна. Она возникла исторически, поскольку в большинстве естественнонаучных задач, которые привели к появлению понятия случайного процесса, параметр t был временем, а значение $\xi(t)$ было тем, что наблюдалось в момент времени t .

Как случайный процесс можно рассматривать, например, движение молекулы газа во времени, уровень воды в водохранилище, колебания крыла самолета и т. п.

Процессом с непрерывным временем является, например, случайная функция

$$\xi(t) = \sum_{k=0}^{\infty} \frac{\xi_k}{2^k} \sin kt, \quad t \in [0, 2\pi],$$

где ξ_k независимы и одинаково распределены.

Рассмотрим некоторый случайный процесс $\{\xi(t), t \in T\}$.

¹⁾ Наряду с термином «случайный процесс» часто используются также названия «вероятностный» или «стохастический» процесс.

Если фиксировать $\omega \in \Omega$, то мы получим функцию $\xi(t)$, $t \in T$, которую часто называют *выборочной* функцией или *траекторией* процесса. Таким образом, в роли случайных величин здесь выступают *функции*. Как и раньше, можно было бы рассматривать здесь *выборочное* вероятностное пространство. Его можно построить, например, так. Рассмотрим пространство \mathcal{X} функций $x(t)$, $t \in T$, в котором лежат траектории $\xi(t)$. Пусть далее $\mathfrak{B}_{\mathcal{X}}^T$ есть σ -алгебра подмножеств из \mathcal{X} , порожденная множествами вида

$$C = \{x \in \mathcal{X} : x(t_1) \in B_1, \dots, x(t_n) \in B_n\} \quad (1)$$

для любых n , любых t_1, \dots, t_n из T и любых борелевских множеств B_1, \dots, B_n . Множества такого вида называются *цилиндрическими*; всевозможные конечные объединения цилиндрических множеств образуют алгебру, порождающую $\mathfrak{B}_{\mathcal{X}}^T$. Если процесс $\xi(t, \omega)$ задан, то он определяет измеримое отображение $\langle \Omega, \mathfrak{F} \rangle$ в $\langle \mathcal{X}, \mathfrak{B}_{\mathcal{X}}^T \rangle$, так как, очевидно, $\xi^{-1}(C) = \{\omega : \xi(\cdot, \omega) \in C\} \in \mathfrak{F}$ для любого цилиндрического множества C и, следовательно, $\xi^{-1}(B) \in \mathfrak{F}$ для любого $B \in \mathfrak{B}_{\mathcal{X}}^T$. Это отображение индуцирует распределение P_{ξ} на $\langle \mathcal{X}, \mathfrak{B}_{\mathcal{X}}^T \rangle$, определяемое равенством $P_{\xi}(B) = P(\xi^{-1}(B))$. Тройка $\langle \mathcal{X}, \mathfrak{B}_{\mathcal{X}}^T, P_{\xi} \rangle$ и называется *выборочным вероятностным пространством*. В этом пространстве элементарный исход « ω » отождествляется с траекторией процесса, а мера P_{ξ} называется *распределением процесса* ξ .

Если же, рассматривая процесс $\{\xi(t)\}$, фиксировать значения времени t_1, t_2, \dots, t_n , то получим многомерную случайную величину $(\xi(t_1, \omega), \dots, \xi(t_n, \omega))$. Распределения таких величин называют *конечномерными распределениями* процесса.

В качестве пространства \mathcal{X} в теории случайных процессов с континуальным множеством T значений времени t чаще других рассматриваются следующие пространства функций.

1. Пространство всех функций на T

$$\mathcal{X} = \mathbb{R}^T = \prod_{t \in T} \mathbb{R}_t,$$

где \mathbb{R}_t есть экземпляры вещественной прямой $(-\infty, \infty)$. Это пространство обычно рассматривают в паре с σ -алгеброй $\mathfrak{B}_{\mathbb{R}}^T$ подмножеств из \mathbb{R}^T , порожденной цилиндрическими множествами.

2. $C(T)$ — пространство всех непрерывных на T функций (будем писать $C(a, b)$, если $T = [a, b]$). В этом пространстве наряду с σ -алгеброй $\mathfrak{B}_{\mathbb{R}}^T$, порожденной цилиндрическими множествами из $C(T)$ (эта σ -алгебра уже, чем аналогичная σ -алгебра в \mathbb{R}^T), часто рассматривают также σ -алгебру $\mathfrak{B}_{C(T)}$ (борелевскую σ -алгебру), порожденную множествами, открытыми относительно равномерной метрики

$$\rho(x, y) = \sup_{t \in T} |y(t) - x(t)|, \quad x, y \in C(T).$$

Оказывается, что в пространстве $C(T)$ всегда $\mathfrak{B}_{C(T)} = \mathfrak{B}_{\mathbb{R}}^{T(2)}$.

3. Пространство $D(T)$ функций, у которых в каждой точке t существуют пределы $x(t-0)$ и $x(t+0)$, причем значение $x(t)$ совпадает либо с $x(t-0)$, либо с $x(t+0)$. Если $T = [a, b]$, то предполагается также, что $x(a) = x(a+0)$, $x(b) = x(b-0)$. Это пространство называют часто *пространством функций без разрывов второго рода*³⁾. Пространство функций, для которых в остальных точках

²⁾ См., например, [7].

³⁾ Разрыв второго рода связан либо с учащающимися незатухающими колебаниями, либо с уходом на бесконечность.

$x(t) = x(t-0)$ ($x(t) = x(t+0)$), будем обозначать $D_-(T)$ ($D_+(T)$). Пространство $D_+(T)$ ($D_-(T)$) будем называть также пространством функций, непрерывных справа (слева). Например, траектории процессов восстановления, описанных в гл. 9, принадлежат $D_+(0, \infty)$.

В пространстве $D(T)$ также можно строить борелевскую σ -алгебру относительно подходящей метрики, но мы ограничимся использованием σ -алгебры \mathfrak{B}_D^T подмножеств из $D(T)$.

После проведенных рассмотрений можно сформулировать следующее эквивалентное определение случайного процесса. Пусть \mathcal{X} — заданное пространство функций и \mathfrak{G} есть σ -алгебра его подмножеств, содержащая σ -алгебру $\mathfrak{B}_{\mathcal{X}}^T$ цилиндрических множеств.

Определение 2. Случайным процессом $\xi(t) = \xi(t, \omega)$ называется измеримое (по ω) отображение $\langle \Omega, \mathfrak{F}, P \rangle$ на $\langle \mathcal{X}, \mathfrak{G}, P_{\xi} \rangle$ (каждому ω ставится в соответствие $\xi(t) = \xi(t, \omega)$ так, что $\xi^{-1}(G) = \{\omega: \xi(\cdot) \in G\} \in \mathfrak{F}$ для $G \in \mathfrak{G}$).

P_{ξ} называется *распределением процесса*.

Условие $\mathfrak{B}_{\mathcal{X}}^T \subset \mathfrak{G}$ нужно для того, чтобы были определены вероятности цилиндрических множеств и, в частности, вероятности $P(\xi(t) \in B)$, $B \in \mathfrak{B}_{\mathcal{X}}^T$, означающие, что $\xi(t)$ есть случайные величины.

До сих пор как бы предполагалось, что процесс $\xi(t)$ задан и нам известно, что его траектории лежат в \mathcal{X} . Однако это бывает весьма редко. Чаще процесс пытаются описать теми или иными сведениями о его распределении. Можно задавать, например, конечномерные распределения. Из теоремы Колмогорова о согласованных распределениях⁴⁾ (см. Приложение 2) следует, что задание согласованных конечномерных распределений определяет однозначно распределение процесса P_{ξ} на пространстве $\langle \mathbb{R}^T, \mathfrak{B}_{\mathbb{R}^T}^T \rangle$. Эту теорему можно рассматривать как теорему существования случайного процесса в $\langle \mathbb{R}^T, \mathfrak{B}_{\mathbb{R}^T}^T \rangle$ с заданными конечномерными распределениями.

Однако пространство $\langle \mathbb{R}^T, \mathfrak{B}_{\mathbb{R}^T}^T \rangle$ не очень удобно для изучения случайных процессов. Дело в том, что далеко не все часто употребляемые в анализе соотношения для функций порождают события — т. е. множества, которые принадлежат σ -алгебре $\mathfrak{B}_{\mathbb{R}^T}^T$ и о вероятности которых имеет смысл говорить. Ведь на основании определения можно с уверенностью сказать, что событиями являются лишь элементы σ -алгебры, порожденной событиями $\{\xi(t) \in B\}$, $t \in T$, где B — борелевские множества. Множество $\{\sup_{t \in T} \xi(t) < c\}$, например, может и не быть событием, так как мы знаем лишь представление его в виде $\bigcap_{t \in T} \{\xi(t) < c\}$, т. е. в виде пересечения *несчетного* числа измеримых множеств, если T есть интервал числовой оси.

Имеет место и другое неудобство — распределение P_{ξ} на $\langle \mathbb{R}^T, \mathfrak{B}_{\mathbb{R}^T}^T \rangle$ не определяет однозначно свойств траекторий $\xi(t)$. Этот факт связан с тем, что пространство \mathbb{R}^T очень богато и принадлежность $x(\cdot)$ цилиндрическому множеству вида (1) не несет в себе никакой информации о поведении $x(t)$ в точках t ,

⁴⁾ Напомним определение согласованных распределений. Пусть \mathbb{R}_t , $t \in T$, — вещественные прямые, \mathfrak{B}_t — σ -алгебры борелевских множеств на \mathbb{R}_t . Пусть $T_n = \{t_1, \dots, t_n\}$ — конечное подмножество чисел из T . Конечномерное распределение $(\xi(t_1, \omega), \dots, \xi(t_n, \omega))$ есть распределение P_{T_n} на $(\mathbb{R}^{T_n}, \mathfrak{B}^{T_n})$, где $\mathbb{R}^{T_n} = \prod_{t \in T_n} \mathbb{R}_t$, $\mathfrak{B}^{T_n} = \prod_{t \in T_n} \mathfrak{B}_t$. Пусть даны два конечных подмножества T' и T'' из T и $(\mathbb{R}', \mathfrak{B}')$, $(\mathbb{R}'', \mathfrak{B}'')$ — соответствующие подпространства $(\mathbb{R}^T, \mathfrak{B}^T)$. Распределения $P_{T'}$ и $P_{T''}$ на $(\mathbb{R}', \mathfrak{B}')$, $(\mathbb{R}'', \mathfrak{B}'')$ согласованы, если их проекции на общую часть пространств \mathbb{R}' и \mathbb{R}'' (если таковая существует) совпадают.

отличных от t_1, \dots, t_n . Это же относится и к произвольным множествам A из $\mathfrak{B}_{\mathbb{R}}^T$ — грубо говоря, принадлежность $x(\cdot) \in A$ может определять значения $x(t)$ разве лишь в счетном числе точек. (Ниже мы увидим, что даже такое множество, как $\{x(t) \equiv 0\}$, не принадлежит $\mathfrak{B}_{\mathbb{R}}^T$.) Чтобы задать всю траекторию процесса, недостаточно задать распределение на $\mathfrak{B}_{\mathbb{R}}^T$; эту σ -алгебру надо расширять.

Прежде чем приводить соответствующий пример, дадим следующее

Определение 3. Процессы $\xi(t)$ и $\eta(t)$ называются *эквивалентными* (или *стохастически эквивалентными*), если $P(\xi(t) = \eta(t)) = 1$ при всех $t \in T$. Процесс η будем называть *модификацией* ξ .

Очевидно, что конечномерные распределения для эквивалентных процессов совпадают и, стало быть, совпадают распределения P_{ξ} и P_{η} в $\langle \mathbb{R}^T, \mathfrak{B}_{\mathbb{R}}^T \rangle$.

Пример 1. Обозначим

$$x_a(t) = \begin{cases} 0, & t \neq a, \\ 1, & t = a \end{cases}$$

и пополним $\mathfrak{B}_{\mathbb{R}}^T$ элементами $x_a(t)$, $a \in [0, 1]$, и элементом $x^0(t) \equiv 0$. Пусть $\gamma \in U_{0,1}$. Рассмотрим два случайных процесса $\xi_0(t)$ и $\xi_1(t)$, определенных следующим образом: $\xi_0(t) \equiv x^0(t)$, $\xi_1(t) = x_{\gamma}(t)$. Тогда, очевидно,

$$P(\xi_0(t) = \xi_1(t)) = P(\gamma \neq t) = 1,$$

процессы ξ_0 и ξ_1 эквивалентны и, следовательно, их распределения в $\langle \mathbb{R}^T, \mathfrak{B}_{\mathbb{R}}^T \rangle$ совпадают. Однако мы видим, что траектории этих процессов существенно различны.

На приведенном примере легко понять, что множество всех непрерывных функций $C(T)$, множество $\left\{ \sup_{t \in [0,1]} x(t) < x \right\}$, элемент $\{x(t) \equiv 0\}$ и многие другие

не принадлежат $\mathfrak{B}_{\mathbb{R}}^T$. Действительно, если допустить противное, например, что $C(T) \in \mathfrak{B}^T$, то мы в силу эквивалентности ξ_0 и ξ_1 должны были бы иметь $P(\xi_0 \in C(0,1)) = P(\xi_1 \in C(0,1))$, в то время как первая из этих вероятностей равна 1, а вторая — нулю.

Простейший путь преодоления отмеченных выше трудностей и неудобств состоит в том, чтобы задавать процессы в пространствах $C(T)$ или $D(T)$, если это возможно. Если, например, $\xi(t) \in C(T)$, $\eta(t) \in C(T)$ и они эквивалентны, то траектории этих процессов будут с вероятностью 1 полностью совпадать, так как в этом случае

$$\bigcap_{t \text{ рациональны}} \{\xi(t) = \eta(t)\} = \bigcap_{t \in T} \{\xi(t) = \eta(t)\} = \{\xi(t) = \eta(t) \text{ при всех } t \in T\},$$

где вероятность события в левой части равенства определена (это есть вероятность пересечения счетного числа событий) и равна 1. Аналогично обстоит дело, например, с событиями $\left\{ \sup_{t \in T} \xi(t) < c \right\} = \bigcap_{t \in T} \{\xi(t) < c\}$.

Точно такие же рассуждения справедливы для пространств $D(T)$, поскольку каждый элемент $x(\cdot)$ из D однозначно определяется значениями $x(t)$ на счетном всюду плотном множестве значений t (например, на множестве рациональных чисел).

Итак, допустим, что удалось тем или иным способом установить, что для исходного процесса $\xi(t)$ (пусть он задан на $\langle \mathbb{R}^T, \mathfrak{B}_{\mathbb{R}}^T \rangle$) существует его непрерывная модификация, т.е. эквивалентный ему процесс $\eta(t)$ такой, что траектории $\eta(t)$

с вероятностью 1 непрерывны (или принадлежат $D(t)$). Сказанное означает прежде всего, что мы расширили каким-то образом σ -алгебру $\mathfrak{B}_{\mathbb{R}}^T$ присоединением, скажем, множества $C(T)$ и рассматриваем теперь распределение ξ на σ -алгебре $\tilde{\mathfrak{B}}^T = \sigma(\mathfrak{B}_{\mathbb{R}}^T, C(T))$ (иначе приведенные слова теряют смысл). Но распределение ξ может быть продолжено с $\langle \mathbb{R}^T, \mathfrak{B}_{\mathbb{R}}^T \rangle$ на $\langle \mathbb{R}^T, \tilde{\mathfrak{B}}^T \rangle$ не однозначно (это мы видели в примере 1; продолжение можно задать, положив, например, $P(\xi \in C(T)) = 0$). Сказанное выше о процессе η означает, что существует продолжение P_η такое, что $P_\eta(C(T)) = P(\eta \in C(T)) = 1$.

Далее, часто бывает лучше совсем не иметь дела с неудобным пространством $\langle \mathbb{R}^T, \mathfrak{B}_{\mathbb{R}}^T \rangle$. Для этого можно задать распределение процесса η на суженном пространстве $\langle C(T), \mathfrak{B}_C^T \rangle$. Ведь ясно, что $\mathfrak{B}_C^T \subset \tilde{\mathfrak{B}}^T = \sigma(\mathfrak{B}_{\mathbb{R}}^T, C(T))$, $\mathfrak{B}_C^T = \tilde{\mathfrak{B}}^T \cap C(T)$ (первая из этих σ -алгебр порождена множествами вида (1), пересеченными с $C(T)$). Поэтому, рассматривая распределение η , сосредоточенное на $C(T)$, можно ограничиться сужением пространства $\langle \mathbb{R}^T, \tilde{\mathfrak{B}}^T \rangle$ до $\langle C(T), \mathfrak{B}_C^T \rangle$, определив на последнем вероятность как значение $P_\eta(A) = P(\eta \in A)$, $A \in \mathfrak{B}_C^T \subset \tilde{\mathfrak{B}}^T$. Тем самым построен процесс η с непрерывными выборочными траекториями, эквивалентный исходному процессу ξ (если их распределения рассматривать в $\langle \mathbb{R}^T, \mathfrak{B}_{\mathbb{R}}^T \rangle$).

Чтобы осуществлять эту конструкцию, надо теперь уметь выяснять по распределению процесса ξ , существует для него непрерывная модификация η или нет.

Прежде чем переходить к формулировке и доказательству соответствующих теорем, отметим еще раз, что названные выше трудности носят главным образом *математический характер*, т.е. связаны с математической моделью случайного процесса. В реальных прикладных задачах, как правило, бывает заранее ясно, является рассматриваемый процесс непрерывным или нет. Если он «физически» непрерывен и мы хотим построить адекватную модель, то, конечно, из всех эквивалентных модификаций процесса надо взять непрерывную.

Точно такие же рассуждения остаются справедливыми, если вместо непрерывных траекторий рассматривать траектории из $D(T)$. Существо проблемы при этом остается тем же: трудности устраняются, если есть возможность описать всю траекторию процесса $\xi(\cdot)$ значениями $\xi(t)$ лишь на каком-нибудь счетном множестве значений t . Процессы, обладающие таким свойством, будем называть *регулярными*.

§ 2. Условия регулярности процессов

Выясним сначала условия, при которых процесс имеет непрерывную модификацию. В качестве множества T , не ограничивая общности, будем рассматривать отрезок $T = [0, 1]$.

Очень простой критерий существования непрерывной модификации, принадлежащей Колмогорову, основан на знании лишь *двумерных* распределений $\xi(t)$.

Теорема 1 (Колмогоров). Пусть $\xi(t)$ — случайный процесс, заданный на $\langle \mathbb{R}, \mathfrak{B}_{\mathbb{R}}^T \rangle$ при $T = [0, 1]$. Если при всех $t, t+h$ из отрезка $[0, 1]$

$$E|\xi(t+h) - \xi(t)|^a \leq c|h|^{1+b} \quad (2)$$

при каких-нибудь $a > 0$, $b > 0$, $c < \infty$ то $\xi(\cdot)$ имеет непрерывную модификацию.

Мы получим это утверждение в качестве следствия более общей теоремы, условия которой воспринимаются несколько труднее, но имеют тот же смысл, что и (2).

Теорема 2. Пусть при всех $t, t+h$ из отрезка $[0, 1]$

$$P(|\xi(t+h) - \xi(t)| > \varepsilon(h)) \leq q(h),$$

где $\varepsilon(h), q(h)$ — убывающие четные функции от h , обладающие свойством

$$\sum_{n=1}^{\infty} \varepsilon(2^{-n}) < \infty, \quad \sum_{n=1}^{\infty} 2^n q(2^{-n}) < \infty.$$

Тогда $\xi(\cdot)$ имеет непрерывную модификацию.

Доказательство. Воспользуемся приближениями процесса $\xi(t)$ с помощью непрерывных процессов. Положим

$$t_{n,r} = r2^{-n}, \quad r = 0, 1, \dots, 2^n,$$

$$\xi_n(t) = \xi(t_{n,r}) + 2^n(t - t_{n,r})[\xi(t_{n,r+1}) - \xi(t_{n,r})]$$

при $t \in [t_{n,r}, t_{n,r+1}]$. Из рис. 15 видно, что

$$|\xi_{n+1}(t) - \xi_n(t)| \leq \left| \xi(t_{n+1,2r+1}) - \frac{1}{2}[\xi(t_{n+1,2r}) + \xi(t_{n+1,2r+2})] \right| \leq \frac{1}{2}(\alpha + \beta),$$

где $\alpha = |\xi(t_{n+1,2r+1}) - \xi(t_{n+1,2r})|$, $\beta = |\xi(t_{n+1,2r+1}) - \xi(t_{n+1,2r+2})|$. Отсюда следует, что

$$Z_n \equiv \max_{t \in [t_{n,r}, t_{n,r+1}]} |\xi_{n+1}(t) - \xi_n(t)| \leq \frac{1}{2}(\alpha + \beta),$$

$$P(Z_n > \varepsilon(2^{-n})) \leq P(\alpha > \varepsilon(2^{-n})) + P(\beta > \varepsilon(2^{-n})) \leq 2q(2^{-n}).$$

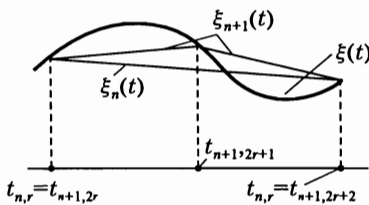


Рис. 15.

(Отметим, что так как процессы $\xi_n(t)$ непрерывны, то здесь $\{Z_n > \varepsilon\} \in \mathfrak{B}_{\mathbb{R}}^T$. Если в определении Z_n вместо $\xi_n(t)$ подставить процессы более общего вида, то такое вложение может и не иметь места.) Так как здесь всего 2^n отрезков типа $[t_{n,r}, t_{n,r+1}]$, $r = 0, 1, \dots, 2^n - 1$, то

$$P\left(\max_{t \in [0,1]} |\xi_{n+1}(t) - \xi_n(t)| > \varepsilon(2^{-n})\right) \leq 2^{n+1} q(2^{-n}).$$

Поскольку $\sum_{n=1}^{\infty} 2^n q(2^{-n}) < \infty$, то по критерию Бореля—Кантелли для п.в. ω (т.е. для $\omega \in A$, $P(A) = 1$) существует $n(\omega)$ такое, что при всех $n \geq n(\omega)$

$$\max_{t \in [0,1]} |\xi_{n+1} - \xi_n(t)| \equiv \rho(\xi_{n+1}, \xi_n) < \varepsilon(2^{-n}).$$

Отсюда следует п.н. фундаментальность последовательности ξ_n , поскольку

$$\rho(\xi_n, \xi_m) \leq \varepsilon_n \equiv \sum_{k=n}^{\infty} \varepsilon(2^{-k}) \rightarrow 0$$

при $n \rightarrow \infty$ и всех $m > n$, $\omega \in A$. Следовательно, для $\omega \in A$ существует $\lim_{n \rightarrow \infty} \xi_n(t) = \eta(t)$, $|\xi_n(t) - \eta(t)| \leq \varepsilon_n$, так что сходимость $\xi_n(t) \rightarrow \eta(t)$ является равномерной. Вместе с непрерывностью $\xi_n(t)$ это означает непрерывность $\eta(t)$ (здесь мы доказываем, собственно, полноту пространства $C(0, 1)$).

Осталось убедиться в эквивалентности ξ и η . Для $t = t_{n,r}$ имеем $\xi_{n+k}(t) = \xi(t)$ при всех $k \geq 0$, так что $\eta(t) = \xi(t)$. Если $t \neq t_{n,r}$ при всех n и r , то существует последовательность r_n такая, что $t_{n,r_n} \rightarrow t$, $0 < t - t_{n,r_n} < 2^{-n}$ и, следовательно,

$$\begin{aligned} P(|\xi(t_{n,r_n}) - \xi(t)| > \varepsilon(t - t_{n,r_n})) &\leq q(t - t_{n,r_n}), \\ P(|\xi(t_{n,r_n}) - \xi(t)| > \varepsilon(2^{-n})) &\leq q(2^{-n}). \end{aligned}$$

В силу критерия Бореля—Кантелли это означает, что $\xi_{n,r_n} \rightarrow \xi(t)$ с вероятностью 1. Одновременно в силу непрерывности $\eta(t)$ выполняется $\eta(t_{n,r_n}) \rightarrow \eta(t)$. Так как $\xi(t_{n,r_n}) = \eta(t_{n,r_n})$ то $\xi(t) = \eta(t)$ с вероятностью 1. ■

Из теоремы 1 вытекает

Следствие 1. Если

$$E|\xi(t+h) - \xi(t)|^a \leq \frac{c|h|}{|\log|h||^{1+b}} \quad (3)$$

при некоторых $b > a > 0$, $c < \infty$, то выполнены условия теоремы 2 и, следовательно, для $\xi(t)$ существует непрерывная модификация.

Условие (3), конечно, будет выполнено, если выполнено (2), так что теорема Колмогорова является следствием теоремы 2.

Доказательство. Положим $\varepsilon(h) = |\log|h||^{-\beta}$, $1 < \beta < b/a$. Тогда

$$\sum_{n=1}^{\infty} \varepsilon(2^{-n}) = \sum_{n=1}^{\infty} n^{-\beta} < \infty,$$

и по неравенству типа Чебышева

$$P(|\xi(t+h) - \xi(t)| > \varepsilon(h)) \leq \frac{c|h|}{|\log|h||^{1+b}} (\varepsilon(h))^{-a} = \frac{c|h|}{|\log|h||^{1+\delta}} \equiv g(h),$$

где $\delta = b - a\beta > 0$. Остается заметить, что

$$\sum_{n=1}^{\infty} 2^n q(2^{-n}) = \sum_{n=1}^{\infty} |\log 2^{-n}|^{-1-\delta} = \sum_{n=1}^{\infty} n^{-1-\delta} < \infty. \quad \blacksquare$$

Критерий существования модификации $\xi(t)$, принадлежащей пространству $D(t)$, формулируется и доказывается сложнее и связан с более слабыми условиями, накладываемыми на процесс. Ограничимся лишь формулировкой следующего утверждения:

Теорема 3 (Колмогоров—Ченцов). Если при некоторых $\alpha \geq 0$, $\beta \geq 0$, $b > 0$ и при всех t , $h_1 \leq t \leq 1 - h_2$, $h_1 \geq 0$, $h_2 \geq 0$, выполняется

$$E|\xi(t) - \xi(t - h_1)|^\alpha |\xi(t + h_2) - \xi(t)|^\beta < ch^{1+b}, \quad h = h_1 + h_2, \quad (4)$$

то существует модификация $\xi(t)$ из $D(0, 1)^5$.

Условие (4) допускает следующее обобщение:

$$P(|\xi(t + h_2) - \xi(t)| |\xi(t) - \xi(t - h_1)| \geq \varepsilon(h)) \leq q(h), \quad (5)$$

при котором утверждение теоремы также сохранится (здесь $\varepsilon(h)$, $q(h)$ имеют тот же смысл, что и в теореме 1).

Следующие два примера иллюстрируют в какой-то мере характер условий теорем 1–3.

⁵⁾ Подробнее об этом см., например, [14].

Пример 2. Допустим, что случайный процесс $\xi(t)$ имеет вид $\xi(t) = \sum_{k=1}^r \xi_k \varphi_k(t)$,

где $\varphi_k(t)$ удовлетворяют условию Гёльдера $|\varphi_k(t+h) - \varphi_k(t)| \leq c|h|^\alpha$, $\alpha > 0$, (ξ_1, \dots, ξ_r) — произвольный случайный вектор, у которого конечны $E|\xi_k|^l$ при некотором $l > 1/\alpha$. Тогда процесс $\xi(t)$ (он является, очевидно, непрерывным) удовлетворяет условию (2). Действительно,

$$E|\xi(t+h) - \xi(t)|^l \leq c_1 \sum_{k=1}^r E|\xi_k|^l |h|^{\alpha l} \leq c_2 |h|^{\alpha l}.$$

Пример 3. Пусть $\gamma \in U_{0,1}$, $\xi(t) = 0$ при $t < \gamma$, $\xi(t) = 1$ при $t \geq \gamma$. Тогда $E|\xi(t+h) - \xi(t)|^l = P(\gamma \in (t, t+h)) = h$ при любом $l > 0$. Здесь условие (2) не выполнено, хотя $|\xi(t+h) - \xi(t)| \rightarrow 0$ при $h \rightarrow 0$. Условие (4), очевидно, выполнено, так как здесь

$$E|\xi(t) - \xi(t-h_1)| |\xi(t+h_2) - \xi(t)| = 0. \quad (6)$$

Примерно те же результаты получим, если в качестве $\xi(t)$ рассмотрим процесс восстановления для последовательности $\gamma_1, \gamma_2, \dots$, где распределение γ_j имеет плотность. В этом случае вместо (6) будет выполняться

$$E|\xi(t) - \xi(t-h_1)| |\xi(t+h) - \xi(t)| \leq ch_1 h_2 \leq ch^2.$$

В общем случае, когда мы не располагаем данными для построения модификаций процесса ξ из пространств $C(T)$ или $D(T)$, трудности, отмеченные в §1, позволяет преодолеть введение понятия сепарабельности.

Определение 4. Процесс $\xi(t)$ называется *сепарабельным*, если существует всюду плотное в T счетное множество S такое, что

$$P\left(\overline{\lim_{u \rightarrow t} \xi(u)} \geq \xi(t) \geq \underline{\lim_{u \rightarrow t} \xi(u)} \text{ для всех } t \in T\right) = 1. \quad (7)$$

Это свойство эквивалентно тому, что для любого интервала $I \subset T$

$$P\left(\sup_{u \in I \cap S} \xi(u) = \sup_{u \in I} \xi(u); \inf_{u \in I \cap S} \xi(u) = \inf_{u \in I} \xi(u)\right) = 1.$$

Дубу принадлежит теорема⁶⁾, утверждающая, что любой случайный процесс имеет сепарабельную модификацию.

Построение сепарабельной модификации процесса, так же, как и построение модификаций из $C(T)$ или $D(T)$, означает расширение σ -алгебры $\mathfrak{B}_{\mathbb{R}}^T$, куда присоединяются несчетные пересечения вида

$$A = \bigcap_{u \in I} \{\xi(u) \in [a, b]\} = \left\{ \sup_{u \in I} \xi(u) \leq b, \inf_{u \in I} \xi(u) \geq a \right\},$$

и распространение на это расширение меры P с помощью равенств

$$P(A) = P\left(\bigcap_{u \in I \cap S} \{\xi(u) \in [a, b]\}\right),$$

где под знаком вероятности стоит уже элемент $\mathfrak{B}_{\mathbb{R}}^T$.

⁶⁾ См. [7, 16].

Для сепарабельных процессов такие множества, как множество всех неубывающих функций, множества $C(T)$, $D(T)$ и др., являются событиями. Процессы из $C(T)$ или $D(T)$, очевидно, автоматически будут сепарабельными. И наоборот, если процесс сепарабелен и допускает непрерывную модификацию (или модификацию из $D(T)$), то он сам будет непрерывным (принадлежать $D(T)$). Действительно, если η есть непрерывная модификация ξ , то

$$P(\xi(t) = \eta(t) \text{ для всех } t \in S) = 1.$$

Отсюда на основании (7) получаем

$$P\left(\overline{\lim}_{\substack{u \rightarrow t \\ u \in S}} \eta(u) \geq \xi(t) \geq \underline{\lim}_{\substack{u \rightarrow t \\ u \in S}} \eta(u) \text{ для всех } t \in T\right) = 1.$$

Так как $\overline{\lim}_{u \rightarrow t} \eta(u) = \underline{\lim}_{u \rightarrow t} \eta(u) = \eta(t)$, то получаем, что

$$P(\xi(t) = \eta(t) \text{ при всех } t \in T) = 1.$$

В примере 1 процесс $\xi_1(t)$, очевидно, не является сепарабельным. Процесс $\xi_0(t)$ есть его сепарабельная модификация.

Наряду с непрерывностью выборочных траекторий существует еще один способ характеризовать свойство непрерывности случайного процесса.

Определение 5. Случайный процесс $\xi(t)$ называется *стохастически непрерывным*, если при всех $t \in T$ и $h \rightarrow 0$

$$\xi(t+h) \xrightarrow{p} \xi(t) \quad (P(|\xi(t+h) - \xi(t)| > \varepsilon) \rightarrow 0).$$

Здесь мы имеем дело лишь с двумерными распределениями $\xi(t)$.

Ясно, что все процессы с непрерывными траекториями являются стохастически непрерывными. Но не только. Разрывные процессы в примерах 1 и 3 также являются стохастически непрерывными. Разрывный процесс будет стохастически непрерывным в точке t_0 , если точка разрыва $\tau(\xi(\tau+0) \neq \xi(\tau-0))$ обладает свойством $P(\tau = t_0) = 0$.

Определение 6. Процесс $\xi(t)$ называется *непрерывным в среднем порядка r* (в среднеквадратичном при $r = 2$), если при всех $t \in T$ и $h \rightarrow 0$

$$\xi(t+h) \xrightarrow{(r)} \xi(t)$$

или, что то же, $E|\xi(t+h) - \xi(t)|^r \rightarrow 0$.

Разрывный процесс $\xi(t)$ в примере 3 будет непрерывным в среднем любого порядка. Поэтому сами по себе непрерывность в среднем и стохастическая непрерывность не очень много говорят о свойствах траекторий (они говорят лишь, что скачок в окрестности каждой фиксированной точки t маловероятен). Как показывает теорема Колмогорова, для того чтобы характеризовать свойства траекторий, нужны *количественные* оценки для $E|\xi(t+h) - \xi(t)|^r$ или для $P(|\xi(t+h) - \xi(t)| > \varepsilon)$.

Из теорем непрерывности для моментов следует, что для *стохастически непрерывного процесса* $\xi(t)$ и любой непрерывной и ограниченной функции $g(x)$ функция $Eg(\xi(t))$ непрерывна. Это утверждение сохранится, если ограниченность $g(x)$ заменить условием $\sup_t E|g(\xi(t))|^\alpha < \infty$ при каком-нибудь $\alpha > 1$.

Последующие гл. 18, 20, 21 будут посвящены изучению тех типов случайных процессов, которые можно задать явной формой конечномерных распределений. К ним относятся:

1. Процессы с независимыми приращениями.
2. Марковские процессы.
3. Гауссовские процессы.

В гл. 21, кроме того, будут рассмотрены некоторые вопросы теории случайных процессов с конечными моментами второго порядка. Гл. 19 содержит предельные теоремы для случайных процессов, порожденных последовательными суммами независимых случайных величин.

Глава 18

Процессы с независимыми приращениями

§ 1. Общие свойства

Определение 1. Процесс $\{\xi(t), t \in [a, b]\}$, определенный на отрезке $[a, b]$, называется *процессом с независимыми приращениями*, если для любых $t_0 < t_1 < \dots < t_n$, $a \leq t_0$, $t_n \leq b$, случайные величины $\xi(t_0)$, $\xi(t_1) - \xi(t_0)$, \dots , $\xi(t_n) - \xi(t_{n-1})$ независимы.

Процесс с независимыми приращениями называется *однородным*, если распределение $\xi(t_1) - \xi(t_0)$ определяется лишь длиной интервала $t_1 - t_0$ и не зависит от t_0 .

В дальнейшем для простоты везде будем считать, что $a = 0$, $\xi(0) = 0$, $b = 1$.

Определение 2. Распределение случайной величины ξ называется *безгранично делимым* (ср. с § 8.9), если при любом n величину ξ можно представить в виде суммы независимых одинаково распределенных случайных величин: $\xi = \xi_{1,n} + \dots + \xi_{n,n}$. Если $\varphi(\lambda)$ есть х.ф. ξ , то отмеченное свойство эквивалентно тому, что $\varphi^{1/n}$ при любом n является х.ф.

Из приведенных определений ясно, что распределение $\xi(t)$ для однородного процесса с независимыми приращениями ξ безгранично делимо, так как $\xi(t) = \xi_{1,n} + \dots + \xi_{n,n}$, где $\xi_{k,n} = \xi(kt/n) - \xi((k-1)t/n)$ независимы и распределены как $\xi(t/n)$.

Теорема 1.

1) Пусть $\{\xi(t), t \geq 0\}$ — стохастически непрерывный однородный процесс с независимыми приращениями и пусть $\varphi_t(\lambda) = E e^{i\lambda \xi(t)}$ есть х.ф. $\xi(t)$, $\varphi(\lambda) = \varphi_1(\lambda)$. Тогда

$$\varphi_t(\lambda) = \varphi^t(\lambda), \quad (1)$$

$\varphi(\lambda) \neq 0$ ни при каком λ . ((1) означает, что $\varphi^t(\lambda)$ обязано быть х.ф. при любом t .)

2) Пусть $\varphi(\lambda)$ есть х.ф. безгранично делимого распределения. Тогда существует случайный процесс $\{\xi(t), t \geq 0\}$, удовлетворяющий условиям 1), и такой, что $E e^{i\lambda \xi(1)} = \varphi(\lambda)$.

Обозначим через \mathcal{L} класс х.ф. всех безгранично делимых распределений и через \mathcal{L}_1 класс х.ф. распределений $\xi(t)$ для стохастически непрерывных однородных процессов с независимыми приращениями. Тогда из теоремы 1 следует, что $\mathcal{L} = \mathcal{L}_1$. Класс \mathcal{L} будет охарактеризован в § 5.

Доказательство.

1) Пусть $\xi(t)$ удовлетворяет условиям п. 1) теоремы. Тогда $\xi(t)$ можно представить в виде суммы независимых приращений

$$\xi(t) = \sum_{j=1}^n [\xi(t_j) - \xi(t_{j-1})], \quad t_0 = 0, \quad t_n = t, \quad t_j > t_{j-1}.$$

Отсюда следует, в частности, что

$$\varphi_t(\lambda) = [\varphi_{1/n}(\lambda)]^{[tn]} \varphi_{t-[tn]/n}(\lambda) = \varphi(\lambda)^{[tn]/n} \varphi_{t-[tn]/n}(\lambda).$$

Здесь $t - [tn]/n \rightarrow 0$, $\varphi_{t-[tn]/n}(\lambda) \rightarrow 1$ (в силу стохастической непрерывности), $\varphi^{[tn]/n}(\lambda) \rightarrow \varphi^t(\lambda)$ при $n \rightarrow \infty$. Отсюда следует, что $\varphi_t(\lambda) \rightarrow \varphi^t(\lambda)$ при $n \rightarrow \infty$. Но $\varphi_t(\lambda)$ и $\varphi^t(\lambda)$ от n не зависят, значит, $\varphi_t(\lambda) = \varphi^t(\lambda)$.

Далее, $\varphi_t(\lambda) = \varphi^t(\lambda) \rightarrow 1$ при $t \rightarrow 0$ и при любом λ . Это возможно, если $\varphi(\lambda) \neq 0$ ни при каком λ .

2) Заметим предварительно, что если $\varphi \in \mathcal{L}$, то φ^t при любом $t > 0$ есть снова х. ф. Действительно,

$$\varphi^t(\lambda) = \lim_{n \rightarrow \infty} \varphi^{[tn]/n}(\lambda),$$

так что $\varphi^t(\lambda)$ есть непрерывный в точке $\lambda = 0$ предел х. ф. По теореме непрерывности для х. ф. это снова есть х. ф.

Построим теперь случайный процесс $\xi(t)$ с независимыми приращениями, задав его конечномерные распределения. Положим $0 = t_0 < t_1 < \dots < t_k$, $\Delta_j = \xi(t_j) - \xi(t_{j-1})$, $\delta_j = t_j - t_{j-1}$ и заметим, что

$$\sum_{j=1}^k \lambda_j \xi(t_j) = \sum_{j=1}^k \lambda_j \sum_{l=1}^j \Delta_l = \sum_{j=1}^k \Delta_j \sum_{l=j}^k \lambda_l.$$

Определим х. ф. совместного распределения $\xi(t_1), \dots, \xi(t_k)$ равенством (постулирующим независимость Δ_j)

$$\mathbb{E} \exp \left\{ i \sum_{j=1}^k \lambda_j \xi(t_j) \right\} = \mathbb{E} \exp \left\{ i \sum_{j=1}^k \Delta_j \sum_{l=j}^k \lambda_l \right\} = \prod_{j=1}^k \varphi \left(\sum_{l=j}^k \lambda_l \right)^{\delta_j}.$$

Таким образом, мы с помощью φ определили конечномерные распределения $\xi(t)$ в $(\mathbb{R}^T, \mathcal{B}_{\mathbb{R}}^T)$, которые, как нетрудно видеть, согласованы. По теореме Колмогорова существует распределение случайного процесса $\xi(t)$ в $(\mathbb{R}^T, \mathcal{B}_{\mathbb{R}}^T)$. Этот процесс по определению является однородным процессом с независимыми приращениями.

Чтобы доказать стохастическую непрерывность $\xi(t)$, заметим, что при $h \rightarrow 0$,

$$\mathbb{E} e^{i\lambda(\xi(t+h) - \xi(t))} = \varphi^h(\lambda) \rightarrow \varphi_0(\lambda),$$

где

$$\varphi_0(\lambda) = \begin{cases} 1, & \text{если } \varphi(\lambda) \neq 0, \\ 0, & \text{если } \varphi(\lambda) = 0. \end{cases}$$

Таким образом, предельная функция $\varphi_0(\lambda)$ может принимать лишь два значения: 1 и 0. Но она обязана быть х. ф., так как непрерывна в точке $\lambda = 0$ ($\varphi(\lambda) \neq 0$ в окрестности точки $\lambda = 0$) и является пределом х. ф. Следовательно, $\varphi_0(\lambda)$ непрерывна, $\varphi_0(\lambda) \equiv 1$, $\varphi^h(\lambda) \rightarrow 1$,

$$\xi(t+h) - \xi(t) \xrightarrow{p} 0 \quad \text{при } h \rightarrow 0.$$

Теорема 2. Однородные стохастически непрерывные процессы с независимыми приращениями $\{\xi(t); t \in T\}$ имеют модификации из пространства $D(T)$, т. е. процесс $\xi(t)$ может быть задан в $\langle D(T), \mathfrak{B}_D^T \rangle$ и, следовательно, не иметь разрывов второго рода.

Доказательство. Чтобы упростить рассуждения, предположим, что существует $E\xi^2(1)$ или, что то же, существует вторая производная $\varphi''(\lambda)$. Тогда

$$E(\xi(t) - \xi(t-h))^2 = -\varphi_h''(0) = -h(h-1)[\varphi'(0)]^2 - h\varphi''(0) \leq c|h|,$$

$$E(|\xi(t+h_2) - \xi(t)|^2 |\xi(t) - \xi(t-h_1)|^2) \leq c^2 h_1 h_2 \leq c^2 (h_1 + h_2)^2,$$

так что выполнен критерий теоремы 17.3. ■

В общем случае доказательство усложняется: надо пользоваться критерием (17.5) и оценками для

$$P(|\xi(t) - \xi(t-h)| > \varepsilon).$$

Рассмотрим теперь два наиболее важных представителя процессов с независимыми приращениями — так называемые винеровские и пуассоновские процессы.

§ 2. Винеровские процессы, свойства траекторий

Определение 3. Винеровским процессом называется однородный процесс с независимыми приращениями, для которого распределение $\xi(1)$ нормально.

Другими словами, это процесс, для которого

$$\varphi(\lambda) = e^{i\lambda\alpha - \sigma^2\lambda^2/2}, \quad \varphi_t(\lambda) = \varphi^t(\lambda) = e^{i\lambda t\alpha - \sigma^2\lambda^2 t/2}$$

при некоторых α и $\sigma^2 \geq 0$. Последнее равенство означает, что приращения $\xi(t+u) - \xi(u)$ распределены нормально с параметрами $(\alpha t, \sigma^2 t)$. Все совместные распределения $\xi(t_1), \dots, \xi(t_n)$, очевидно, также будут нормальными.

Числа α и σ называются соответственно коэффициентами сноса и диффузии. Введя в рассмотрение процесс $\xi_0(t) = (\xi(t) - \alpha t)/\sigma$, отличающийся от $\xi(t)$ линейным преобразованием, получим, что для него х. ф. равна

$$E e^{i\lambda \xi_0(t)} = e^{-i\lambda \alpha t / \sigma} \varphi_t\left(\frac{\lambda}{\sigma}\right) = e^{-\lambda^2 t / 2}.$$

Такой процесс с параметрами $(0, t)$ часто называют *стандартным винеровским процессом*. Рассмотрим его подробнее.

Теорема 3. Существует непрерывная модификация винеровского процесса.

Это означает, как мы знаем, что винеровский процесс $\{\xi(t), t \in [0, 1]\}$ можно рассматривать, как заданный на измеримом пространстве $\langle C(0, 1), \mathfrak{B}_C^{[0, 1]} \rangle$ непрерывных функций.

Доказательство. Имеем $\xi(t+h) - \xi(t) \in \Phi_{0,h}$, $(\xi(t+h) - \xi(t))/\sqrt{h} \in \Phi_{0,1}$. Поэтому

$$E(\xi(t+h) - \xi(t))^4 = h^2 E\xi(1)^4 = 3h^2.$$

Это означает, что выполнены условия теоремы 17.1. ■

Итак, можно считать, что $\xi(\cdot) \in C(0, 1)$. Стандартный винеровский процесс с непрерывными траекториями будем обозначать $\{w(t), t \in T\}$.

Отметим теперь, что *траектории винеровского процесса* $w(t)$, обладая свойствами непрерывности, с вероятностью 1 не дифференцируемы ни в какой точке t .

В силу однородности процесса достаточно доказать его недифференцируемость в точке 0. Если бы с положительной вероятностью, т.е. на множестве $A \subset \Omega$, $P(A) > 0$, существовала производная

$$w'(0) = w'(0, \omega) = \lim_{t \rightarrow 0} \frac{w(t)}{t},$$

то тогда на том же множестве существовал бы

$$\lim_{k \rightarrow \infty} \frac{w(2^{-k+1}) - w(2^{-k})}{2^{-k}} = \lim_{k \rightarrow \infty} \frac{2w(2^{-k+1})}{2^{-k+1}} - \lim_{k \rightarrow \infty} \frac{w(2^{-k})}{2^{-k}} = 2w'(0) - w'(0) = w'(0).$$

Но это невозможно по следующей причине. Независимые разности $w(2^{-k+1}) - w(2^{-k})$ распределены так же, как $w(2^{-k})$, и с положительной вероятностью $p = 1 - \Phi(1)$ превосходят $\sqrt{2^{-k}}$. Т.е. независимые события $B_k = \{w(2^{-k+1}) - w(2^{-k}) > \sqrt{2^{-k}}\}$ обладают тем свойством, что $\sum_{k=1}^{\infty} P(B_k) = \infty$. По критерию Бореля—Кантелли это означает, что с вероятностью 1 произойдет бесконечно много событий B_k , так что

$$P\left(\limsup_{k \rightarrow \infty} \frac{w(2^{-k+1}) - w(2^{-k})}{\sqrt{2^{-k}}} > 1\right) = 1.$$

Точно так же обнаруживаем, что

$$P\left(\liminf_{k \rightarrow \infty} \frac{w(2^{-k+1}) - w(2^{-k})}{\sqrt{2^{-k}}} < -1\right) = 1.$$

Отсюда следует, что

$$P\left(\limsup_{k \rightarrow \infty} \frac{w(2^{-k+1}) - w(2^{-k})}{2^{-k}} = \infty; \liminf_{k \rightarrow \infty} \frac{w(2^{-k+1}) - w(2^{-k})}{2^{-k}} = -\infty\right) = 1,$$

и, следовательно, процесс $w(t)$ в каждой точке t с вероятностью 1 недифференцируем.

Имеет место и более сильный факт: с вероятностью 1 не существует ни одной точки t , где траектория $w(t)$ имела бы производную. Другими словами, винеровский процесс с вероятностью 1 нигде не дифференцируем. Доказательство этого факта значительно сложнее и выходит за рамки этой книги.

Читатель может без труда проверить, что $w(t)$ в известном смысле обладает свойством параболы — именно, процесс $w^*(t) = 1/\sqrt{c} w(ct)$ при любом $c > 0$ снова является стандартным винеровским процессом.

§ 3. Законы повторного логарифма

С помощью рассуждений, подобных тем, что применялись в конце предыдущего параграфа, можно установить и значительно более сильный факт: траектория $w(t)$ в окрестности точки $t = 0$, образно говоря, «полностью зачерчивает» внутренность области, образованной двумя ветвями кривых $y = \pm \sqrt{2t \ln \ln \frac{1}{t}}$. Внешность этой области остается нетронутой. Это так называемый *локальный закон повторного логарифма*.

Теорема 4.

$$P\left(\limsup_{t \rightarrow 0} \frac{w(t)}{\sqrt{2t \ln \ln \frac{1}{t}}} = 1\right) = 1,$$

$$P\left(\liminf_{t \rightarrow 0} \frac{w(t)}{\sqrt{2t \ln \ln \frac{1}{t}}} = -1\right) = 1.$$

Таким образом, если рассмотреть последовательность случайных величин $w(t_n)$, $t_n \downarrow 0$, то последовательности $\sqrt{2t_n \ln \ln \frac{1}{t_n}}(1 \pm \varepsilon)$ при любом $\varepsilon > 0$ будут для нее соответственно верхней и нижней последовательностями.

Для процессов можно было бы естественным образом ввести понятия *верхней* и *нижней функций*. Если, например, процесс $\xi(t)$ принадлежит $C(0, \infty)$ или $D(0, \infty)$ (или сепарабелен на $(0, \infty)$), то соответствующее определение для $t \rightarrow \infty$ имеет следующий вид.

Определение 4. Функция $a(t)$ называется *верхней* (*нижней*) для процесса $\xi(t)$, если для какой-нибудь последовательности $t_n \uparrow \infty$ события $A_n = \{\sup_{u \geq t_n} (\xi(t) - a(t)) > 0\}$ с вероятностью 1 наступают лишь конечное число раз (бесконечное число раз).

Наряду с теоремой 4 мы получим здесь обычный закон повторного логарифма. Доказываются оба эти утверждения практически одинаково. Докажем второе, а первое получим в качестве его следствия.

Теорема 5 (закон повторного логарифма).

$$P\left(\limsup_{t \rightarrow \infty} \frac{w(t)}{\sqrt{2t \ln \ln t}} = 1\right) = 1,$$

$$P\left(\liminf_{t \rightarrow \infty} \frac{w(t)}{\sqrt{2t \ln \ln t}} = -1\right) = 1.$$

Таким образом, при $t \rightarrow \infty$ функции $\sqrt{2t \ln \ln t}(1 \pm \varepsilon)$ при любом $\varepsilon > 0$ являются соответственно верхней и нижней функциями для $w(t)$.

Обозначим $\bar{w}(t) = \sup_{u \leq t} w(u)$. Для доказательства теоремы потребуется

Лемма 1.

$$P(\bar{w}(t) \geq x) = 2P(w(t) \geq x) = \frac{2}{\sqrt{2\pi t}} \int_x^\infty e^{-v^2/2t} dv; \quad x \geq 0.$$

Здесь справа стоит так называемый *удвоенный нормальный закон*.

Доказательство. Введем в рассмотрение случайную величину $\tau_x = \inf\{u : w(u) \geq x\}$ (напомним, что $w(\cdot) \in C(0, \infty)$). Это есть момент остановки в том же смысле, что и раньше:

$$\{\tau_x > v\} = \{\bar{w}(v) < x\} \in \mathfrak{F}_v = \sigma\{\xi(u); u \leq v\},$$

$$\{\tau_x = v\} = \bigcap_{n=1}^{\infty} \left\{ \bar{w}\left(v - \frac{1}{n}\right) < x, \bar{w}(v) = x \right\} \in \mathfrak{F}_v.$$

Следовательно, при фиксированном $\tau_x = v$, $v \leq t$, приращение $w(t) - w(\tau_x) = w(t) - w(v)$ не зависит от $\{\tau_x = v\} \in \mathcal{F}_v$ и распределено как $w(t - v) \in \Phi_{0, t-v}$. Поэтому на множестве $\{\tau_x \leq t\}$

$$P(w(t) \geq x | \tau_x) = P(w(t) - w(\tau_x) \geq 0 | \tau_x) = \frac{1}{2},$$

$$P(w(t) \geq x) = E[P(w(t) \geq x | \tau_x); \tau_x \leq t] = \frac{1}{2}P(\tau_x \leq t) = \frac{1}{2}P(\bar{w}(t) \geq x).$$

Доказательство теоремы 5. Заметим предварительно, что в силу правила Лопиталя

$$P(w(t) > x) = \frac{1}{\sqrt{2\pi t}} \int_x^\infty e^{-u^2/2t} du = \frac{1}{\sqrt{2\pi}} \int_{x/\sqrt{t}}^\infty e^{-u^2/2} du \sim \frac{\sqrt{t}}{\sqrt{2\pi}x} e^{-x^2/2t} \quad (2)$$

при $x/\sqrt{t} \rightarrow \infty$.

Пусть $a > 1$ и $x_k = \sqrt{2a^k \ln \ln a^k}$. Надо убедиться, что при любом $\varepsilon > 0$

$$P\left(\limsup_{t \rightarrow \infty} \frac{w(t)}{\sqrt{2t \ln \ln t}} < 1 + \varepsilon\right) = 1, \quad (3)$$

т. е. что с вероятностью 1 *при всех* достаточно больших t

$$w(t) < (1 + \varepsilon)\sqrt{2t \ln \ln t}.$$

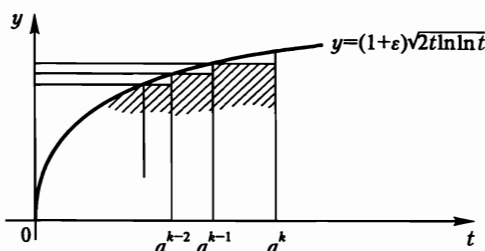


Рис. 16.

Для этого достаточно установить, что с вероятностью 1 происходит лишь конечное число событий

$$B_k = \left\{ \sup_{a^{k-1} < u \leq a^k} w(u) > (1 + \varepsilon)x_{k-1} \right\}.$$

Рассмотрим события $A_k = \left\{ \sup_{u \leq a^k} w(u) > (1 + \varepsilon)x_{k-1} \right\}$ (см. рис. 16). Тогда, так как $x_k/\sqrt{a^k} \rightarrow \infty$ при $k \rightarrow \infty$, то

$$\begin{aligned} P(A_k) &= 2P(w(a^k) > (1 + \varepsilon)x_{k-1}) \sim \\ &\sim \sqrt{\frac{2}{\pi}} \frac{\sqrt{a^k}}{(1 + \varepsilon)x_{k-1}} \exp \left\{ -\frac{2(1 + \varepsilon)^2 a^{k-1} \ln \ln a^{k-1}}{2a^k} \right\} = \\ &= \frac{1}{1 + \varepsilon} \sqrt{\frac{1}{\pi a \ln \ln a^{k-1}}} \frac{1}{(\ln a^{k-1})^{(1+\varepsilon)^2/a}} = \\ &= c(a, \varepsilon) \frac{1}{(k-1)^{(1+\varepsilon)^2/a} \sqrt{\ln(k-1) + \ln \ln a}}. \end{aligned}$$

Положим $a = 1 + \varepsilon > 1$. Тогда, очевидно, при $k \rightarrow \infty$

$$P(A_k) \sim \frac{c(\varepsilon)}{k^{1+\varepsilon} \sqrt{\ln k}}.$$

В этих формулах $c(a, \varepsilon)$ и $c(\varepsilon)$ — некоторые постоянные, зависящие от своих аргументов. Полученное соотношение означает, что $\sum_{k=1}^{\infty} P(A_k) < \infty$, $\sum_{k=1}^{\infty} P(B_k) < \infty$

$(B_k \subset A_k)$ и, значит, по критерию Бореля—Кателли (теорема 10.1) события B_k с вероятностью 1 происходят лишь конечное число раз.

Докажем теперь, что при произвольном $\varepsilon > 0$

$$P\left(\limsup_{t \rightarrow \infty} \frac{w(t)}{\sqrt{2t \ln \ln t}} > 1 - \varepsilon\right) = 1. \quad (4)$$

Очевидно, вместе с (3) это и будет означать, что выполнено первое утверждение теоремы.

Рассмотрим при $a > 1$ независимые приращения $w(a^k) - w(a^{k-1})$ и обозначим B_k событие

$$B_k = \left\{ w(a^k) - w(a^{k-1}) > \left(1 - \frac{\varepsilon}{2}\right) x_k \right\}.$$

Так как $w(a^k) - w(a^{k-1})$ распределено так же, как $w(a^k(1 - a^{-1}))$, то в силу (2), как и прежде, находим

$$P(B_k) \sim \frac{\sqrt{a^k(1 - a^{-1})}}{\sqrt{2\pi(1 - \varepsilon/2)x_k}} \exp \left\{ -\frac{(1 - \varepsilon/2)^2 2a^k \ln \ln a^{-k}}{2a^k(1 - a^{-1})} \right\} \sim \frac{c_1(a, \varepsilon)}{\sqrt{\ln k}} k^{-(1 - \varepsilon/2)^2/(1 - a^{-1})}.$$

Отсюда следует, что при $a \geq 2/\varepsilon$ ряд $\sum_{k=1}^{\infty} P(B_k)$ расходится, и, значит, по критерию Бореля—Кателли с вероятностью 1 наступает бесконечное число событий B_k .

Далее, в силу симметрии процесса $w(t)$ из соотношения (3) следует, что при всех достаточно больших k и любом $\delta > 0$

$$w(a^k) > -(1 + \delta)x_k.$$

Вместе с предыдущим это означает, что бесконечное число раз будет наступать событие

$$w(a^{k-1}) + [w(a^k) - w(a^{k-1})] = w(a^k) > -(1 + \delta)x_{k-1} + \left(1 - \frac{\varepsilon}{2}\right)x_k.$$

Но правая часть этого неравенства может быть сделана выбором a больше $(1 - \varepsilon)x_k$. Действительно,

$$-(1 + \delta)x_{k-1} + \frac{\varepsilon}{2}x_k > 0,$$

если только $(1 + \delta)\sqrt{\frac{\ln \ln a^{k-1}}{a \ln \ln a^k}} < \varepsilon/2$, чего в свою очередь легко добиться, выбрав a достаточно большим. Таким образом, соотношение (4) доказано.

Второе утверждение теоремы, очевидно, следует из первого в силу симметрии распределения $w(t)$. ■

В качестве следствия можно получить теперь локальный закон повторного логарифма для $t \rightarrow 0$.

Доказательство теоремы 4. Рассмотрим процесс $\{W(u) = uw(1/u); u \geq 0\}$, где принято $W(0) = 0$. Замечательный факт состоит в том, что процесс $\{W(u); u \geq 0\}$ также является стандартным винеровским процессом. Действительно, при $t > u$

$$\begin{aligned} E \exp \{i\lambda(W(t) - W(u))\} &= E \exp \left\{ i\lambda \left[tw\left(\frac{1}{t}\right) - uw\left(\frac{1}{u}\right) \right] \right\} = \\ &= E \exp \left\{ i\lambda \left[w\left(\frac{1}{t}\right)(t - u) - u \left(w\left(\frac{1}{u}\right) - w\left(\frac{1}{t}\right) \right) \right] \right\} = \\ &= \exp \left\{ -\frac{\lambda^2}{2}(t - u)^2 \frac{1}{t} - \frac{\lambda^2 u^2}{2} \left(\frac{1}{u} - \frac{1}{t} \right) \right\} = \exp \left\{ -\frac{\lambda^2}{2}(t - u) \right\}. \end{aligned}$$

Независимость приращений проще всего доказать, установив их некоррелированность. Действительно,

$$\begin{aligned} E[W(u)(W(t) - W(u))] &= E\left[uw\left(\frac{1}{u}\right)\left(tw\left(\frac{1}{t}\right) - uw\left(\frac{1}{u}\right)\right)\right] = \\ &= E\left[uw\left(\frac{1}{t}\right)tw\left(\frac{1}{t}\right) - u^2w^2\left(\frac{1}{u}\right)\right] = u - u = 0. \end{aligned}$$

Для доказательства теоремы остается заметить, что

$$\limsup_{t \rightarrow \infty} \frac{w(t)}{\sqrt{2t \ln \ln t}} = \limsup_{u \rightarrow 0} \frac{uw\left(\frac{1}{u}\right)}{\sqrt{2u \ln \ln \frac{1}{u}}} = \limsup_{u \rightarrow 0} \frac{W(u)}{\sqrt{2u \ln \ln \frac{1}{u}}}.$$

Доказательство можно было бы провести, повторив рассуждения теоремы 5, где надо выбрать $a < 1$.

В заключение отметим, что винеровские процессы играют весьма важную роль во многих теоретико-вероятностных рассмотрениях и служат моделями для описания многих процессов в природе. Они представляют собой, например, хорошую модель для описания движения диффундирующей частицы. В связи с этим винеровские процессы часто называют также *процессами броуновского движения*.

Винеровские процессы оказываются в известном смысле предельными для случайных ломаных, построенных по точкам $(k/n, S_k/\sqrt{n})$, где S_k — суммы случайных величин ξ_j , $E\xi_j = 0$, $D\xi_j = 1$. Об этом подробнее будет сказано в гл. 19. С винеровским процессом тесно связано также понятие стохастического интеграла и большое количество других конструкций и результатов.

§ 4. Пуассоновские процессы

Определение 5. Однородный процесс $\xi(t)$ с независимыми приращениями называется *пуассоновским*, если $\xi(t) - \xi(0)$ имеет распределение Пуассона.

Положим для простоты $\xi(0) = 0$. Если $\xi(1) \in \Pi_\mu$, то $\varphi(\lambda) = Ee^{i\lambda\xi(1)} = \exp\{\mu(e^{i\lambda} - 1)\}$ и, как мы знаем,

$$\varphi_t(\lambda) = Ee^{i\lambda\xi(t)} = \varphi^t(\lambda) = \exp\{\mu t(e^{i\lambda} - 1)\},$$

так что $\xi(t) \in \Pi_{\mu t}$.

Рассмотрим свойства пуассоновского процесса. Прежде всего, $\xi(t)$ при каждом t принимает лишь целочисленные значения $0, 1, 2, \dots$. Разобьем полуинтервал $[0, t)$ на полуинтервалы $[0, t_1), [t_1, t_2), \dots, [t_{n-1}, t_n)$ длиной $\Delta_i = t_i - t_{i-1}$, $i = 1, \dots, n$. При малых Δ_i распределение приращений $\xi(t_i) - \xi(t_{i-1})$ будет обладать свойствами

$$\begin{aligned} P(\xi(t_i) - \xi(t_{i-1}) = 0) &= P(\xi(\Delta_i) = 0) = e^{-\mu\Delta_i} = 1 - \mu\Delta_i + O(\Delta_i^2), \\ P(\xi(t_i) - \xi(t_{i-1}) = 1) &= \mu\Delta_i e^{-\mu\Delta_i} = \mu\Delta_i + O(\Delta_i^2), \\ P(\xi(t_i) - \xi(t_{i-1}) \geq 2) &= O(\Delta_i^2). \end{aligned} \quad (5)$$

Рассмотрим «вложенные» рациональные разбиения $\mathcal{R}(n) = \{t_1, \dots, t_n\}$ отрезка $[0, t]$ такие, что $\mathcal{R}(n) \subset \mathcal{R}(n+1)$, $\bigcup \mathcal{R}(n) = \mathcal{R}_t$ есть множество всех рациональных чисел отрезка $[0, t]$.

Отметим следующие три обстоятельства.

1) Пусть $\nu(n)$ — число интервалов разбиения $\mathcal{R}(n)$, на которых приращение процесса ξ отлично от 0. При каждом ω $\nu(n)$ не убывает при $n \rightarrow \infty$. Кроме того, число $\nu(n)$ представимо в виде суммы независимых случайных величин, равных 1, если на i -м интервале есть приращение, и нулю в противном случае. Поэтому в силу (5)

$$\begin{aligned} P(\nu(n) \neq \xi(t)) &= P\left(\bigcup_{t_j \in \mathcal{R}(n)} \{\xi(t_i) - \xi(t_{i-1}) \geq 2\} \cup \{\xi(t) - \xi(t_n) \geq 1\}\right) = \\ &= O\left(\sum_{j=1}^n \Delta_j^2 + (t - t_n)\right), \end{aligned}$$

где $\sum_{j=1}^n \Delta_j^2 \leq t \max \Delta_j \rightarrow 0$ при $n \rightarrow \infty$, так что при измельчении разбиения п.в. $\nu(n) \uparrow \xi(t) \in \Pi_{\mu t}$.

2) Так как длина максимального интервала Δ_j сходится к 0 при $n \rightarrow \infty$, то суммарная длина интервалов, содержащих скачки, сходится к 0.

Следовательно, объединяя оставшиеся (т.е. без приращений ξ) и примыкающие друг к другу интервалы Δ_j , при каждом ω получим при $n \rightarrow \infty$ в пределе $\xi(t) + 1$ интервалов $(0, a_1), (a_1, a_2), \dots, (a_{\xi(t)}, t)$, на которых приращения процесса ξ равны нулю.

3) Наконец, в силу (5) вероятность того, что хотя бы один скачок на интервалах Δ_j превосходит единицу, меньше $\sum O(\Delta_j^2) = o(1)$ при $n \rightarrow \infty$, так что с вероятностью 1 приращения в точках a_k равны 1.

Мы получили, таким образом, что на отрезке $[0, t]$ для каждого ω существует конечное число $\xi(t)$ точек $a_1, \dots, a_{\xi(t)}$ таких, что $\xi(u)$ в рациональных точках интервалов (a_k, a_{k+1}) принимает одно и то же постоянное значение, равное k . Это означает, что можно доопределить траектории процесса $\xi(u)$, скажем, по непрерывности справа так, что $\xi(u) = k$ при всех u , $a_k \leq u < a_{k+1}$.

Теми самым для исходного процесса $\xi(t)$ построена модификация $\bar{\xi}(t)$ с траекториями в $D_+(T)$. Эквивалентность $\bar{\xi}$ и ξ следует из самого построения, так как в силу 1)

$$P(\bar{\xi}(t) = \xi(t)) = P\left(\lim_{n \rightarrow \infty} \nu(n) = \xi(t)\right) = 1.$$

Такие непрерывные справа (или слева) модификации пуассоновского процесса обычно и рассматривают. С такого рода процессами мы уже имели дело в гл. 9, где определялись более общие процессы — процессы восстановления — сразу с помощью траекторий. Тот факт, что пуассоновский процесс есть процесс восстановления, вытекает из следующих соображений. Из соотношений (5) легко установить, что распределения случайных величин $a_1, a_2 - a_1, a_3 - a_2, \dots$ совпадают и что сами эти величины независимы. Действительно, $a_j - a_{j-1}$, $j \geq 1$, $a_0 = 0$, приближается суммой $(\gamma_j - \gamma_{j-1})\Delta$ одинаковых интервалов $\Delta_i = \Delta$, где γ_j есть номер интервала, на котором расположено j -е по счету ненулевое приращение ξ . Так как

$$P((\gamma_j - \gamma_{j-1})\Delta > u) = P\left(\gamma_1 > \frac{u}{\Delta}\right) = (e^{-\mu\Delta})^{[u/\Delta]} \rightarrow e^{-\mu u}$$

при $\Delta \rightarrow 0$, то величины $\alpha_j = a_j - a_{j-1}$, $j = 1, 2, 3, \dots$, распределены показательно, а значение $\xi(t) + 1$ можно рассматривать как время первого прохождения уровня t суммами a_j :

$$\xi(t) = \max\{k: a_k \leq t\}, \quad \xi(t) + 1 = \min\{k: a_k > t\}.$$

Мы получили, таким образом, что пуассоновский процесс $\xi(t)$ совпадает с процессами восстановления $\eta(t)$ (см. гл. 9) для показательного распределенных величин $\alpha_1, \alpha_2, \dots$; $P(\alpha_j > u) = e^{-\mu u}$.

Из сказанного и из свойств пуассоновского процесса вытекает также следующее замечательное свойство показательного распределенных случайных величин. Количества точек скачков (т. е. сумм a_k), попавших в непересекающиеся интервалы времени δ_j , независимы, причем эти количества имеют распределения Пуассона с параметрами $\mu\delta_j$.

Пользуясь последним обстоятельством, можно построить и более общую модель скачкообразного однородного процесса с независимыми приращениями. Рассмотрим произвольную, не зависящую от σ -алгебры, порожденной процессом $\xi(t)$, последовательность независимых одинаково распределенных случайных величин ζ_1, ζ_2, \dots с х.ф. $\beta(\lambda)$. Построим теперь новый процесс $\zeta(t)$ следующим образом. Каждому ω поставим в соответствие новую траекторию, полученную из траектории $\xi(t)$ заменой первого единичного скачка на величину ζ_1 , второго — на величину ζ_2 и т. д. Легко видеть, что $\zeta(t)$ также будет процессом с независимыми приращениями. При этом значение $\zeta(t)$ будет равно сумме

$$\zeta(t) = \zeta_1 + \dots + \zeta_{\xi(t)} \quad (6)$$

случайного числа $\xi(t)$ случайных величин ζ_1, ζ_2, \dots , где $\xi(t)$ по построению не зависит от $\{\zeta_k\}$.

Поэтому по формуле полной вероятности

$$\begin{aligned} Ee^{i\lambda\zeta(t)} &= \sum P(\xi(t) = k) Ee^{i\lambda(\zeta_1 + \dots + \zeta_k)} = \\ &= \sum \frac{(\mu t)^k}{k!} e^{-\mu t} (\beta(\lambda))^k = e^{-\mu t + \mu t \beta(\lambda)} = e^{\mu t(\beta(\lambda) - 1)}. \end{aligned} \quad (7)$$

Определение 6. Процесс $\zeta(t)$, определенный формулой (6) или х.ф. (7), называется *обобщенным пуассоновским процессом*.

Как уже отмечалось, это снова есть однородный процесс с независимыми приращениями. В формуле (7) μ определяет интенсивность скачков процесса $\zeta(t)$, $\beta(\lambda)$ — их распределение. Если к $\zeta(t)$ добавить постоянный «снос» at , то процесс $\tilde{\zeta}(t) = \zeta(t) + at$, очевидно, также будет процессом с независимыми приращениями с х.ф. $Ee^{i\lambda\tilde{\zeta}(t)} = e^{t(i\lambda a + \mu(\beta(\lambda) - 1))}$.

Наконец, если на том же вероятностном пространстве задан винеровский процесс $w(t)$ с коэффициентами $(0, \sigma)$, независимый от $\tilde{\zeta}(t)$, и мы каждому ω поставим в соответствие траекторию $\tilde{\zeta}(t) + w(t)$, то вновь получим процесс с независимыми приращениями с х.ф. $e^{t(i\lambda a - \lambda^2 \sigma^2 / 2 + \mu(\beta(\lambda) - 1))}$.

Следует отметить, однако, что этими построениями далеко не исчерпан весь класс процессов с независимыми приращениями (и, следовательно, класс безгранично делимых распределений).

Описание всего класса будет дано в следующем параграфе.

Пуассоновские процессы так же, как и винеровские, часто используются в качестве математических моделей в разного рода приложениях.

Пуассоновским процессом описывается, например, процесс появления космических частиц определенной энергии в заданном объеме или процесс возникновения соударений элементарных частиц в ускорителе. Им же описывается процесс поступления вызовов на телефонную станцию и многое другое.

Изучение обобщенных пуассоновских процессов в силу представления (6) во многом сводится к исследованию свойств сумм независимых случайных величин.

§ 5. Описание распределений всего класса процессов с независимыми приращениями

Мы видели в теореме 1 §1, что для описания класса распределений стохастически непрерывных однородных процессов с независимыми приращениями достаточно описать класс всех безгранично делимых распределений. Пусть, как и прежде, \mathcal{L} есть класс х.ф. безгранично делимых распределений.

Лемма 2. *Класс \mathcal{L} замкнут относительно операций умножения и предельного перехода (если предел есть снова х.ф.).*

Доказательство.

1) Пусть $\varphi_1 \in \mathcal{L}$, $\varphi_2 \in \mathcal{L}$. Тогда $\varphi_1 \varphi_2 = (\varphi_1^{1/n} \cdot \varphi_2^{1/n})^n$, где $\varphi_1^{1/n}$, $\varphi_2^{1/n}$ есть х.ф.

2) Пусть $\varphi_n \in \mathcal{L}$, $\varphi_n \rightarrow \varphi$, φ есть х.ф. Тогда для любого m $\varphi_n^{1/m} \rightarrow \varphi^{1/m}$, где $\varphi^{1/m}$ непрерывна в нуле и, следовательно, есть х.ф. ■

Обозначим через $\mathcal{L}_n \subset \mathcal{L}$ класс х.ф. вида

$$\ln \varphi(\lambda) = i\lambda a + \sum c_k (e^{i\lambda b_k} - 1), \quad c_k \geq 0, \quad \sum c_k < \infty.$$

Этот класс будем называть *пуассоновским*. Мы уже знаем, что он соответствует пуассоновским процессам со сносом a и интенсивностью c_k скачков величины b_k ($\sum c_k (e^{i\lambda b_k} - 1) = \sum c_k E(e^{i\lambda \zeta} - 1)$, где ζ имеет скачки b_k с вероятностью $c_k / \sum c_k$).

Лемма 3. *Х.ф. φ принадлежит \mathcal{L} тогда и только тогда, когда $\varphi = \lim \varphi_n$, $\varphi_n \in \mathcal{L}_n$.*

Доказательство.

Достаточность. Пусть

$$\ln \varphi_n = \psi_n = \sum_k (i\lambda a_{k,n} + c_{k,n} (e^{i\lambda b_{k,n}} - 1)),$$

$\varphi = \lim \varphi_n$ есть х.ф. Очевидно, что $\varphi_n^{1/m} \in \mathcal{L}_n \subset \mathcal{L}$, $\varphi_n^{1/m} \rightarrow \varphi^{1/m}$. Поэтому $\varphi^{1/m}$, будучи непрерывным в нуле пределом х.ф., есть х.ф., $\varphi \in \mathcal{L}$.

Необходимость. Пусть $\varphi \in \mathcal{L}$. Тогда $\varphi(\lambda) \neq 0$, существует $\psi = \ln \varphi$, $n(\varphi^{1/n} - 1) \rightarrow \psi$,

$$\varphi^{1/n} - 1 = \int (e^{i\lambda x} - 1) dF_n(x).$$

Интеграл от непрерывной функции, стоящий справа, можно рассматривать как интеграл Римана—Стилтьеса. Это означает, что для F_n существует разбиение вещественной оси на интервалы Δ_{nk} такое, что при $x_{nk} \in \Delta_{nk}$, $r_n < cn^{-2}$

$$\int (e^{i\lambda x} - 1) dF_n(x) = \sum_k (e^{i\lambda x_{nk}} - 1) P_n(\Delta_{nk}) + r_n$$

($P_n(\Delta)$ есть вероятность попадания в Δ , соответствующая F_n). Получаем

$$\psi = \lim n(\varphi^{1/n} - 1) = \lim \left[n \sum_k (e^{i\lambda x_{nk}} - 1) P_n(\Delta_{nk}) \right].$$

Теорема 6 (Леви, Хинчин). $\varphi \in \mathcal{L}$ тогда и только тогда, когда функция $\psi = \ln \varphi$ представима в виде

$$\psi = \psi(\lambda; a, \Psi) = i\lambda a + \int \left(e^{i\lambda x} - 1 - \frac{i\lambda x}{1+x^2} \right) \frac{1+x^2}{x^2} d\Psi(x), \quad (8)$$

где Ψ есть неубывающая функция ограниченной вариации (функция распределения с точностью до постоянного множителя), подынтегральную функцию при $x = 0$ следует считать равной $-\lambda^2/2$ (доопределенной по непрерывности).

Доказательство. Пусть ψ имеет вид (8). Тогда $\psi(\lambda)$ есть непрерывная функция, так как она представляет собой (с точностью до непрерывного слагаемого $i\lambda a$) равномерно сходящийся интеграл от непрерывной ограниченной функции. Далее, пусть $x_{kn} \neq 0$, $k = 1, \dots, n$, — точки измельчающегося разбиения полуинтервала $[-\sqrt{n}, \sqrt{n})$. Тогда $\psi^0(\lambda) = \psi(\lambda) - i\lambda a$ можно представить в виде $\psi^0 = \lim \psi_n$,

$$\psi_n(\lambda) = \sum_{k=1}^n [i\lambda a_{kn} + c_{kn}(e^{i\lambda b_{kn}} - 1)] \in \mathcal{L}_n,$$

где при естественных соглашениях относительно обозначений следует положить

$$c_{kn} = \frac{1+x_{kn}^2}{x_{kn}^2} \Psi([x_{kn}, x_{k+1,n})), \quad a_{kn} = -\frac{\Psi([x_{kn}, x_{k+1,n}))}{x_{kn}}, \quad b_{kn} = x_{kn},$$

здесь буквой Ψ мы обозначаем также меру $\Psi(A) = \int_A d\Psi(x)$. Мы получили, что φ есть непрерывный предел последовательности х.ф. $\varphi_n \in \mathcal{L}_n$. Остается воспользоваться леммой 3.

Пусть теперь $\varphi \in \mathcal{L}$. Тогда

$$\begin{aligned} \psi &= \lim n(\varphi^{1/n} - 1) = \lim \int (e^{i\lambda x} - 1) n dF_n(x) = \\ &= \lim \left[i\lambda \int \frac{nx}{1+x^2} dF_n(x) + \int \left(e^{i\lambda x} - 1 - \frac{i\lambda x}{1+x^2} \right) \frac{1+x^2}{x^2} \frac{nx^2}{1+x^2} dF_n(x) \right]. \end{aligned} \quad (9)$$

Если положить

$$a_n = \int \frac{nx}{1+x^2} dF_n(x), \quad d\Psi_n(x) = \frac{nx^2}{1+x^2} dF_n(x), \quad (10)$$

то справа будет стоять $\lim \psi_n$, $\psi_n = \psi(\lambda; a_n, \Psi_n)$.

Допустим теперь на минуту, что для функций из \mathcal{L} справедлива теорема непрерывности:

Лемма 4. Если $\psi_n = \psi(\lambda; a_n, \Psi_n) \rightarrow \psi$ и ψ непрерывна в точке $\lambda = 0$, то $\psi(\lambda)$ имеет вид $\psi(\lambda; a, \Psi)$, $a_n \rightarrow a$, $\Psi_n \Rightarrow \Psi$ (последнее, как и для функций распределения, означает сходимость в точках непрерывности, $\Psi_n(\pm\infty) \rightarrow \Psi(\pm\infty)$).

Если эта лемма верна, то требуемое утверждение очевидным образом следует из (9), (10). Осталось лишь провести

Доказательство леммы 4. Отметим сначала, что соответствие $\psi(\lambda; a, \Psi) \leftrightarrow (a, \Psi)$ взаимно однозначно. Так как в одну сторону это очевидно, то надо убедиться

лишь, что ψ однозначно определяет a и Ψ . Каждому ψ поставим в соответствие функцию

$$\begin{aligned}\gamma(\lambda) &= \int_0^1 \left[\psi(\lambda) - \frac{\psi(\lambda+h) - \psi(\lambda-h)}{2} \right] dh = \\ &= \int_0^1 \int \left(e^{i\lambda x} - \frac{e^{i(\lambda+h)x} - e^{i(\lambda-h)x}}{2} \right) \frac{1+x^2}{x^2} d\Psi(x) dh,\end{aligned}$$

где

$$\begin{aligned}\frac{e^{i(\lambda+h)x} + e^{i(\lambda-h)x}}{2} &= e^{i\lambda x} \cos hx, \\ \int_0^1 e^{i\lambda x} (1 - \cos hx) dh &= e^{i\lambda x} \left(1 - \frac{\sin x}{x} \right), \\ 0 < c_1 < \left(1 - \frac{\sin x}{x} \right) \frac{1+x^2}{x^2} < c_2 < \infty.\end{aligned}$$

Поэтому $\gamma(\lambda) = \int e^{i\lambda x} d\Gamma(x)$, где

$$\Gamma(x) = \int_{-\infty}^x \left(1 - \frac{\sin u}{u} \right) \frac{1+u^2}{u^2} d\Psi(u) \quad (\Psi(x) = \int_{-\infty}^x \frac{u^2}{1+u^2} \left(1 - \frac{\sin u}{u} \right)^{-1} d\Gamma(u))$$

ведет себя (с точностью до постоянного множителя) как функция распределения, для которой $\gamma(\lambda)$ играет роль х.ф. Итак, мы получили цепочку однозначных соответствий $\psi \rightarrow \gamma \rightarrow \Gamma \rightarrow \Psi$, доказывающих утверждение.

Вернемся к доказательству леммы 4. Так как $e^{\psi_n} \rightarrow e^{\psi}$, e^{ψ_n} есть х.ф. и e^{ψ} непрерывна в точке $\lambda = 0$, то e^{ψ} есть х.ф. и, следовательно, непрерывная функция. Это значит, что сходимость $\psi_n \rightarrow \psi$ равномерна на любом отрезке,

$$\begin{aligned}\gamma_n(\lambda) &= \int_0^1 \left[\psi_n(\lambda) - \frac{\psi_n(\lambda+h) + \psi_n(\lambda-h)}{2} \right] dh \rightarrow \\ &\rightarrow \int_0^1 \left[\psi(\lambda) - \frac{\psi(\lambda+h) + \psi(\lambda-h)}{2} \right] dh = \gamma(\lambda),\end{aligned}$$

и функция $\gamma(u)$ непрерывна. По теореме непрерывности для х.ф. это означает, что $\gamma(u)$ есть х.ф. (конечной меры Γ), $\Gamma_n \Rightarrow \Gamma$ (Γ_n есть прообраз γ_n), $\Psi_n \Rightarrow \Psi$, $a_n \rightarrow a$. Итак, установлено, что

$$\begin{aligned}\psi &= \lim \psi_n = \lim \left[i\lambda a_n + \int \left(e^{i\lambda x} - 1 - \frac{i\lambda x}{1+x^2} \right) d\Psi_n(x) \right] = \\ &= i\lambda a + \int \left(e^{i\lambda x} - 1 - \frac{i\lambda x}{1+x^2} \right) d\Psi(x) = \psi(\lambda; a, \Psi).\end{aligned}$$

Сделаем несколько замечаний, относящихся к структуре процесса $\xi(t)$ и его связи с представлением (8). Функция Ψ в (8) соответствует так называемой *спектральной мере процесса* $\xi(t)$. Ее можно представить в виде $\mu\Psi_1(x)$, где

$\mu = \Psi(\infty) - \Psi(-\infty)$, $\Psi_1(x)$ есть функция распределения. (Мы используем также запись $\Psi(A) = \int_A d\Psi(x)$.)

1) Винеровскому процессу соответствует спектральная мера, сосредоточенная в точке 0. Если $\Psi(\{0\}) = \sigma^2$, то $\xi(1) \in \Phi_{a, \sigma^2}$.

2) Пуассоновскому процессу (обобщенному) соответствует мера Ψ такая, что $\int \frac{d\Psi(x)}{|x|^2} < \infty$. В этом случае

$$G(x) = \int_{-\infty}^x \frac{1+u^2}{u^2} d\Psi(u)$$

обладает свойствами функции распределения и $\psi(\lambda; a\Psi)$ можно записать в виде

$$i\lambda a_1 + \int (e^{i\lambda x} - 1) dG(x),$$

где $a_1 = a - \int \frac{d\Psi(x)}{x}$.

3) Рассмотрим теперь общий случай, но при условии $\Psi(\{0\}) = 0$. Как мы знаем, ψ можно приблизить при малых Δ ($\Delta_k = ((k-1)\Delta, k\Delta)$) с помощью выражений вида

$$i\lambda a + \sum_{\substack{k=-\infty \\ k \neq 0}}^{\infty} \left[-\frac{i\lambda}{k\Delta} \Psi(\Delta_k) + (e^{i\lambda k\Delta} - 1) \frac{1 + (k\Delta)^2}{(k\Delta)^2} \Psi(\Delta_k) \right],$$

что соответствует сумме пуассоновских процессов со скачками $k\Delta$ и интенсивностью $\frac{1+(k\Delta)^2}{(k\Delta)^2} \Psi(\Delta_k)$. Если, скажем, $\int_{+0}^{\infty} \frac{d\Psi(x)}{x^2} = \infty$, то суммарная интенсивность скачков из $(0, \varepsilon)$ этих процессов при любом $\varepsilon > 0$ будет неограниченно возрастать при $\Delta \rightarrow 0$. Это значит, что с вероятностью 1 на любом временном интервале δ появится хотя бы один скачок меньше любого заданного $\varepsilon > 0$, так что траектории $\xi(t)$ будут всюду разрывны. Для «компенсации» этих скачков добавляется снос величиной $\frac{\Psi(\Delta_k)}{k\Delta}$, который также может быть «в сумме» неограниченным (если $\int_{+0}^{\infty} \frac{d\Psi(x)}{x} = \infty$).

4) Устойчивым процессам (см. § 8.9) соответствуют гладкие на полуосях степенные функции $\Psi(x)$, обладающие свойством $c_1 \Psi'(x) = \Psi'(c_2 x)$ при подходящих c_1, c_2 .

Г л а в а 19

Функциональные предельные теоремы

§ 1. Сходимость к винеровскому процессу (принцип инвариантности)

Уже отмечалось в §18.2, что винеровские процессы являются в известном смысле предельными для случайных ломаных, построенных по точкам $(k/n, S_k/\sqrt{n})$, где $S_k = \xi_1 + \dots + \xi_k$ — суммы независимых одинаково распределенных случайных величин ξ_1, ξ_2, \dots . Придадим теперь этому высказыванию более точный и общий смысл.

Пусть

$$\xi_{1,n}, \dots, \xi_{n,n} \quad (1)$$

— независимые случайные величины в схеме серий

$$\zeta_{k,n} = \sum_{j=1}^k \xi_{j,n}, \quad E\xi_{k,n} = 0, \quad E\xi_{k,n}^2 = \sigma_{k,n}^2$$

с конечными третьими моментами $E|\xi_{k,n}|^3 = \mu_{k,n} < \infty$.

Будем предполагать, не ограничивая общности (см. §8.4), что $D\zeta_{n,n} = \sum_{j=1}^n \sigma_{j,n}^2 = 1$. Положим $t_{k,n} = \sum_{j=1}^k \sigma_{j,n}^2$, так что $t_{0,n} = 0$, $t_{n,n} = 1$, и рассмотрим случайную непрерывную ломаную с узлами в точках (t_k, ζ_k) , где для краткости опущен второй индекс n : $t_k = t_{k,n}$, $\zeta_k = \zeta_{k,n}$.

Мы получили случайный процесс на $[0, 1]$ с непрерывными траекториями, который обозначим $s_n = s_n(t)$ (рис. 17). Принцип инвариантности (или функциональная центральная предельная теорема; смысл первого названия будет прокомментирован позже) состоит в том, что для любого функционала f , заданного на пространстве $C(0, 1)$ и непрерывного в равномерной метрике, распределение $f(s_n)$ слабо сходится:

$$f(s_n) \Rightarrow f(w), \quad (2)$$

к распределению $f(w)$, где $w = w(t)$ — стандартный винеровский процесс. Обычная центральная предельная теорема является частным случаем этого факта (надо в качестве $f(x)$ взять $x(1)$).

Высказанное утверждение эквивалентно каждому из следующих двух фактов:

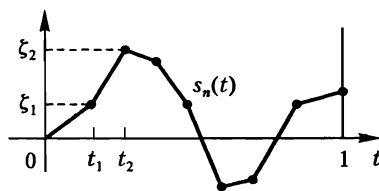


Рис. 17.

1. Для любого ограниченного и непрерывного функционала f

$$Ef(s_n) \rightarrow Ef(w). \quad (3)$$

2. Для любого G из σ -алгебры борелевских множеств $\mathfrak{B}_{C(0,1)}$ в пространстве $C(0,1)$ ($\mathfrak{B}_{C(0,1)}$ порождена открытыми множествами в метрическом пространстве $C(0,1)$ с равномерной метрикой; уже отмечалось, что $\mathfrak{B}_{C(0,1)} = \mathfrak{B}_C^{[0,1]}$); такого, что $P(w \in \partial G) = 0$, где ∂G — граница множества G , выполняется

$$P(s_n \in G) \rightarrow P(w \in G). \quad (4)$$

Соотношения (3), (4) суть эквивалентные определения слабой сходимости распределений P_n процессов s_n к распределению W винеровского процесса w в пространстве $\langle C(0,1), \mathfrak{B}_{C(0,1)} \rangle$. Более подробные сведения об этом можно найти в Приложении 3 и в [1, 7].

Основное содержание этого параграфа составляют следующие теоремы.

Обозначим, как и прежде, $L_3 = \sum_{k=1}^n \mu_{k,n}$.

Теорема 1. Пусть $L_3 \rightarrow 0$ при $n \rightarrow \infty$ (условие Ляпунова). Тогда справедливы соотношения сходимости (2)–(4).

Замечание 1. Условие $L_3 \rightarrow 0$ здесь можно ослабить до условия Линдеберга. В таком виде высказанная теорема сходимости носит наименование *принципа инвариантности Донскера—Прохорова*.

Наряду с теоремой 1 получим и более точное утверждение.

Определение 1. Множество G назовем *липщицевым*, если $\Delta W_G(\varepsilon) \equiv W(G^{(\varepsilon)}) - W(G) \leq c\varepsilon$, где $G^{(\varepsilon)}$ есть ε -окрестность G , W — мера, соответствующая винеровскому процессу.

В дальнейшем буквой c (с индексами или без) будем обозначать абсолютные постоянные, не обязательно одни и те же.

Теорема 2. Если G липщицево, то

$$|P(s_n \in G) - P(w \in G)| < cL_3^{1/4}. \quad (5)$$

В случае $\xi_{k,n} = \xi_k/\sqrt{n}$ (ξ_k от n не зависят и одинаково распределены, $E\xi_k = 0$, $D\xi_k = 1$) правая часть в (5) превращается в $cn^{-1/8}$.

Аналогичная оценка может быть получена для функционалов. Функционал f в $C(0,1)$ назовем *липщицевым*, если выполнены два условия:

- 1) $|f(x) - f(y)| < c\rho(x, y)$,
- 2) распределение $f(w)$ имеет ограниченную плотность.

Следствие 1. Если f — липщицев функционал, то множество $G_v = \{f(x) < v\}$ липщицево (с одной и той же постоянной при всех v), так что в силу теоремы 2

$$\sup_v |P(f(w) < v) - P(f(s_n) < v)| \leq cL_3^{1/4}.$$

Сформулированные выше теоремы будут следствиями формулируемой далее теоремы 3.

Пусть

$$\eta_{1,n}, \dots, \eta_{n,n} \quad (6)$$

— любая другая последовательность независимых случайных величин в схеме серий с теми же первыми двумя моментами $E\eta_{k,n} = 0$, $E\eta_{k,n}^2 = \sigma_{k,n}^2$ и с конечными третьими моментами. Обозначим $F_{k,n}$, $\Phi_{k,n}$ функции распределения $\xi_{k,n}$ и $\eta_{k,n}$ соответственно,

$$\begin{aligned}\nu_{k,n} &= E|\eta_{k,n}|^3 < \infty, \quad N_3 = \sum_{k=1}^n \nu_{k,n}, \\ \mu_{k,n}^0 &= \int |x|^3 |d(F_{k,n}(x) - \Phi_{k,n}(x))| \leq \mu_{k,n} + \nu_{k,n}, \\ L_3^0 &= \sum_{k=1}^n \mu_{k,n}^0 \leq L_3 + N_3.\end{aligned}$$

Обозначим через $s'_n(t)$ случайный процесс, построенный так же, как процесс $s_n(t)$, но по последовательности $\{\eta_{k,n}\}$.

Теорема 3. Для любого $A \in \mathfrak{B}_{C(0,1)}$

$$P(s_n \in A) \leq P(s'_n \in A^{(2\varepsilon)}) + \frac{cL_3^0}{\varepsilon^3}.$$

Чтобы доказать теорему 3, получим сначала ее конечномерный аналог. Обозначим через ζ и η векторы $\zeta = (\zeta_1, \dots, \zeta_n)$, $\eta = (\eta_1, \dots, \eta_n)$, где $\zeta_k = \sum_{j=1}^k \xi_{j,n}$, $\eta_k = \sum_{j=1}^k \eta_{j,n}$, и обозначим $B^{(\varepsilon)}$ ε -окрестность множества $B \in \mathbb{R}^n$:

$$B^{(\varepsilon)} = \bigcup_{\substack{x \in B \\ |v| \leq \varepsilon}} (x + v),$$

где $x = (x_1, \dots, x_n)$, $v = (v_1, \dots, v_n)$, $|v| = \max_k |v_k|$.

Лемма 1. Пусть B — любое борелевское множество из \mathbb{R}^n . Тогда

$$P(\zeta \in B) \leq P(\eta \in B^{(2\varepsilon)}) + \frac{cL_3^0}{\varepsilon^3}.$$

*Доказательство*¹⁾. Введем в рассмотрение совокупность расширяющихся окрестностей

$$B^{(\varepsilon)}(k) = \bigcup_{\substack{x \in B \\ |v| \leq \varepsilon}} (x_1, \dots, x_k, x_{k+1} + v_{k+1}, \dots, x_n + v_n), \quad k = 0, \dots, n,$$

$$B = B^{(\varepsilon)}(n) \subset B^{(\varepsilon)}(n-1) \subset \dots \subset B^{(\varepsilon)}(1) \subset B^{(\varepsilon)}(0) = B^{(\varepsilon)}$$

и обозначим через e_k вектор $(0, \dots, 0, 1, 0, \dots, 0)$, где 1 стоит на месте k -й координаты. Очевидно, что, если $x \in B^{(\varepsilon)}(k)$, то

$$x + e_k v_k \in B^{(\varepsilon)}(k-1) \quad \text{при} \quad |v_k| \leq \varepsilon. \quad (7)$$

¹⁾ В основе приводимого ниже доказательства лежит предложенное А. И. Саханенко обобщение подхода к доказательству центральной предельной теоремы, изложенного в § 8.5.

Далее, наряду с сериями (1), (6) рассмотрим совокупность «переходных» последовательностей

$$\xi_{1,n}, \dots, \xi_{k,n}, \eta_{k+1,n}, \dots, \eta_{n,n}, \quad k = 0, \dots, n. \quad (8)$$

Обозначим $\zeta(k) = (\zeta_1(k), \dots, \zeta_n(k))$ векторы, образованные накопленными суммами k -й последовательности, так что

$$\begin{aligned} \zeta_j(k) &= \zeta_j \quad \text{при } j \leq k \\ \zeta_j(k) &= \zeta_k + \eta_{k+1,n} + \dots + \eta_{j,n} \quad \text{при } j > k. \end{aligned}$$

Лемма 2. Для любой случайной величины δ такой, что $P(|\delta| \leq \varepsilon) = 1$, выполняется

$$P(\zeta \in B) \leq P(\eta \in B^{(2\varepsilon)}) + \sum_{k=1}^n \Delta_k, \quad (9)$$

где

$$\begin{aligned} \Delta_k &= P(\zeta(k) + \delta e(k-1) \in B^{(\varepsilon)}(k-1)) - P(\zeta(k-1) + \delta e(k-1) \in B^{(\varepsilon)}(k-1)), \\ e(r) &= \sum_{r+1}^n e_j = (0, \dots, 0, 1, \dots, 1). \end{aligned}$$

Доказательство. Действительно, в силу (7)

$$\begin{aligned} P(\zeta \in B) &= P(\zeta(n) \in B^{(\varepsilon)}(n)) \leq P(\zeta(n) + e(n)\delta \in B^{(\varepsilon)}(n-1)) = \\ &= P(\zeta(n) + e(n-1)\delta \in B^{(\varepsilon)}(n-1)) \equiv P(\zeta(n-1) + e(n-1)\delta \in B^{(\varepsilon)}(n-1)) + \Delta_n. \end{aligned}$$

Применяя снова те же выкладки к правой части, получим

$$\begin{aligned} P(\zeta(n-1) + e(n-1)\delta \in B^{(\varepsilon)}(n-1)) &\leq \\ &\leq P(\zeta(n-1) + e(n-1)\delta + e_{n-1}\delta \in B^{(\varepsilon)}(n-2)) = \\ &= P(\zeta(n-1) + e(n-2)\delta \in B^{(\varepsilon)}(n-2)) \equiv \\ &\equiv P(\zeta(n-2) + e(n-2)\delta \in B^{(\varepsilon)}(n-2)) + \Delta_{n-1}. \end{aligned}$$

$$\begin{aligned} &\dots \dots \dots \\ P(\zeta(1) + e(1)\delta \in B^{(\varepsilon)}(1)) &\leq P(\zeta(1) + e(1)\delta + e_1\delta \in B^{(\varepsilon)}(0)) = \\ &= P(\zeta(1) + e(0)\delta \in B^{(\varepsilon)}(0)) \equiv P(\zeta(0) + e(0)\delta \in B^{(\varepsilon)}(0)) + \Delta_1. \end{aligned}$$

Так как $\zeta(0) = \eta$, $P(\eta + e(0)\delta \in B^{(\varepsilon)}) \leq P(\eta \in B^{(2\varepsilon)})$, то (9) доказано. ■

Чтобы получить лемму 1, надо оценить теперь Δ_k . Нам будет удобно вместе с (8) рассматривать также последовательности

$$\xi_{1,n}, \dots, \xi_{k-1,n}, y, \eta_{k+1,n}, \dots, \eta_{n,n}$$

и обозначать $\zeta(k, y) = (\zeta_1(k, y), \dots, \zeta_n(k, y))$ соответствующий вектор накопленных сумм, так что

$$\begin{aligned} \zeta(k, \xi_{k,n}) &= \zeta(k) = \zeta(k, 0) + \xi_{k,n}e_{(k-1)}, \\ \zeta(k, \eta_{k,n}) &= \zeta(k-1) = \zeta(k, 0) + \eta_{k,n}e_{(k-1)}. \end{aligned}$$

Тогда Δ_k можно записать в виде

$$\Delta_k = P(\zeta(k, 0) + (\delta + \xi_{k,n})e(k-1) \in B^{(\epsilon)}(k-1)) - \\ - P(\zeta(k, 0) + (\delta + \eta_{k,n})e(k-1) \in B^{(\epsilon)}(k-1)). \quad (10)$$

Возьмем в качестве δ случайную величину, не зависящую от ζ и η . Тогда вероятности, входящие в (10), удобно оценивать, пользуясь у.м.о., так как, например, в равенстве

$$P(\zeta(k, 0) + (\delta + \xi_{k,n})e(k-1) \in B^{(\epsilon)}(k-1)) = \\ = EP((\delta + \xi_{k,n})e(k-1) \in B^{(\epsilon)}(k-1) - \zeta(k, 0) | \zeta(k, 0)). \quad (11)$$

множество $C = B^{(\epsilon)}(k-1) - \zeta(k, 0)$ можно считать фиксированным (см. свойства у.м.о., δ и $\xi_{k,n}$ от $\zeta(k, 0)$ не зависят). Обозначим через D множество y , для которых $ye(k-1) \in C$. Тогда надо оценить разность

$$P(\delta + \xi_{k,n} \in D) - P(\delta + \eta_{k,n} \in D). \quad (12)$$

Воспользуемся леммой 8.3. Чтобы привести (12) к виду, удобному для применения леммы, выберем в качестве δ случайную величину, имеющую трижды непрерывно дифференцируемую плотность $g(t)$, и положим для краткости $\xi_{k,n} = \xi$, $\eta_{k,n} = \eta$. Тогда $\delta + \xi$ будет иметь плотность, равную $\int dF_\xi(t)g(y-t) = Eg(y-\xi)$, так что

$$P(\delta + \xi \in D) = \int_D Eg(y-\xi)dy = E \int_D g(y-\xi)dy.$$

Если положим теперь $f(x) = \int g(y-x)dy$, то получим, что

$$P(\delta + \xi \in D) = Ef(\xi),$$

где f — трижды непрерывно дифференцируемая функция,

$$|f'''(x)| \leq \int |g'''(y)|dy \equiv f_3.$$

Применяя теперь лемму 8.3, получим, что

$$|P(\delta + \xi \in D) - P(\delta + \eta \in D)| = |E(f(\xi) - f(\eta))| \leq \frac{f_3}{6} \mu_{k,n}^0, \\ \mu_{k,n}^0 = \int |x^3| |d(F_{k,n}(x) - \Phi_{k,n}(x))|.$$

Так как правая часть здесь от $\zeta(k, 0)$ и D никак не зависит, то, возвращаясь к (10), (11), мы получим ту же оценку

$$|\Delta_k| \leq \frac{f_3}{6} \mu_{n,k}^0. \quad (13)$$

Пусть теперь $g_1(x)$ — некоторая гладкая плотность, сосредоточенная на $[-1, 1]$. Тогда, полагая $g(x) = 1/\epsilon g_1(\frac{x}{\epsilon})$, получим

$$f_3 = \int \frac{1}{\epsilon^4} \left| g_1''' \left(\frac{y}{\epsilon} \right) \right| dy = \frac{1}{\epsilon^3} \int |g_1'''(y)| dy = \frac{c_1}{\epsilon^3}, \quad c_1 = \text{const}. \quad (14)$$

Сопоставляя (9), (13) и (14), мы получим утверждение леммы 1. ■

Доказательство теоремы 3. Эта теорема есть следствие леммы 1. Действительно, пусть $B \in \mathbb{R}^n$ таково, что события $\{s_n \in A\}$ и $\{\zeta \in B\}$ эквивалентны (s_n полностью определяется по ζ), тогда, очевидно, $\{s_n \in A^{(\varepsilon)}\} = \{\zeta \in B^{(\varepsilon)}\}$, и утверждение теоремы 3 повторяет утверждение леммы 1. ■

Доказательство теоремы 1. Пусть $w(t)$ — стандартный винеровский процесс. Положим $\eta_{k,n} = w(t_{k,n}) - w(t_{k-1,n})$. Тогда последовательность $\eta_{1,n}, \dots, \eta_{n,n}$ удовлетворяет всем наложенным выше условиям, так как

$$E\eta_{k,n} = 0, \quad E\eta_{k,n}^2 = \sigma_{k,n}^2, \quad \nu_{k,n} = E|\xi_{k,n}|^3 = c_3\sigma_{k,n}^3 < \infty.$$

Отметим также, что $\sigma_{k,n}^2 = (E|\xi_{k,n}|^2)^{3/2} \leq E|\xi_{k,n}|^3 = \mu_{k,n}$, так что

$$N_3 = \sum_{k=1}^n \nu_{k,n} = c_3 \sum_{k=1}^n \sigma_{k,n}^3 \leq c_3 L_3 \rightarrow 0.$$

Нам понадобится

Лемма 3. $P(\rho(w, s'_n) > \varepsilon) \leq cN_3/\varepsilon^3$.

Доказательство. Событие $\{\rho(w, s'_n) > \varepsilon\}$ равно $\bigcup_k A_k$, где

$$A_k = \left\{ \sup_{\delta_k} |w(t) - s'(t)| > \varepsilon \right\} \subset \left\{ \sup_{\delta_k} |w(t)| > \frac{\varepsilon}{2} \right\},$$

$$\delta_k = [t_{k-1}, t_k].$$

Поэтому ($t_k - t_{k-1} = \sigma_{k,n}^2$, $w(t) = \sigma w(t/\sigma^2)$)

$$P(A_k) \leq P\left(\sup_{[0,1]} |w(t)| > \frac{\varepsilon}{2\sigma_{k,n}}\right) \leq 2\left(1 - \Phi\left(\frac{\varepsilon}{2\sigma_{k,n}}\right)\right).$$

Функция $(1 - \Phi(t))$ убывает при $t \rightarrow \infty$ значительно быстрее, чем t^{-3} . Поэтому

$$2\left(1 - \Phi\left(\frac{\varepsilon}{2\sigma_{k,n}}\right)\right) \leq c \frac{\sigma_{k,n}^3}{\varepsilon^3},$$

$$P\left(\bigcup_k A_k\right) \leq \frac{cN_3}{\varepsilon^3}.$$

■

Из доказательства видно, что оценка в лемме 3 является весьма грубой. Вернемся к доказательству теоремы 1. Так как

$$P(s'_n \in G) = P(s'_n \in G, \rho(w, s'_n) \leq \varepsilon) + P(s'_n \in G, \rho(w, s'_n) > \varepsilon),$$

то

$$P(s'_n \in G) \leq P(w \in G^{(\varepsilon)}) + \frac{cN_3}{\varepsilon^3} \quad (15)$$

и в силу теоремы 3

$$P(s_n \in G) \leq P(w \in G^{(3\varepsilon)}) + \frac{c(L_3^0 + N_3)}{\varepsilon^3}.$$

Получим теперь обратное неравенство. Обозначим $G^{(-\varepsilon)}$ множество $G^{-\varepsilon} = G - (\partial G)^{(\varepsilon)}$, тогда $[G^{(-\varepsilon)}]^{(\varepsilon)} \equiv G^0 \subset G$. Меняя местами s_n и s'_n в теореме 3 и применяя ее к множеству $G^{(-2\varepsilon)}$, получим

$$P(s_n \in G^0) \geq P(s'_n \in G^{(-2\varepsilon)}) - \frac{c(L_3^0 + N_3)}{\varepsilon^3}. \quad (16)$$

Меняя местами w и s'_n в (15) и применяя это соотношение к $G^{(-3\varepsilon)}$, находим

$$P(s'_n \in G^{(-2\varepsilon)}) \geq P(w \in G^{(-3\varepsilon)}) - \frac{cN_3}{\varepsilon^3}.$$

Отсюда и из (16) получаем

$$P(s_n \in G) \geq P(s_n \in G^0) \geq P(w \in G^{(-3\varepsilon)}) - \frac{c(L_3^0 + N_3)}{\varepsilon^3}.$$

Если обозначить

$$P(w \in G^{(\varepsilon)}) - P(w \in G) = W(G^{(\varepsilon)}) - W(G) = \Delta W_G(\varepsilon)$$

и учесть, что $N_3 \leq cL_3$, $L_3^0 \leq L_3 + N_3$, то мы получили

$$-\Delta W_G(-3\varepsilon) + \frac{cL_3}{\varepsilon^3} \leq P(s_n \in G) - W(G) \leq \Delta W_G(3\varepsilon) + \frac{cL_3}{\varepsilon^3}. \quad (17)$$

Если $W(\partial G) = 0$, то очевидно, что

$$W(G^{(3\varepsilon)}) - W(G^{(-3\varepsilon)}) \rightarrow 0$$

при $\varepsilon \rightarrow 0$, $\Delta W_G(\pm 3\varepsilon) \rightarrow 0$. Отсюда и из (17) уже легко получить, что $P(s_n \in G) \rightarrow P(w \in G)$.

Сходимость $f(s_n) \Rightarrow f(w)$ для непрерывных функционалов следует из (4), так как множество $G_v = \{x \in C(0, 1): f(x) < v\}$ для точки v , являющейся точкой непрерывности распределения $f(w)$, обладает свойством

$$W(\partial G_v) = P(f(w) = v) = 0$$

и, следовательно,

$$P(f(s_n) < v) \rightarrow P(f(w) < v).$$

Доказательство теоремы 2. Если множество G липшицево, то

$$|\Delta W_G(\pm 3\varepsilon)| < c\varepsilon$$

и в силу (17)

$$|P(s_n \in G) - W(G)| < c\left(\varepsilon + \frac{L_3}{\varepsilon^3}\right).$$

Полагая $\varepsilon = L_3^{1/4}$, получим требуемое утверждение. ■

Название «принцип инвариантности», относящееся к основным утверждениям этого параграфа, лучше всего иллюстрируется теоремой 3, в силу которой для приближенного вычисления $P(s_n \in A)$ можно пользоваться значением $P(s'_n \in A)$ для любой другой последовательности (6), обладающей лишь теми же первыми двумя моментами, что и (1). В этом смысле главная часть $P(s_n \in A)$ инвариантна относительно конкретных распределений рассматриваемых последовательностей.

Например, вычисление $P(s_n \in G)$ или $P(w(t) \in G)$ можно заменить вычислением $P(s'_n \in G)$ для последовательности в схеме Бернулли, к которой удобно применять различные численные методы. С другой стороны, вероятности $P(w \in G)$ для целого класса областей G найдены в явном виде (см., например, [20]). Мы знаем, например, что $P(\sup w(t) > y) = 2(1 - \Phi(y))$. (Отсюда следует, в частности, что множество $G = \{x \in C(0, 1) : \sup x(t) > y\}$ является липшицевым.) Поэтому для распределения максимума сумм $\bar{S}_n = \max_{k \leq n} S_k$, $S_k = \sum_{j=1}^k \xi_j$, $E\xi_k = 0$, $D\xi_k = \sigma^2$ будем иметь

$$P(\bar{S}_n > x\sigma\sqrt{n}) \rightarrow 2(1 - \Phi(x)),$$

и мы можем использовать это соотношение для приближенного вычисления распределения \bar{S}_n , представляющего, как мы видели в гл. 11, значительный интерес в приложениях.

Точно так же можно приближать совместное распределение S_n , \bar{S}_n и $\underline{S}_n = \min_{k \leq n} S_k$ (т.е. вероятности вида $P(\bar{S}_n < x\sqrt{n}$, $\underline{S}_n > y\sqrt{n}$, $S_n \in B)$) с помощью соответствующих формул для винеровского процесса, приведенных в [20].

Замечание 2. В заключение этого параграфа отметим, что все утверждения, высказанные выше, сохраняются, если в них вместо $s_n(t)$ рассматривать не непрерывную, а ступенчатую ломаную $s_n^*(t) = \zeta_{k,n}$ при $t \in [t_k, t_{k+1})$. В этом можно убедиться, проводя заново все рассмотрения применительно к s_n^* . Другой путь для получения, скажем, теорем 1, 2 для s_n^* состоит в том, чтобы воспользоваться уже полученными результатами и оценить расстояние $\rho(s_n, s_n^*)$. Так как $\{\rho(s_n, s_n^*) > \varepsilon\} \subset \bigcup_{k=1}^n \{|\xi_{k,n}| > \varepsilon\}$, то

$$P(\rho(s_n, s_n^*) > \varepsilon) \leq \sum_{k=1}^n P(|\xi_{k,n}| > \varepsilon) \leq \sum_{k=1}^n \frac{\mu_{k,n}}{\varepsilon^3} = \frac{L_3}{\varepsilon^3}.$$

Напомним, что оценку такого же вида мы получили для $\rho(s'_n, w)$ и это позволило заменить, где это требовалось, s'_n на w . Следовательно, пользуясь точно такими же рассуждениями, можно менять s_n на s_n^* . В этом случае можно рассматривать сходимость распределений функционалов $f(s_n^*)$, определенных на $D(0, 1)$ (и непрерывных в равномерной метрике ρ). В ряде случаев использование s_n^* бывает технически более удобным, чем s_n . Например, в случае, когда надо найти предельное распределение $\sum_{k=1}^n \varphi(\zeta_{k,n}) = n \int \varphi(s_n^*(t)) dt$ ($\zeta_{k,n}$ одинаково распределены). Из выписанного представления следует, что $\frac{1}{n} \sum_{k=1}^n \varphi(\zeta_{k,n}) \Rightarrow \int \varphi(w(t)) dt$.

§ 2. Закон повторного логарифма

Пусть ξ_1, ξ_2, \dots — последовательность независимых случайных величин,

$$E\xi_k = 0, \quad E\xi_k^2 = \sigma_k^2, \quad E|\xi_k|^3 = \mu_k, \\ S_n = \sum_{k=1}^n \xi_k, \quad B_n^2 = \sum_{k=1}^n \sigma_k^2, \quad M_n = \sum_{k=1}^n \mu_k.$$

В этих обозначениях отношение Ляпунова равно

$$L_3 = L_{3,n} = \frac{M_n}{B_n^3}.$$

В этом параграфе мы покажем, что из закона повторного логарифма для винеровского процесса и из теоремы 2 следует

Теорема 4 (закон повторного логарифма для сумм случайных величин). Если $B_n \rightarrow \infty$, $L_{3,n} < c / \ln B_n$ при некотором $c < \infty$, то

$$\begin{aligned} P\left(\overline{\lim}_{n \rightarrow \infty} \frac{S_n}{B_n \sqrt{2 \ln \ln B_n}} = 1\right) &= 1, \\ P\left(\underline{\lim}_{n \rightarrow \infty} \frac{S_n}{B_n \sqrt{2 \ln \ln B_n}} = -1\right) &= 1. \end{aligned}$$

Таким образом, все последовательности, располагающиеся выше $(1 + \varepsilon) \times B_n \sqrt{2 \ln \ln B_n}$, будут верхними для последовательности S_n , а все последовательности ниже $(1 - \varepsilon) B_n \sqrt{2 \ln \ln B_n}$ — нижними.

Условия теоремы, очевидно, будут выполнены для одинаково распределенных ξ_k , так как в этом случае $L_{3,n} = \mu_1 / (\sigma_1^3 \sqrt{n})$, $B_n^2 = \sigma_1^2 n$.

Доказательство. Обратимся к доказательству закона повторного логарифма в теореме 18.5 и применим его к последовательности S_n . При этом не потребуются вносить никаких существенных изменений. Надо лишь вместо $w(a^k)$ рассматривать S_{n_k} , где $n_k = \min\{n: B_n^2 \geq a^k\}$, и заменить, где потребуется, a^k на $B_{n_k}^2$. В силу условия Ляпунова $\max_{k \leq n} \sigma_k^2 = o(B_n^2)$, так что $B_{n_{k-1}}^2 \sim B_{n_k}^2 \sim a^k$ при $k \rightarrow \infty$.

Ключевым моментом доказательства теоремы 18.5 является доказательство сходимости ряда (при любом $\varepsilon > 0$)

$$\sum_k P\left(\sup_{u \leq a^k} w(u) > (1 + \varepsilon) x_{k-1}\right) \quad (18)$$

и расходимости ряда

$$\sum_k P\left(w(a^k) - w(a^{k-1}) > \left(1 - \frac{\varepsilon}{2}\right) x_k\right), \quad (19)$$

где

$$x_k = \sqrt{2a^k \ln \ln a^k}, \quad w(a^k) - w(a^{k-1}) = w(a^k(1 - a^{-1})).$$

В рассматриваемом случае, если следовать тому же пути, надо доказать сходимость ряда

$$\sum_k P(\bar{S}_{n_k} > (1 + \varepsilon) y_{k-1}) \quad (18a)$$

и расходимость ряда

$$\sum_k P(S_{n_k} - S_{n_{k-1}} > \left(1 - \frac{\varepsilon}{2}\right) y_k), \quad (19a)$$

где $y_k = \sqrt{2B_{n_k}^2 \ln \ln B_{n_k}^2} \sim x_k$. Но асимптотическое поведение вероятностей событий в (18), (18a) и в (19), (19a) при выполнении условия $L_{3,n} < c / \ln B_n$ будет по существу одинаковым. Чтобы установить это, воспользуемся неравенством

$$\left|P\left(\frac{S_n}{B_n} \in G\right) - P(w \in G^{(\pm \delta)})\right| < \frac{cL_{3,n}}{\delta^3}, \quad (20)$$

вытекающим из доказательства теоремы 3. В силу этого неравенства

$$\begin{aligned} P\left(\frac{\bar{S}_n}{B_n} > (1 + 3\varepsilon)x\right) &\leq P\left(\sup_{u \leq 1} w(u) > (1 + 2\varepsilon)x\right) + \frac{cL_{3,n}}{(\varepsilon x)^3} = \\ &= P\left(\sup_{u \leq B_n^2} w(u) > (1 + 2\varepsilon)x B_n\right) + \frac{cL_{3,n}}{(\varepsilon x)^3}. \end{aligned}$$

Поэтому (см. (18a)), полагая $n = n_k$, $x = y_{k-1}/B_n$, получим

$$P\left(\bar{S}_{n_k} > (1 + 3\varepsilon)y_{k-1}\right) \leq P\left(\sup_{u \leq B_{n_k}^2} w(u) > (1 + 2\varepsilon)y_{k-1}\right) + \frac{c}{\varepsilon^3} L_{3,n_k} (\ln \ln B_{n_k}^2)^{-3/2}.$$

Здесь $y_{k-1} \sim x_{k-1}$, $B_{n_k}^2 \geq a^k$, $\ln \ln B_{n_k}^2 \sim \ln \ln a^k \sim \ln k$, $L_{3,n_k} \leq \frac{c}{\ln B_{n_k}} \sim \frac{c}{\ln a^k} \sim \frac{c_1}{k}$. Следовательно, при всех достаточно больших k (буквой c обозначены различные постоянные)

$$P\left(\bar{S}_{n_k} > (1 + 3\varepsilon)y_{k-1}\right) \leq P\left(\sup_{u \leq a^k} w(u) > (1 + \varepsilon)x_{k-1}\right) + \frac{c}{\varepsilon^3 k (\ln k)^{3/2}}.$$

Так как

$$\sum_{k=1}^{\infty} \frac{1}{k (\ln k)^{3/2}} < \infty, \quad (21)$$

то полученное неравенство означает, что из сходимости ряда (18) вытекает сходимость ряда (18a). Первая часть теоремы доказана.

Вторая часть доказывается аналогично. Рассмотрим ряд (19a). В силу (20) находим

$$\begin{aligned} P(S_{n_k} - S_{n_{k-1}} > (1 - 3\varepsilon)y_k) &= P\left(s_{n_k}(1) - s_{n_k}(r_k) > (1 - 3\varepsilon)\frac{y_k}{B_{n_k}}\right) \geq \\ &\geq P\left(w(1) - w(r_k) > (1 - 2\varepsilon)\frac{y_k}{B_{n_k}}\right) - \frac{cL_{3,n_k}}{\varepsilon^3} (\ln \ln B_{n_k}^2)^{-3/2}, \quad (22) \end{aligned}$$

где $r_k = B_{n_{k-1}}^2 B_{n_k}^{-2} \rightarrow a^{-1}$ в силу того, что

$$B_{n_k}^2 = a^k + \theta_k \sigma_{n_k}^2, \quad 0 \leq \theta_k \leq 1, \quad \sigma_{n_k}^2 = o(B_{n_k}^2).$$

Первое слагаемое в (22) равно

$$\begin{aligned} P\left(w(1 - r_k) > (1 - 2\varepsilon)\frac{y_k}{B_{n_k}}\right) &\geq P\left(w(a^k(1 - r_k)) > (1 - \varepsilon)x_k\right) = \\ &= P\left(w(a^k(1 - a^{-1})) > (1 - \varepsilon)x_k \sqrt{\frac{1 - a^{-1}}{1 - r_k}}\right) \geq \\ &\geq P\left(w(a^k) - w(a^{k-1}) > \left(1 - \frac{\varepsilon}{2}\right)x_k\right). \end{aligned}$$

Как и прежде, в силу (21) ряд, составленный из вторых слагаемых в (22), сходится. Поэтому установленные неравенства означают, что доказанная нами расходимость ряда (19) влечет за собой расходимость ряда (19a). Теорема доказана. ■

На самом деле условия выполнения закона повторного логарифма связаны не столько с существованием третьих моментов, сколько со скоростью сходимости в центральной предельной теореме. В какой-то мере этот факт иллюстрируется следующим утверждением (см. [19]):

Теорема. Если $B_n \rightarrow \infty$, $B_{n+1}/B_n \rightarrow 1$ при $n \rightarrow \infty$

$$\sup_x \left| P\left(\frac{S_n}{B_n} < x\right) - \Phi(x) \right| \leq c(\ln B_n)^{-1-\delta}$$

при некоторых $\delta > 0$, $c < \infty$, то имеет место закон повторного логарифма.

§ 3. Сходимость к пуассоновскому процессу

Приведенные в § 1, 2 теоремы показывают, что винеровский процесс достаточно хорошо описывает эволюцию накопленных сумм при суммировании «обычных» случайных величин $\xi_{k,n}$, удовлетворяющих условию Ляпунова. Оказывается, что подобным же образом пуассоновский процесс описывает эволюцию накопленных сумм, когда случайные величины $\xi_{k,n}$ соответствуют появлению редких событий.

Рассмотрим независимые случайные величины в схеме серий $\xi_{1,n}, \dots, \xi_{n,n}$, которые для простоты будем считать одинаково распределенными. Пусть

$$P(\xi_{k,n} = 0) = 1 - p_n, \quad P(\xi_{k,n} = 1) = p_n, \quad np_n \equiv \mu_n \rightarrow \mu > 0. \quad (23)$$

Читателю предлагается проверить, что в этом случае условие Линдберга (и, стало быть, Ляпунова) выполнено не будет.

Образует случайную функцию $s_n(t)$ на $[0, 1]$ следующим образом. Положим

$$s_n(t) = \zeta_{k,n} = \sum_{j=1}^k \xi_{j,n}, \quad t \in \left[\frac{k-1}{n}, \frac{k}{n}\right), \quad k = 1, \dots, n.$$

Здесь удобнее использовать ступенчатую, а не непрерывную траекторию $s_n(t)$ (ср. с замечанием 2). Утверждение, которое мы получим в этом разделе, аналогично принципу инвариантности и состоит в том, что процесс $s_n(t)$ сходится в известном смысле к пуассоновскому процессу $\xi(t)$ с параметром μ на $[0, 1]$. Эту сходимость можно было бы трактовать, конечно, как слабую сходимость распределений в $D(0, 1)$. Но в рамках настоящей книги это, по-видимому, нецелесообразно, по крайней мере, по двум причинам:

1. Нам пришлось бы для этого вводить метрику в $D(0, 1)$ и изучать ее свойства, что само по себе сложно.

2. Траектории процессов $s_n(t)$ и $\xi(t)$ устроены просто, и характеризовать их близость можно проще и точнее, не привлекая общих концепций. Действительно, как мы видели, траектория $\xi(t)$ на $[0, 1]$ полностью определяется набором случайных величин $(\xi(1); \alpha_1, \dots, \alpha_{\xi(1)})$, где α_k есть точка k -го скачка процесса, $\alpha_{k+1} - \alpha_k \in \Gamma_\mu$. Такую же характеризацию допускают траектории $s_n(t)$: они определяются вектором $(s_n(1); \gamma_1/n, \dots, \gamma_{s_n(1)}/n)$, где $\gamma_1, \gamma_2, \dots$ суть те значения j , для которых $\xi_{j,n} = 1$. Будем считать распределения $s_n(t)$ и $\xi(t)$ близкими, если близки распределения двух названных векторов. Это будет отвечать сходимости процессов в весьма естественном и сильном смысле.

Из сказанного ранее (см. § 18.4) о процессах Пуассона нетрудно понять, что названная сходимость распределений точек скачков $s_n(t)$ эквивалентна сходимости конечномерных распределений $s_n(t)$ к распределениям $\xi(t)$ (ступенчатый характер траекторий $s_n(t)$ известен).

Теорема 5. Можно так построить $s_n(t)$ и $\xi(t)$ на одном вероятностном пространстве, что

$$P\left(\xi(1) = s_n(1); \left|\alpha_k - \frac{\gamma_k}{n}\right| < \frac{1}{n}, k = 1, \dots, \xi(1)\right) \geq 1 - \frac{\mu_n^2}{n}. \quad (24)$$

Другими словами, траектории $\xi(t)$ и $s_n(t)$ с вероятностью, не меньшей, чем $1 - \mu_n^2/n$, отличаются лишь малыми ($< 1/n$) отклонениями точек скачков.

В отличие от теоремы 1 это утверждение уже трудно называть принципом инвариантности, так как разнообразие распределений $\xi_{k,n}$, приводящее к такому результату, не очень велико. По сравнению с теоремой оно может быть расширено до всех распределений таких, что (см. теорему 7.4)

$$P(\xi_{k,n} = 1) = p_n, \quad 1 - P(\xi_{k,n} = 0) \sim p_n, \\ P(\xi_{k,n} \neq 0, \xi_{k,m} \neq 1) = o(p_n).$$

Следствие 2. Пусть B есть множество в \mathbb{R}^k , обладающее свойством (его также естественно называть липшицевым)

$$|P((\alpha_1, \dots, \alpha_k) \in B^{(\varepsilon)}) - P((\alpha_1, \dots, \alpha_k) \in B)| < c\varepsilon.$$

Тогда из теоремы 5 вытекает, что

$$\left|P\left(s_n(1) = k, \left(\frac{\gamma_1}{n}, \dots, \frac{\gamma_k}{n}\right) \in B\right) - P(\xi(1) = k, (\alpha_1, \dots, \alpha_k) \in B)\right| < \frac{c}{n}.$$

Доказательство. Из (24) следует, что

$$P\left(s_n(1) = k, \left(\frac{\gamma_1}{n}, \dots, \frac{\gamma_k}{n}\right) \in B\right) \leq P(\xi(1) = k, (\alpha_1, \dots, \alpha_k) \in B^{(1/n)}) + \frac{\mu_n^2}{n} \leq \\ \leq P(\xi(1) = k, (\alpha_1, \dots, \alpha_k) \in B) + \frac{c}{n} + \frac{\mu_n^2}{n}.$$

Точно так же устанавливается обратное неравенство. ■

Доказательство теоремы 5. Пусть $\eta_{k,n} = \xi\left(\frac{k}{n}\right) - \xi\left(\frac{k-1}{n}\right)$, $k = 1, \dots, n$. Мы докажем утверждение теоремы, если осуществим построение $\xi_{k,n}$, $\eta_{k,n}$ на одном пространстве так, что

$$P\left(\bigcup_{k=1}^n \{\xi_{k,n} \neq \eta_{k,n}\}\right) < \frac{\mu_n^2}{n} \quad (25)$$

(таким построением $\xi(t)$ будет определен не на всех интервалах $((k-1)/n, k/n)$. Однако дальнейшая «достройка» $\xi(t)$ нас интересовать не будет — она может быть любой). Конструкция, приводящая к (25), по существу уже была использована в теореме 5.9. Мы получим нужное построение, если рассмотрим независимые $\omega_1, \dots, \omega_n$, $\omega_k \in U_{0,1}$ и положим

$$\xi_{k,n} = \begin{cases} 0, & \text{если } \omega_k < 1 - p_n, \\ 1, & \text{если } \omega_k \geq 1 - p_n, \end{cases} \quad \eta_{k,n} = \begin{cases} 0, & \text{если } \omega_k < e^{-p_n}, \\ k \geq 1, & \text{если } \omega_k \in [\pi_{k-1}, \pi_k), \end{cases}$$

где $\pi_k = \Pi_\mu([0, k])$, $k = 0, 1, \dots$. Тогда $\xi_{k,n}$ и $\eta_{k,n}$ будут иметь требуемые совместные распределения,

$$\{\xi_{k,n} \neq \eta_{k,n}\} = \{\omega_k \in [1 - p_n, e^{-p_n}) \cup [e^{-p_n} + p_n e^{-p_n}, 1]\}.$$

Так как $P(\xi_{k,n} \neq \eta_{k,n}) \leq p_n^2$, то теорема доказана. ■

Пуассоновский процесс может выступать в роли предельного и в нескольких качестве — как предельный процесс для суммы большого числа однородных «медленных» процессов восстановления.

Сформулируем постановку задачи более точно. Пусть $\eta_i(t)$, $i = 1, 2, \dots, n$, — независимые между собой *произвольные однородные процессы восстановления* в «схеме серий» (т.е. зависящие от n), для которых (см. гл. 9)

$$\begin{aligned} E\eta_i(t) &= \frac{t}{a_i}, \quad a_i = a_{i,n} = E\xi_2^{(i)} \rightarrow \infty, \\ \sum_{i=1}^n \frac{1}{a_i} &\rightarrow \lambda, \quad \text{и при фиксированном } t, \\ F_i(t) &= P(\xi_2^{(i)} < t) \leq r_{t,n} \rightarrow 0 \end{aligned}$$

при $n \rightarrow \infty$, где $r_{t,n}$ от i не зависит, $\xi_2^{(i)}, \xi_3^{(i)}, \dots$ — случайные величины, порождающие процесс $\eta_i(t)$.

Теорема 6. При выполнении сформулированных условий конечномерные распределения процесса

$$\zeta_n(t) = \sum_{i=1}^n \eta_i(t)$$

сходятся к распределению пуассоновского процесса $\zeta(t)$ с параметром λ :

$$P(\zeta_n(t_1) = k_1, \dots, \zeta_n(t_l) = k_l) \rightarrow P(\zeta(t_1) = k_1, \dots, \zeta(t_l) = k_l).$$

(О сходимости к пуассоновскому процессу см. замечание, предшествующее теореме 5.)

Доказательство. Докажем сначала сходимость распределений приращений $\zeta_n(t+u) - \zeta_n(u)$ к распределению Пуассона с параметром λt . Обозначим $\Delta_i = \eta_i(t+u) - \eta_i(u)$, $p_i = t/a_i$. Имеем ($\chi_i(u)$ — эксцесс для процесса η_i ; см. гл. 9)

$$\begin{aligned} E\Delta_i &= p_i, \\ P(\Delta_i \geq l) &\leq P(\chi_i(u) < t) [P(\xi_2^{(i)} < t)]^{l-1} \leq \\ &\leq \frac{1}{a_i} \int_0^1 P(\xi_2^{(i)} > z) dz \cdot F_i(t)^{l-1} \leq \frac{t}{a_i} (r_{t,n})^{l-1} = p_i r_{t,n}^{l-1}. \end{aligned}$$

Отсюда следует, что

$$\begin{aligned} E\Delta_i &= p_i = \sum_{l=1}^{\infty} l P(\Delta_i = l) = P(\Delta_i = 1) + o(p_i), \\ P(\Delta_i = 1) &= p_i + o(p_i), \quad P(\Delta_i = 0) = 1 - p_i + o(p_i). \end{aligned} \quad (26)$$

Мы находимся в условиях следствия 5.5, из которого следует, что

$$\zeta_n(t+u) - \zeta_n(u) = \sum_{i=1}^n \Delta_i \in \Pi_{\lambda t}. \quad (27)$$

Осталось доказать асимптотическую независимость приращений. Рассмотрим для простоты лишь два приращения на интервалах $(0, u)$ и $(u, u+t)$ и предположим, что $\zeta_n(u) = k$. Более того, предположим, что произошло событие A , состоящее в том, что восстановления произошли в процессах с номерами i_1, \dots, i_k . Нам достаточно убедиться, что при этом условии по-прежнему будет справедливо (27). Пусть B есть событие, состоящее в том, что в процессах η_i с номерами i_1, \dots, i_k на интервале $(u, u+t)$ вновь произошли восстановления. Очевидно, что

$$P(B|A) \leq \sum_{i=1}^k P(\xi_2^{(i)} < t+u) \leq k r_{t+u, n} \rightarrow 0.$$

Таким образом, вклад процессов η_i , $i = 1, \dots, k$, в сумму (27) при условии A пренебрежимо мал. Рассмотрим оставшиеся $n - k$ процессов. Для них

$$\begin{aligned} P(\Delta_i \geq 1|A) &= \frac{P(\chi_i(0) \in (u, u+t))}{P(\chi_i(0) > u)} = \\ &= \frac{1}{a_i} \int_u^{u+t} (1 - F_i(z)) dz \left[1 - \frac{1}{a_i} \int_0^u (1 - F_i(z)) dz \right]^{-1} = p_i + o(p_i). \end{aligned} \quad (28)$$

Так как соотношение (26) для условных (при условии A и при $i \neq i_l$, $l = 1, \dots, k$) распределений Δ_i сохраняется, то получим аналогично предыдущему (роль равенства $\sum_{l=1}^{\infty} l P(\Delta_i = l) = p_i$ теперь играет равенство $\sum_{l=1}^{\infty} l P(\Delta_i = l|A) = p_i + o(p_i)$, вытекающее из (28)):

$$P(\Delta_i = 1|A) = p_i + o(p_i), \quad P(\Delta_i = 0|A) = 1 - p_i + o(p_i).$$

Остается вновь воспользоваться следствием 5.5. ■

Марковские процессы

§ 1. Определения и общие свойства

Марковские процессы с дискретным временем и дискретным множеством состояний (цепи Маркова) рассматривались в гл. 12. Напомним, что главным в определении таких процессов было свойство независимости «будущего» процесса от его «прошлого», если фиксировано его «настоящее». Этот же принцип лежит в основе определения марковского процесса и в общем случае.

1. Определения и общие свойства. Пусть $\langle \Omega, \mathfrak{F}, P \rangle$ — вероятностное пространство, на котором задан случайный процесс $\{\xi(t) = \xi(t, \omega), t \geq 0\}$. Обозначим

$$\mathfrak{F}_t = \sigma(\xi(u); u \leq t), \quad \mathfrak{F}_{[t, \infty)} = \sigma(\xi(u); u \geq t),$$

так что величина $\xi(u)$ измерима относительно \mathfrak{F}_t при $u \leq t$ и относительно $\mathfrak{F}_{[t, \infty)}$ при $u \geq t$. σ -алгебра $\sigma(\mathfrak{F}_t, \mathfrak{F}_{[t, \infty)})$ порождена значениями $\xi(u)$ при всех u и для выборочного вероятностного пространства может совпадать с \mathfrak{F} .

Определение 1. Случайный процесс $\xi(t)$ называется *марковским*, если при любых $t, A \in \mathfrak{F}_t, B \in \mathfrak{F}_{[t, \infty)}$

$$P(AB|\xi(t)) = P(A|\xi(t))P(B|\xi(t)). \quad (1)$$

Это точно выраженная независимость будущего от прошлого при фиксированном настоящем (условная независимость \mathfrak{F}_t и $\mathfrak{F}_{[t, \infty)}$ при фиксированном $\xi(t)$).

Покажем, что приведенное определение эквивалентно следующему:

Определение 2. Процесс $\xi(t)$ называется *марковским*, если для любой ограниченной случайной величины η , измеримой относительно $\mathfrak{F}_{[t, \infty)}$,

$$E(\eta|\mathfrak{F}_t) = E(\eta|\xi(t)). \quad (2)$$

В качестве η достаточно брать функции вида $\eta = f(\xi(s))$ при $s \geq t$.

Доказательство эквивалентности. Пусть верно (1). В силу теоремы о монотонной сходимости достаточно доказать (2) для простых функций η . Для этого в свою очередь достаточно доказать (2) для $\eta = I_B$ — индикатора множества $B \in \mathfrak{F}_{[t, \infty)}$. Пусть $A \in \mathfrak{F}_t$. Тогда в силу (1)

$$\begin{aligned} P(AB) &= EP(AB|\xi(t)) = EP(A|\xi(t))P(B|\xi(t)) = \\ &= EE[I_A P(B|\xi(t))|\xi(t)] = EI_A P(B|\xi(t)). \end{aligned} \quad (3)$$

С другой стороны,

$$P(AB) = EI_A I_B = EI_A P(B|\mathfrak{F}_t). \quad (4)$$

Так как (3), (4) верны для любого $A \in \mathfrak{F}_t$, то это означает, что $P(B|\mathfrak{F}_t) = P(B|\xi(t))$.

Наоборот, пусть теперь верно (2). Тогда для $A \in \mathfrak{F}_t$, $B \in \mathfrak{F}_{[t, \infty)}$

$$\begin{aligned} P(AB|\xi(t)) &= E[E(I_A I_B | \mathfrak{F}_t) | \xi(t)] = E[I_A E(I_B | \mathfrak{F}_t) | \xi(t)] = \\ &= E[I_A E(I_B | \xi(t)) | \xi(t)] = P(B|\xi(t))P(A|\xi(t)). \end{aligned}$$

Приведем еще одно эквивалентное определение марковского процесса.

Определение 3. Процесс $\xi(t)$ является *марковским*, если для любой ограниченной функции f и для любых $t_1 < t_2 < \dots < t_n \leq t$

$$E(f(\xi(t)) | \xi(t_1), \dots, \xi(t_n)) = E(f(\xi(t)) | \xi(t_n)). \quad (5)$$

Доказательство эквивалентности. Соотношение (5) очевидным образом следует из (2). Пусть теперь выполнено (5). Тогда для любого $A \in \sigma(\xi(t_1), \dots, \xi(t_n))$

$$E(f(\xi(t)); A) = E[E(f(\xi(t)) | \xi(t_n)); A]. \quad (6)$$

Обе части в (6) являются мерами, совпадающими на алгебре цилиндрических множеств. Следовательно, по теореме единственности продолжения меры они совпадают и на σ -алгебре, порожденной этими множествами, т.е. на \mathfrak{F}_{t_n} . Другими словами, (6) верно при любом $A \in \mathfrak{F}_{t_n}$, что эквивалентно равенству

$$E[f(\xi(t)) | \mathfrak{F}_{t_n}] = E[f(\xi(t)) | \xi(t_n)]$$

при любом $t_n \leq t$. Соотношение (2) для $\eta = f(\xi(t))$ доказано.

Осталось убедиться, что в (2) достаточно брать $\eta = f(\xi(s))$, $s \geq t$. Пусть $t \leq u_1 < \dots < u_n$. Докажем, что тогда (2) верно для

$$\eta = \prod_{i=1}^n f_i(\xi(u_i)). \quad (7)$$

Воспользуемся индукцией и предположим, что (2) выполнено для функций $\gamma = \prod_{i=1}^{n-1} f_i(\xi(u_i))$ (при $n=1$ (2) выполнено). Тогда, обозначив $g(u_{n-1}) = E[f_n(\xi(u_n)) | \xi(u_{n-1})]$, получим

$$\begin{aligned} E(\eta | \mathfrak{F}_t) &= E[E(\eta | \mathfrak{F}_{u_{n-1}}) | \mathfrak{F}_t] = E[\gamma E(f_n(\xi(u_n)) | \mathfrak{F}_{u_{n-1}}) | \mathfrak{F}_t] = \\ &= E[\gamma E(f_n(\xi(u_n)) | \xi(u_{n-1})) | \mathfrak{F}_t] = E[\gamma g(\xi(u_{n-1})) | \mathfrak{F}_t]. \end{aligned}$$

По предположению индукции отсюда следует, что

$$E(\eta | \mathfrak{F}_t) = E[\gamma g(\xi(u_{n-1})) | \xi(t)]$$

и, значит, $E(\eta | \mathfrak{F}_t)$ измеримо относительно $\sigma(\xi(t))$,

$$E(\eta | \xi(t)) = E(E(\eta | \mathfrak{F}_t) | \xi(t)) = E(\eta | \mathfrak{F}_t).$$

Мы доказали выполнение (2) для функций (7), измеримых относительно $\sigma(\xi(u_1), \dots, \xi(u_n))$. Путем предельного перехода справедливость (2) устанавливается для простых, а затем и для любых $\mathfrak{F}_{[t, \infty)}$ -измеримых функций. ■

2. Переходная вероятность. Мы выяснили, что для марковского процесса $\xi(t)$ условная вероятность

$$P(\xi(t) \in B | \mathfrak{F}_s) = P(\xi(t) \in B | \xi(s)) \quad \text{при } t > s$$

есть борелевская функция от $\xi(s)$, которую обозначим

$$P(\xi(t) \in B | \xi(s)) = P(s, \xi(s); t, B).$$

Можно считать, что $P(s, x; t, B)$ как функция от B и x есть условное распределение (см. § 4.9) $\xi(t)$ при условии, что $\xi(s) = x$. В силу свойства марковости они удовлетворяют соотношению ($s < u < t$)

$$P(s, x; t, B) = \int P(s, x; u, dy) P(u, y; t, B), \quad (8)$$

которое вытекает из равенства

$$P(\xi(t) \in B | \xi(s) = x) = E[P(\xi(t) \in B | \mathfrak{F}_u) | \xi(s) = x] = E[P(u, \xi(u); t, B) | \xi(s) = x].$$

Уравнение (8) носит название *уравнения Колмогорова—Чепмэна*.

Функцию $P(s, x; t, B)$ можно использовать для аналитического задания марковского процесса. Предварительно надо выяснить, какими свойствами должна обладать функция $P_{x,B}(s, t)$ для того, чтобы нашелся марковский процесс $\xi(t)$, для которого

$$P_{x,B}(s, t) = P(s, x; t, B).$$

Пусть $\langle \mathcal{X}, \mathfrak{B}_{\mathcal{X}} \rangle$ — некоторое измеримое пространство.

Определение 4. Функция $P_{x,B}(s, t)$ называется *переходной функцией* на $\langle \mathcal{X}, \mathfrak{B}_{\mathcal{X}} \rangle$, если она удовлетворяет следующим условиям:

- 1) $P_{x,B}(s, t)$ как функция от B есть распределение вероятностей при каждом $s \leq t$, $x \in \mathcal{X}$.
- 2) $P_{x,B}(s, t)$ измерима по x при каждом $s \leq t$, $B \in \mathfrak{B}_{\mathcal{X}}$.
- 3) При $0 \leq s < u < t$ и при всех x, B

$$P_{x,B}(s, t) = \int P_{x,dy}(s, u) P_{y,B}(u, t)$$

(уравнение Колмогорова—Чепмэна).

- 4) $P_{x,B}(s, t) = I_k(B)$ при $s = t$.

Здесь свойства 1), 2) обеспечивают функции $P_{x,B}(s, t)$ возможность быть условным распределением (ср. с § 4.9).

Определим теперь с помощью $P_{x,B}(s, t)$ конечномерные распределения некоторого процесса $\xi(t)$ с начальным условием $\xi(0) = a$ по формуле

$$P(\xi(t_1) \in dy_1, \dots, \xi(t_n) \in dy_n) = P_{a,dy_1}(0, t_1) P_{y_1,dy_2}(t_1, t_2) \dots P_{y_{n-1},dy_n}(t_{n-1}, t_n). \quad (9)$$

В силу свойств 3), 4) эти распределения согласованы и, значит, по теореме Колмогорова определяют $\xi(t)$ в $\langle \mathbb{R}^T, \mathfrak{B}_{\mathbb{R}}^T \rangle$.

По формуле (9) и правилу (6)

$$\begin{aligned} P(\xi(t_n) \in B_n | (\xi(t_1), \dots, \xi(t_{n-1})) = (y_1, \dots, y_{n-1})) = \\ = P_{y_{n-1}, B_n}(t_{n-1}, t_n) = P(\xi(t_n) \in B_n | \xi(t_{n-1}) = y_{n-1}) = P(t_{n-1}, y_{n-1}; t_n, B_n). \end{aligned}$$

Можно проверить это равенство и более формально, пользуясь тем, что интегралы от обеих его частей по множеству $\{\xi(t_1) \in B_1, \dots, \xi(t_{n-1}) \in B_{n-1}\}$ совпадают.

Итак, в силу определения 3 мы построили марковский процесс $\xi(t)$, для которого

$$P(s, x; t, B) = P_{x, B}(s, t).$$

Эту функцию мы будем называть также *переходной функцией* (или *вероятностью*) *процесса* $\xi(t)$.

Определение 5. Марковский процесс $\xi(t)$ называется *однородным*, если $P(s, x; t, B)$ как функция s и t зависит лишь от разности $t - s$:

$$P(s, x; t, B) = P(t - s; x, B).$$

Это есть вероятность за время $t - s$ попасть из x в B . Если $P(u; x, B) = \int_B p(u; x, y) dy$, то функция $p(u; x, y)$ называется *переходной плотностью*.

Нетрудно видеть, что винеровский и пуассоновский процессы являются марковскими однородными процессами. Например, для винеровского процесса

$$p(u; x, y) = \frac{1}{\sqrt{2\pi u}} e^{-(x-y)^2/(2u)}.$$

§ 2. Марковские процессы со счетным множеством состояний. Примеры

1. Основные свойства процесса. Не ограничивая общности, будем считать, что «дискретное» фазовое пространство \mathcal{X} совпадает с множеством целых чисел $\{0, 1, 2, \dots\}$. Для простоты будем рассматривать лишь однородные марковские процессы.

Переходная функция такого процесса определяется набором функций $P(t; i, j) = p_{ij}(t)$, образующих стохастическую матрицу $P(t) = \|p_{ij}(t)\|$ ($p_{ij}(t) > 0$, $\sum_j p_{ij}(t) = 1$). Уравнение Колмогорова—Чепмэна здесь принимает вид

$$p_{ij}(t+s) = \sum_k p_{ik}(t) p_{kj}(s),$$

или, что то же, в матричной форме

$$P(t+s) = P(t)P(s) = P(s)P(t). \quad (10)$$

Будем рассматривать в дальнейшем лишь *стохастически непрерывные* процессы, для которых $\xi(t+s) \xrightarrow{p} \xi(t)$ при $s \rightarrow 0$, что в рассматриваемых условиях эквивалентно тому, что

$$\begin{aligned} P(\xi(t+s) \neq \xi(t)) &\rightarrow 0, \\ P(t+s) &\rightarrow P(t), \\ P(s) &\rightarrow P(0) \equiv E \end{aligned} \quad (11)$$

при $s \rightarrow 0$ (поэлементно). Мы будем предполагать также, что сходимость в (11) *равномерна* (для конечного \mathcal{X} это всегда так).

В соответствии с требованием сепарабельности будем считать, что $\xi(t)$ не может менять свое состояние за «нулевое» время более одного раза (исключаем эффекты, проиллюстрированные в примере 17.1, т. е. будем считать, что если $\xi(t) = j$, то с вероятностью 1 $\xi(t+s) = j$ при $s \in [0, \tau)$, $\tau = \tau(\omega) > 0$). В этом случае траектории процесса будут кусочно-постоянными (для определенности, непрерывными

справа), т. е. ось времени разбивается на полуинтервалы $[0, \tau_1), [\tau_1, \tau_1 + \tau_2), \dots$, на которых $\xi(t)$ постоянно.

Обозначим

$$q_j(t) = P(\xi(u) = j, 0 \leq u < t | \xi(0) = j) = P(\tau_1 \geq t).$$

Теорема 1. При выполнении сделанных предположений (стохастическая непрерывность, сепарабельность)

$$q_i(t) = e^{-q_i t},$$

где $q_i < \infty$; $q_i > 0$, если $p_{ii}(t) \neq 1$. Существуют

$$\lim_{t \rightarrow 0} \frac{1 - p_{ii}(t)}{t} = q_i, \quad \lim_{t \rightarrow 0} \frac{p_{ij}(t)}{t} = q_{ij}, \quad i \neq j, \quad (12)$$

где $\sum_{j \neq i} q_{ij} = q_i$.

Доказательство. В силу свойства марковости

$$q_i(t+s) = q_i(t)q_i(s),$$

при этом $q_i(t) \downarrow$. Поэтому существует единственное решение этого уравнения $q_i(t) = e^{-q_i t}$, где $q_i < \infty$, т. к. $P(\tau_1 > 0) = 1$ и $q_i > 0$, т. к. $q_i(t) < 1$ при $p_{ii}(t) \neq 1$.

Пусть далее $0 < t_0 < t_1 < \dots < t_n < t$. Так как события

$$\{\xi(u) = i \text{ при } u \leq t_r, \quad \xi(t_{r+1}) = j\} \quad r = 0, \dots, n-1; \quad j \neq i,$$

несовместны, то

$$p_{ii}(t) = q_i(t) + \sum_{r=0}^{n-1} \sum_{j \neq i} q_i(t_r) p_{ij}(t_{r+1} - t_r) p_{ji}(t - t_{r+1}). \quad (13)$$

Здесь по условию (11) $p_{ji}(t - t_{r+1}) < \varepsilon_t$ при всех $j \neq i$, $\varepsilon_t \rightarrow 0$ при $t \rightarrow 0$, так что сумма в (13) не превосходит

$$\varepsilon_t \sum_{r=0}^{n-1} \sum_{j \neq i} q_i(t_r) p_{ij}(t_{r+1} - t_r) = \varepsilon_t P\left(\bigcup_{r=1}^n \{\xi(t_r) \neq i\} | \xi(0) = i\right) < \varepsilon_t (1 - q_i(t)),$$

$$p_{ii}(t) \leq q_i(t) + \varepsilon_t (1 - q_i(t)).$$

Вместе с очевидным неравенством $p_{ii}(t) \geq q_i(t)$ это дает

$$1 - q_i(t) \geq 1 - p_{ii}(t) \geq (1 - q_i(t))(1 + \varepsilon_t)$$

(т. е. асимптотически $1 - q_i(t)$ и $1 - p_{ii}(t)$ при $t \rightarrow 0$ ведут себя одинаково). Отсюда следует второе утверждение теоремы.

Пусть теперь $t_r = rt/n$. Рассмотрим переходные вероятности

$$\begin{aligned} p_{ij}(t) &\geq \sum_{r=0}^{n-1} q_i(t_r) p_{ij}\left(\frac{t}{n}\right) q_i(t - t_{r+1}) \geq \\ &\geq (1 - \varepsilon_t) p_{ij}\left(\frac{t}{n}\right) \sum_{r=0}^{n-1} e^{-q_i r t/n} \geq (1 - \varepsilon_t) p_{ij}\left(\frac{t}{n}\right) \frac{1 - e^{-q_i t}}{q_i t/n}. \end{aligned}$$

Отсюда следует, что

$$p_{ij}(t) \geq (1 - \varepsilon_t) \left(\frac{1 - e^{-q_i t}}{q_i} \right) \lim_{\delta \rightarrow 0} \frac{p_{ij}(\delta)}{\delta}$$

и что верхний предел, стоящий справа, ограничен. Переходя к пределу при $t \rightarrow 0$, получим

$$\lim_{t \rightarrow 0} \frac{p_{ij}(t)}{t} \geq \lim_{\delta \rightarrow 0} \frac{p_{ij}(\delta)}{\delta}.$$

Так как $\sum_{j \neq i} p_{ij}(t) = 1 - p_{ii}(t)$, то $\sum_{j \neq i} q_{ij} = q_i$. ■

Теорема показывает, что числа

$$p_{ij} = \frac{q_{ij}}{q_i}, \quad j \neq i, \quad p_{ii} = 0$$

образуют стохастическую матрицу и имеют смысл вероятностей перехода из i в j за бесконечно малый интервал Δ при условии, что процесс $\xi(t)$ на этом интервале покинул состояние i :

$$P(\xi(t + \Delta) = j | \xi(t) = i, \xi(t + \Delta) \neq i) = \frac{p_{ij}(\Delta)}{1 - p_{ii}(\Delta)} \rightarrow \frac{q_{ij}}{q_i}.$$

Таким образом, эволюцию $\xi(t)$ можно представлять себе следующим образом. Если $\xi(0) = X_0$, то $\xi(t)$ находится в состоянии X_0 время $\tau_1 \in \Gamma_{q_{X_0}}$. Затем $\xi(t)$ переходит в состояние X_1 с вероятностью $p_{X_0 X_1}$. Далее $\xi(t) = X_1$ на интервале времени $[\tau_1, \tau_1 + \tau_2)$, $\tau_2 \in \Gamma_{q_{X_1}}$, после чего система меняет свое состояние на X_2 , и т. д. Ясно, что последовательность X_0, X_1, \dots связана в однородную цепь Маркова с матрицей переходных вероятностей $\|p_{ij}\|$. Поэтому дальнейшее изучение $\xi(t)$ во многом может сведено к изучению цепи Маркова $\{X_n; n \geq 0\}$, проведенному достаточно подробно в гл. 12.

Мы видим, что эволюция $\xi(t)$ целиком описывается с помощью чисел q_{ij} , q_i , образующих матрицу

$$Q = \|q_{ij}\| = \lim_{t \rightarrow 0} \frac{P(t) - P(0)}{t}, \quad (14)$$

где принято $q_{ii} = -q_i$, так что $\sum_j q_{ij} = 0$. К этому факту можно подойти и аналитически. Чтобы упростить техническую сторону рассуждений, будем предполагать, где это потребуется, что элементы матрицы Q ограничены, а сходимость в (12) происходит равномерно по i .

Обозначим через e^A матричнозначную функцию

$$e^A = E + \sum_{k=1}^{\infty} \frac{A^k}{k!}.$$

Теорема 2. *Переходные вероятности $p_{ij}(t)$ удовлетворяют системе дифференциальных уравнений*

$$P'(t) = P(t)Q, \quad (15)$$

$$P'(t) = QP(t). \quad (16)$$

Каждая из систем (15), (16) имеет единственное решение

$$P(t) = e^{Qt}.$$

Ясно, что решение получается немедленно, если формально проинтегрировать (15).

Доказательство. В силу (10), (11), (14)

$$P'(t) = \lim_{s \rightarrow 0} \frac{1}{s} (P(t+s) - P(t)) = \lim_{s \rightarrow 0} \frac{1}{s} P(t) (P(s) - E) = P(t)Q. \quad (17)$$

Точно так же из равенства

$$P(t+s) - P(t) = (P(s) - E)P(t)$$

получается второе уравнение (16). Законность предельных переходов обеспечивается сделанными предположениями.

Далее, из (15) следует, что функция $P(t)$ является бесконечно дифференцируемой, при этом

$$\begin{aligned} P^{(k)}(t) &= P(t)Q^k, \\ P(t) - P(0) &= \sum_{k=1}^{\infty} P^{(k)}(0) \frac{t^k}{k!} = \sum_{k=1}^{\infty} \frac{Q^k t^k}{k!}, \\ P(t) &= P(0)e^{Qt}. \end{aligned}$$

В связи с методом получения уравнение (15) называется *обратным уравнением Колмогорова*, а уравнение (16) — *прямым* (приращение времени берется *после* или *до* основного интервала времени).

Разница между этими уравнениями становится еще более наглядной в случае неоднородных марковских процессов, когда переходные вероятности $P(\xi(t) = j | \xi(s) = i) = p_{ij}(s, t)$, $s \leq t$, зависят от двух временных аргументов: s и t . В этом случае (10) превращается в равенство $P(s, t+u) = P(s, t)P(t, t+u)$, обратное уравнение имеет вид $\partial P(s, t)/\partial s = P(s, t)Q(s)$, прямое — $\partial P(s, t)/\partial t = Q(t)P(s, t)$. Читатель может получить эти соотношения самостоятельно.

Каковы общие условия существования стационарного предельного распределения? Можно воспользоваться здесь подходом, аналогичным тому, который использовался в гл. 12.

Пусть $\xi^{(i)}(t)$ есть процесс с начальным значением $\xi^{(i)}(0) = i$ и с непрерывными справа траекториями. Для заданного i_0 обозначим

$$\begin{aligned} \nu^{(i)} &= \min\{t \geq 0: \xi^{(i)}(t) = i_0\} = \nu_0, \\ \nu_k &= \min\{t \geq \nu_{k-1} + 1: \xi^{(i)}(t) = i_0\}, \quad k = 1, 2, \dots \end{aligned}$$

Здесь во второй формуле мы рассматриваем значения $t \geq \nu_{k-1} + 1$, так как при $t \geq \nu_{k-1}$ было бы $\nu_k \equiv \nu_{k-1}$. Очевидно, что $P(\nu_k - \nu_{k-1} = 1) > 0$, $P(\nu_k - \nu_{k-1} \in (t, t+h)) > 0$ при любых $t > 1$, $h > 0$, если $p_{i_0 i_0}(t) \neq 1$.

Заметим также, что величины ν_k , $k = 0, 1, \dots$, определены не для всех элементарных исходов. Мы положим $\nu_0 = \infty$, если $\xi^{(i)}(t) \neq i_0$ при всех $t \geq 0$. Аналогичное соглашение действует и для ν_k при $k \geq 1$.

Теорема 3. Пусть существует состояние i_0 такое, что $E\nu_1 < \infty$, $P(\nu^{(i)} < \infty) = 1$ при любом $i \in \mathcal{X}_0 \subset \mathcal{X}$. Тогда существуют

$$\lim_{t \rightarrow \infty} p_{ij}(t) = p_j, \quad (18)$$

не зависящие от $i \in \mathcal{X}_0$.

Доказательство. Так же, как и для цепей Маркова, моменты времени ν_1, ν_2, \dots разбивают ось времени на одинаково устроенные независимые циклы по возвращению (спустя время 1) в состояние i_0 . Рассмотрим процесс восстановления, порожденный суммами ν_k , $k = 0, 1, \dots$, независимых случайных величин $\nu_0, \nu_k - \nu_{k-1}$, $k = 1, 2, \dots$. Пусть

$$\eta(t) = \min\{k: \nu_k > t\}, \quad \gamma(t) = t - \nu_{\eta(t)-1}, \quad H(t) = \sum_{k=0}^{\infty} P(\nu_k < t).$$

Событие $A_{dv} = \{\gamma(t) \in [v, v + dv)\}$ представимо в виде произведения событий $B_{dv} = \bigcup_{k \geq 0} \{\nu_k \in (t - v - dv, t - v]\} \in \mathfrak{F}_{t-v}$ и $C_v = \{\xi(u) \neq i_0 \text{ при } u \in [t - v + 1, t]\} \in \mathfrak{F}_{[t-v, \infty)}$. Имеем

$$\begin{aligned} p_{ij}(t) &= \int_0^t P(\xi^{(i)}(t) = j, \gamma(t) \in [v, v + dv)) = \\ &= \int_0^t P(\xi^{(i)}(t) = j, B_{dv} C_v) = \int_0^t E I_{B_{dv}} P(\xi^{(i)}(t) = j, C_v | \mathfrak{F}_{t-v}) = \\ &= \int_0^t E I_{B_{dv}} P(\xi^{(i)}(t) = j, C_v | \xi(t - v)). \end{aligned}$$

На множестве B_{dv} выполняется $\xi(t - v) = i_0$ и, следовательно, вероятность под интегралом равна

$$P(\xi^{(i_0)}(v) = j, \xi(u) \neq i_0 \text{ при } u \in [1, v]) = g(v)$$

и не зависит от t и i . Так как $P B_{dv} = dH(t - v)$, то

$$p_{ij}(t) = \int_0^t g(v) P(B_{dv}) = \int_0^t g(v) dH(t - v).$$

По основной теореме восстановления этот интеграл сходится к $1/(E\nu_1) \int_0^{\infty} g(v) dv$. Существование последнего интеграла вытекает из того, что $g(v) \leq P(\nu_1 > v)$. ■

Теорема 4. Если существует стационарное распределение

$$P = \lim_{t \rightarrow \infty} P(t),$$

где все строки матрицы P одинаковы, то оно является единственным решением уравнения

$$PQ = 0. \quad (19)$$

Очевидно, что уравнение (19) получается приравниванием $P'(t) = 0$ в (15). Уравнение (16) дает тривиальное равенство $QP = 0$.

Доказательство. Уравнение (19) получится, если в (17) перейти к пределу сначала при $t \rightarrow \infty$, а затем при $s \rightarrow 0$. Допустим теперь, что P_1 есть решение (19), т. е.

$P_1 Q = 0$. Тогда $P_1 P(t) = P_1$ при $t < 1$, поскольку $P_1(P(t) - P(0)) = P_1 \sum_{k=1}^{\infty} \frac{Q^k t^k}{k!} = 0$.

Далее, $P_1 = P_1 P^k(t) = P_1 P(kt)$, $P(kt) \rightarrow P$ при $k \rightarrow \infty$ и, значит, $P_1 = P_1 P = P$. ■

Рассмотрим цепь Маркова $\{X_n\}$ с дискретным временем и переходными вероятностями $p_{ij} = q_{ij}/q_i$, $i \geq j$, $p_{ii} = 0$. Допустим, что эта цепь эргодична (см. теорему 12.8). Тогда ее стационарные вероятности $\{\pi_j\}$ удовлетворяют уравнениям (12.8). Заметим теперь, что уравнения (19) можно записать в виде

$$p_j q_j = \sum_k p_k q_k p_{kj},$$

которое имеет очевидное решение $p_j = c \pi_j / q_j$, $c = \text{const}$. Стало быть, если

$$\sum \frac{\pi_j}{q_j} < \infty, \quad (20)$$

то существует решение (19), равное

$$p_j = \frac{\pi_j}{q_j} \left(\sum \frac{\pi_j}{q_j} \right)^{-1}. \quad (21)$$

В § 4 мы получим эргодическую теорему для процессов более общего вида, чем те, которые рассматривались в этом параграфе. Из нее будет следовать, в частности, что эргодичность $\{X_n\}$ и сходимости (20) влекут за собой эргодичность (18). Напомним, что для эргодичности $\{X_n\}$ в свою очередь достаточно существования решения $\{\pi_j\}$ уравнений (12.8). Таким образом, существование решения (21) влечет за собой эргодичность $\xi(t)$.

2. Примеры.

Пример 1. Пуассоновский процесс $\xi(t)$ с параметром λ является марковским процессом, для которого $q_i = \lambda$, $q_{i,i+1} = \lambda$, $p_{i,i+1} = 1$, $i = 0, 1, \dots$. Для него стационарное распределение p не существует (траектория уходит на бесконечность).

Пример 2. Процессы рождения и гибели. Так называются процессы, для которых при $i \geq 1$

$$p_{ij}(\Delta) = \begin{cases} \lambda_i \Delta + o(\Delta) & \text{при } j = i + 1, \\ \gamma_i \Delta + o(\Delta) & \text{при } j = i - 1, \\ o(\Delta) & \text{при } |j - i| \geq 2, \end{cases}$$

так что

$$p_{ij} = \begin{cases} \frac{\lambda_i}{\lambda_i + \gamma_i} & \text{при } j = i + 1, \\ \frac{\gamma_i}{\lambda_i + \gamma_i} & \text{при } j = i - 1 \end{cases}$$

есть вероятности, соответственно, рождения и гибели одной частицы в некоторой популяции, когда эта популяция содержала i частиц и изменила свое состояние. При $i = 0$ следует положить $\gamma_0 = 0$. Выяснение условий существования стационарного режима представляет собой весьма сложную задачу (связанную главным образом с выяснением условий ухода траектории на бесконечность). Если таковой существует, то согласно теореме 4 стационарные вероятности p_j могут быть найдены

единственным образом из рекуррентных соотношений (см. уравнение (19), здесь $q_{ii} = -q_i = -(\lambda_i + \gamma_i)$)

$$\begin{aligned} -p_0\lambda_0 + p_1\gamma_1 &= 0, \\ p_0\lambda_0 - p_1(\lambda_1 + \gamma_1) + p_2\gamma_2 &= 0, \\ \dots\dots\dots \\ p_{k-1}\lambda_{k-1} - p_k(\lambda_k + \gamma_k) + p_{k+1}\gamma_{k+1} &= 0, \\ \dots\dots\dots \end{aligned} \quad (22)$$

и условия $\sum p_j = 1$.

Пример 3. Задача о телефонных линиях в теории массового обслуживания. Пусть дана система, состоящая из бесконечного числа каналов связи, по которым ведутся телефонные разговоры. Вероятность окончания разговора в каждой из линий за малый интервал времени $(t, t + \Delta)$ равна $\lambda\Delta + o(\Delta)$. Вероятность поступления нового вызова за этот интервал времени равна $\gamma\Delta + o(\Delta)$. Таким образом, «входящий» поток вызовов есть не что иное, как пуассоновский процесс с параметром λ , а число $\xi(t)$ занятых каналов в момент времени t есть процесс рождения и гибели, для которого $\lambda_i = \lambda$, $\gamma_i = i\gamma$.

В этом случае с помощью теоремы 3 нетрудно убедиться, что всегда существует стационарное предельное распределение (при $i_0 = 0$), уравнения (22), для которого примут вид

$$\begin{aligned} \gamma p_0 &= \lambda p_1, \\ \dots\dots\dots \\ (\lambda + \gamma k)p_k &= \lambda p_{k-1} + (k+1)\gamma p_{k+1}, \\ \dots\dots\dots \end{aligned} \quad (23)$$

Отсюда получаем

$$p_1 = p_0 \frac{\lambda}{\gamma}, \quad p_2 = \frac{p_0}{2} \left(\frac{\lambda}{\gamma} \right)^2, \quad \dots, \quad p_k = \left(\frac{\lambda}{\gamma} \right)^k \frac{p_0}{k!}, \quad (24)$$

так что $p_0 = e^{-\lambda/\gamma}$, а предельное распределение будет распределением Пуассона с параметром λ/γ .

Если число каналов n конечно, то вызовы, заставшие все каналы занятыми, получают отказ и надо положить в (22) $\lambda_n = 0$, $p_{n+1} = p_{n+2} = \dots = 0$. В этом случае последнее уравнение в (23) будет иметь вид $\gamma n p_n = \lambda p_{n-1}$. Так как формулы (24) при $k \leq n$ сохраняются, то получаем для стационарного распределения так называемые формулы Эрланга:

$$p_k = \left(\frac{\lambda}{\gamma} \right)^k \frac{1}{k!} \left[\sum_{j=0}^n \left(\frac{\lambda}{\gamma} \right)^j \frac{1}{j!} \right]^{-1}$$

(усеченное распределение Пуассона).

Следующий пример мы выделим в отдельный раздел.

§ 3. Ветвящиеся процессы

Существо математической модели, описывающей ветвящийся процесс, остается примерно тем же, что в п. 2 § 7.8. В условиях непрерывного времени задавать ветвящийся процесс можно следующим образом. Пусть $\xi^{(i)}(t)$ означает число частиц в момент времени t при начальном условии $\xi^{(i)}(0) = i$. Каждая частица независимо

от других за интервал времени $(t, t + \Delta)$ с вероятностью $\mu\Delta + o(\Delta)$ переходит в случайное число $\eta \neq 1$ частиц (при $\eta = 0$ мы говорим, что частица погибла). Таким образом,

$$\xi^{(i)}(t) = \xi_1^{(1)}(t) + \dots + \xi_i^{(1)}(t), \quad (25)$$

где $\xi_k^{(1)}(t)$ независимы и распределены как $\xi^{(1)}(t)$. Кроме того,

$$\begin{aligned} p_{ij}(\Delta) &= i\mu\Delta p_{j-i+1} + o(\Delta); \quad j \neq i; \quad p_k = P(\eta = k); \quad p_1 = 0; \\ p_{ii}(\Delta) &= 1 - i\mu\Delta + o(\Delta), \end{aligned} \quad (26)$$

так что здесь $q_{ij} = i\mu p_{j-i+1}$, $q_{ii} = -i\mu$.

Формула (26) объясняется тем, что $i\mu\Delta$ есть главная часть вероятности того, что хотя бы одна частица будет делиться. Очевидно, что состояние 0 является поглощающим.

Оно не будет таковым, если рассматривать процессы с *иммиграцией*; когда к $\xi^{(i)}(t)$ добавляется пуассоновский процесс с параметром λ «посторонних» частиц того же типа. Тогда

$$\begin{aligned} p_{ij}(\Delta) &= i\mu\Delta p_{j-i+1} + o(\Delta) \quad \text{при } j - i \neq 0, 1, \\ p_{i,i+1}(\Delta) &= \Delta(i\mu p_2 + \lambda) + o(\Delta). \end{aligned}$$

Вернемся к ветвящемуся процессу (25), (26). В силу (25) имеем

$$\varphi^{(i)}(t, z) = E z^{\xi^{(i)}(t)} = [E z^{\xi^{(1)}(t)}]^i = \varphi^i(t, z) = \sum_{k=0}^{\infty} z^k p_{ik}(t),$$

где

$$\varphi(t, z) = E z^{\xi^{(1)}(t)} = \sum_{k=0}^{\infty} z^k p_{1k}(t). \quad (27)$$

Из второго уравнения (16) следует, что

$$p'_{1k}(t) = \sum_{l=0}^{\infty} q_{1l} p_{lk}(t).$$

Поэтому, дифференцируя (27) по t , получим

$$\begin{aligned} \varphi'(t, z) &= \sum_{k=0}^{\infty} z^k p'_{1k}(t) = \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} q_{1l} p_{lk}(t) z^k = \\ &= \sum_{l=0}^{\infty} q_{1l} \sum_{k=0}^{\infty} z^k p_{lk}(t) = \sum_{l=0}^{\infty} q_{1l} \varphi^l(t, z). \end{aligned} \quad (28)$$

Но $q_{1l} = \mu p_l$ при $l \neq 1$, $q_{11} = -\mu$, и если обозначить

$$f(s) = \sum_{l=0}^{\infty} q_{1l} s^l = \mu(E s^\eta - s) = \mu \left(\sum_{l=0}^{\infty} p_l s^l - s \right),$$

то (28) можно записать в виде

$$\varphi'(t, z) = f(\varphi(t, z)).$$

Мы получили дифференциальное уравнение для $\varphi = \varphi(t, z)$ (эквивалентное (16)), которое удобнее записать в виде

$$\frac{d\varphi}{f(\varphi)} = dt, \quad t = \int_{\varphi(0,z)}^{\varphi(t,z)} \frac{dy}{f(y)} = \int_z^{\varphi(t,z)} \frac{dy}{f(y)}.$$

Рассмотрим поведение функции $f_1(y) = E\eta^y - y$ на $[0, 1]$. Очевидно, что $f_1(0) = P(\eta = 0)$, $f_1(1) = 0$, $f_1'(1) = E\eta - 1$, $f_1''(y) = E\eta(\eta - 1)y^{\eta-2} > 0$. Следовательно, $f_1(y)$ выпукла вниз и не имеет нулей на $(0, 1)$ при $E\eta \leq 1$. При $E\eta > 1$ существует точка $q \in (0, 1)$ такая, что $f_1(q) = 0$, $f_1'(q) < 0$ (рис. 18), и в окрестности этой точки $f_1(y) = (y - q)f_1'(q) + O((y - q)^2)$.

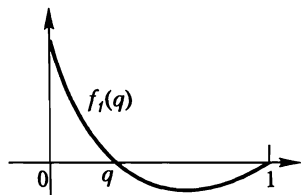


Рис. 18.

Таким образом, если $E\eta > 1$, $z < q$, $\varphi \uparrow q$, то в силу представления $1/f_1(y) = 1/((y - q)f_1'(q)) + O(1)$ получим

$$t = \int_z^{\varphi} \frac{dy}{f(y)} = \frac{1}{\mu f_1'(q)} \ln\left(\frac{\varphi - q}{z - q}\right) + O(1).$$

Отсюда следует, что при $t \rightarrow \infty$,

$$\varphi(t, z) - q = (z - q)e^{\mu t f_1'(q) + O(1)} \sim (z - q)e^{\mu t f_1'(q)},$$

$$\varphi(t, z) = q + O(e^{-\alpha t}), \quad \alpha = -\mu f_1'(q) > 0. \quad (29)$$

В частности, вероятность вырождения

$$p_{10}(t) = \varphi(t, 0) = q + O(e^{-\alpha t})$$

экспоненциально быстро сходится к q , $p_{10}(\infty) = q$. Сравнивая с § 7.8, читатель может убедиться, что такой же была вероятность вырождения для ветвящегося процесса с дискретным временем (к этому выводу можно прийти и непосредственно). Так как $p_{k0}(t) = [p_{10}(t)]^k$, то $p_{k0}(\infty) = q^k$.

Из (29) следует, что оставшаяся «вероятностная масса» распределения $\xi(t)$ быстро уходит при $t \rightarrow \infty$ на бесконечность.

Если $E\eta < 1$, то все приведенные рассуждения сохраняются с заменой q на 1, так что $p_{10}(\infty) = p_{k0}(\infty) = 1$.

Если $E\eta = 1$, то

$$f_1(y) = \frac{(y - 1)^2}{2} f_1''(1) + O((y - 1)^3),$$

$$t = \int_z^{\varphi} \frac{dy}{f(y)} \sim -\frac{2}{\mu f_1''(1)} \frac{1}{\varphi - 1},$$

$$\varphi(t, z) - 1 \sim -\frac{2}{\mu t f_1''(1)}.$$

Таким образом, здесь вероятность вырождения $\varphi(t, 0) = p_{10}(t)$ также сходится к 1.

§ 4. Полумарковские и регенерирующие процессы

1. Полумарковские процессы. Полумарковские процессы описываются следующим образом. Пусть дана цепь Маркова $\{X_n\}$ с дискретным временем n и множеством состояний $\mathcal{X} = \{0, 1, 2, \dots\}$. Каждому состоянию j поставим в соответствие распределение $F_i(t)$ положительной случайной величины $\zeta^{(i)}$:

$$F_i(t) = P(\zeta^{(i)} < t).$$

Рассмотрим далее не зависящие от $\{X_n\}$ и между собой последовательности $\zeta_1^{(i)}, \zeta_2^{(i)}, \dots$ независимых случайных величин с распределением F_i . Пусть, кроме того, задано распределение начального случайного вектора (X_0, ζ_0) , $X_0 \in \mathcal{X}$, $\zeta_0 \geq 0$. Эволюция полумарковского процесса $\xi(t)$ описывается следующим образом.

$$\begin{aligned} \xi(u) &= X_0 \quad \text{при} \quad 0 \leq u < \zeta_0, \\ \xi(u) &= X_1 \quad \text{при} \quad \zeta_0 \leq u < \zeta_0 + \zeta_1^{(X_1)}, \\ \xi(u) &= X_2 \quad \text{при} \quad \zeta_0 + \zeta_1^{(X_1)} \leq u < \zeta_0 + \zeta_1^{(X_1)} + \zeta_2^{(X_2)}, \\ &\text{и т.д. } \dots, \\ \xi(u) &= X_n \quad \text{при} \quad Z_{n-1} \leq u < Z_n, \quad Z_n = \zeta_0 + \zeta_1^{(X_1)} + \dots + \zeta_n^{(X_n)} \end{aligned} \quad (30)$$

и т.д. Таким образом, попав в состояние $X_n = j$, траектория $\xi(u)$ находится в этом состоянии время $\zeta_n^{(X_n)} = \zeta_n^{(j)}$, затем переходит в состояние X_{n+1} и т.д. Очевидно, что такой процесс, вообще говоря, не является марковским. Он будет марковским, только если

$$1 - F_i(t) = e^{-q_i t}, \quad q_i > 0$$

и будет совпадать в этом случае с процессом, описанным в § 2.

Если же распределение F_i не является экспоненциальным, и дано значение $\xi(t) = i$, то расстояние от точки t до момента следующего скачка $\xi(t)$ будет зависеть от момента предыдущего скачка $\xi(t)$, так как

$$P(\zeta^{(i)} > v + u | \zeta^{(i)} > v) = \frac{1 - F_i(v + u)}{1 - F_i(v)}$$

для не экспоненциальных F_i зависит от v . Это и означает отсутствие марковости, т.к. означает, что фиксация «настоящего» (т.е. значения $\xi(t)$) не делает «будущее» процесса $\xi(u)$ независимым от «прошлого» (т.е. от траектории $\xi(u)$ при $u < t$).

Процесс $\xi(t)$ можно «дополнить» до марковского, присоединив к нему компоненту $\chi(t)$, означающую время u , которое траектория $\xi(t + u)$, $u \geq 0$, будет еще находиться в состоянии $\xi(t)$. Другими словами, $\chi(t)$ есть эксцесс уровня t в случайном блуждании Z_0, Z_1, \dots :

$$\chi(t) = Z_{\nu(t)} - t, \quad \nu(t) = \min\{k: Z_k > t\}.$$

Процесс $\chi(t)$ является марковским и имеет «пилообразные» траектории, детерминированные внутри интервалов (Z_k, Z_{k+1}) . Последним свойством обладает, очевидно, и траектория $\xi(t)$. Процесс $X(t) = (\xi(t), \chi(t))$, очевидно, является марковским, т.к. знание $X(t)$ однозначно определяет эволюцию процесса $X(t + u)$ при $u \geq 0$ при любой «предыстории» $X(v)$, $v < t$. Аналогичным образом можно

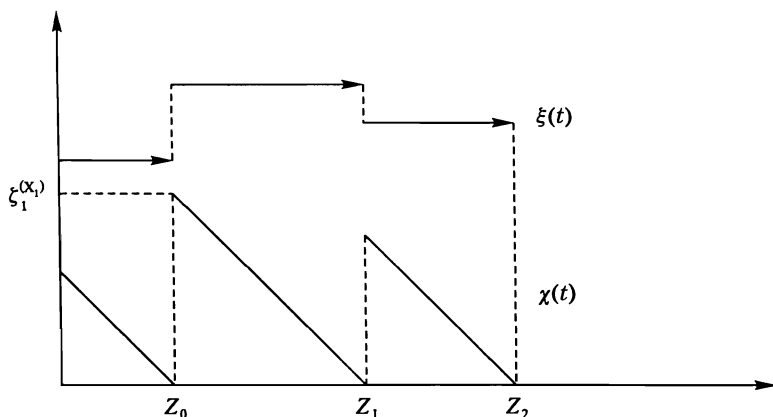


Рис. 19.

рассматривать марковский процесс $Y(t) = (\xi(t), \gamma(t))$, где $\gamma(t)$ — дефект уровня t в блуждании Z_0, Z_1, \dots :

$$\gamma(t) = t - Z_{\nu(t)-1}.$$

В дальнейшем мы будем различать два случая:

(L) — *решетчатый*, когда возможные значения $\zeta^{(i)}$, $i = 0, 1, \dots$ кратны некоторому числу d , которое, не ограничивая общности, можно считать равным 1. В этом случае мы будем считать также, что н.о.д. возможных значений сумм величин $\zeta^{(i)}$ вдоль траектории $\{X_n\}$ тоже равен $d = 1$. Это, очевидно, эквивалентно тому, что н.о.д. возможных значений времени $\theta^{(i)}$ возвращения $\xi(t)$ в состояние i равен 1 при любом фиксированном i .

(NL) — *нерешетчатый* случай, когда условие (L) не выполнено.

Обозначим $a_i = E\zeta^{(i)}$.

Теорема 5. Пусть цепь Маркова $\{X_n\}$ является эргодической (удовлетворяет условиям теоремы 12.8) и $\{\pi_j\}$ есть стационарное распределение этой цепи. Тогда в нерешетчатом случае (NL) для любого начального распределения (ζ_0, X_0) существует

$$\lim_{t \rightarrow \infty} P(\xi(t) = i, \chi(t) > v) = \frac{\pi_i}{\sum \pi_j a_j} \int_v^\infty P(\zeta^{(i)} > u) du. \quad (31)$$

В решетчатом случае (L) значения v в (31) следует брать целочисленными (интеграл при этом перейдет в сумму).

Из (31) следует, что существует

$$\lim_{t \rightarrow \infty} P(\xi(t) = i) = \frac{\pi_i a_i}{\sum \pi_j a_j}.$$

Доказательство. Ограничимся для определенности нерешетчатым случаем (NL). В § 12.4 мы рассматривали времена $\tau^{(i)}$ возвращения $\{X_n\}$ в состояние i . Эти времена можно называть «вложенными», как и цепь $\{X_n\}$ по отношению к процессу $\xi(t)$. Наряду с $\tau^{(i)}$ нам понадобятся «реальные» времена $\theta^{(i)}$ возвращения

процесса $\xi(t)$ в состояние i . Пусть, например, $X_1 = 1$. Тогда

$$\theta^{(1)} = \zeta_1^{(X_1)} + \zeta_2^{(X_2)} + \dots + \zeta_\tau^{(X_\tau)},$$

где $\tau = \tau^{(1)}$. Для сокращения записи и для определенности фиксируем на время $i = 1$ и положим $\theta^{(1)} = \theta$. Пусть сначала

$$\zeta_0 = \zeta_p^{(1)}, \quad X_0 = 1. \quad (32)$$

Тогда вся траектория процесса $X(t)$ при $t \geq 0$ разбивается на одинаково устроенные и независимые циклы по моментам попадания в состояние $\xi(t) = 1$. Длины этих циклов обозначим $\theta_1, \theta_2, \dots$; они независимы и одинаково распределены. Покажем, что

$$E\theta = \frac{1}{\pi_1} \sum a_j \pi_j. \quad (33)$$

Обозначим $\theta(n)$ «реальное» время, затраченное на n переходов «управляющей» цепи $\{X_n\}$. Тогда

$$\theta_1 + \dots + \theta_{\eta(n)-1} \leq \theta(n) \leq \theta_1 + \dots + \theta_{\eta(n)}, \quad (34)$$

где $\eta(n) = \min\{k: T_k > n\}$, $T_k = \sum_{j=1}^k \tau_j$, τ_j независимы и распределены как τ . Докажем, что если $E\theta < \infty$, то

$$E\theta(n) \sim n\pi_1 E\theta. \quad (35)$$

В силу тождества Вальда

$$E\theta(n) \leq E\theta E\eta(n), \quad (36)$$

где $E\eta(n) \sim n/(E\tau) = n\pi_1$.

Оценим теперь $E\theta(n)$ снизу. Положим $m = [n\pi_1 - \varepsilon n]$, $\Theta_n = \sum_{j=1}^n \theta_j$. Тогда

$$E\theta(n) \geq E(\theta(n); \eta(n) > m) \geq E(\Theta_m; \eta(n) > m) = mE\theta - E(\Theta_m; \eta(n) \leq m). \quad (37)$$

Здесь случайная величина $\Theta_m/m \geq 0$ обладает свойствами: $\Theta_m/m \xrightarrow{p} E\theta$, $E\Theta_m/m = E\theta$. Следовательно, она удовлетворяет условиям п. 4 леммы 6.1 и является равномерно интегрируемой. Это, в свою очередь, в силу леммы 6.2 и сходимости $P(\eta(n) \leq m) \rightarrow 0$ означает, что последнее слагаемое в правой части (37) есть $o(m)$. В силу (37) и произвольности $\varepsilon > 0$ мы получаем

$$\liminf n^{-1} E\theta(n) \geq \pi_1 E\theta.$$

Вместе с (36) это доказывает (35).

Подсчитаем теперь значение $E\theta(n)$ из других соображений. $\theta(n)$ допускает представление

$$\theta(n) = \sum_j (\zeta_1^{(j)} + \dots + \zeta_{N(j,n)}^{(j)}),$$

где $N(j, n)$ число попаданий траектории $\{X_k\}$ в состояние j за время n . Так как $\{\zeta_k^{(j)}\}_{k=1}^\infty$ и $N(j, n)$ при каждом j независимы, то

$$\begin{aligned} E\theta(n) &= \sum_j a_j E N(j, n), \\ E N(j, n) &= \sum_{k=1}^n p_{1j}(k). \end{aligned}$$

Так как $p_{1j}(k) \rightarrow \pi_j$ при $k \rightarrow \infty$, то $\lim_{n \rightarrow \infty} \frac{1}{n} E N(j, n) = \pi_j$. Кроме того, $\pi_j = \sum \pi_i p_{ij}(k) \geq \pi_1 p_{1j}(k)$ и, стало быть, $p_{1j}(k) \leq \frac{\pi_j}{\pi_1}$. Поэтому $\frac{1}{n} E N(j, n) \leq \frac{\pi_j}{\pi_1}$, и в случае $\sum a_j \pi_j < \infty$ ряд $\sum a_j \frac{1}{n} E N(j, n)$ сходится равномерно по n . Следовательно, существует

$$\lim_{n \rightarrow \infty} n^{-1} E\theta(n) = \sum_j a_j \pi_j.$$

Сравнивая это с (35), мы получим (33). Если $E\theta = \infty$, то очевидно $E\theta(n) = \infty$, $\sum a_j \pi_j = \infty$ и наоборот, если $\sum a_j \pi_j = \infty$, то $E\theta = \infty$.

Рассмотрим теперь случайное блуждание $\{\Theta_k\}$. Циклу с номером k соответствует число переходов T_k . Поэтому по формуле полной вероятности

$$P(\xi(t) = 1, \chi(t) > v) = \sum_{k=1}^{\infty} \int_0^t P(\Theta_k \in du, \zeta_{T_k+1}^{(1)} > t - u + v),$$

где $\zeta_{T_k+1}^{(1)}$ от Θ_k не зависит и распределено как $\zeta^{(1)}$ (см. лемму 9.1 или строго марковское свойство). Стало быть, если обозначить $H_\theta(u) = \sum_{k=1}^{\infty} P(\Theta_k < u)$ функцию восстановления последовательности $\{\Theta_k\}$, то для нерешетчатого случая (NL) в силу теоремы восстановления (см. теорему 9.4 и (9.10)) будем иметь

$$\begin{aligned} P(\xi(t) = 1, \chi(t) > v) &= \int_0^t dH_\theta(u) P(\zeta^{(1)} > t - u + v) \rightarrow \\ &\rightarrow \frac{1}{E\theta} \int_0^\infty P(\zeta^{(1)} > u + v) dv = \frac{1}{E\theta} \int_v^\infty P(\zeta^{(1)} > u) du. \end{aligned} \quad (38)$$

Мы доказали утверждение (31) для $i = 1$ и начальных условий (32). Переход к произвольным начальным условиям вполне очевиден и происходит точно так же как в доказательстве эргодических теорем в гл. 12.

Если $\sum a_i \pi_i = \infty$, то, как уже отмечалось, $E\theta = \infty$, и в силу теоремы восстановления и (38) $P(\xi(t) = 1, \chi(t) > v) \rightarrow 0$ при $t \rightarrow \infty$. Остается заметить, что вместо $i = 1$ мы могли фиксировать любое другое значение i . Теорема доказана. ■

Совершенно аналогично мы могли бы доказать также, что

$$\lim_{t \rightarrow \infty} P(\xi(t) = i, \gamma(t) > v) = \frac{\pi_i}{\sum_j a_j \pi_j} \int_v^{\infty} P(\zeta^{(i)} > y) dy,$$

$$\lim_{t \rightarrow \infty} P(\xi(t) = i, \chi(t) > u, \gamma(t) > v) = \frac{\pi_i}{\sum_j a_j \pi_j} \int_{u+v}^{\infty} P(\zeta^{(i)} > y) dy$$

(см. теорему 9.6).

Наряду с полумарковскими процессами $\xi(t)$, описанными в начале этого параграфа, иногда рассматривают также полумарковские процессы, «заданные на переходах» цепи $\{X_n\}$. В этом случае задаются распределения F_{ij} случайных величин $\zeta^{(ij)} > 0$ и аналогично (30) при начальном условии (X_0, X_1, ζ_0)

$$\begin{aligned} \xi(u) &= (X_0, X_1) \quad \text{при} \quad 0 \leq u < \zeta_0, \\ \xi(u) &= (X_1, X_2) \quad \text{при} \quad \zeta_0 \leq u < \zeta_0 + \zeta_1^{(X_0, X_1)}, \\ \xi(u) &= (X_2, X_3) \quad \text{при} \quad \zeta_0 + \zeta_1^{(X_0, X_1)} \leq u < \zeta_0 + \zeta_1^{(X_0, X_1)} + \zeta_2^{(X_1, X_2)} \end{aligned} \quad (39)$$

и т.д. Эта модель, на первый взгляд более общая, может быть полностью редуцирована к полумарковским процессам (30). Для этого надо заметить, что «двумерная» последовательность $Y_n = (X_n, X_{n+1})$, $n = 0, 1, \dots$, также образует цепь Маркова. Ее переходные вероятности имеют вид

$$p_{(ij)(kl)} = \begin{cases} p_{jl} & \text{при } k = j, \\ 0 & \text{при } k \neq j, \end{cases}$$

$$p_{(ij)(kl)}(n) = p_{jk}(n) p_{kl} \quad \text{при } n > 1,$$

так что, если цепь $\{X_n\}$ эргодична, то цепь $\{Y_n\}$ также эргодична и

$$p_{(ij)(kl)}(n) \rightarrow \pi_k p_{kl}.$$

Это дает возможность легко переформулировать теорему 5 для полумарковских процессов (39), заданных на переходах цепи $\{X_n\}$, поскольку процесс (39) будет обычным полумарковским процессом, заданным на цепи $\{Y_n\}$.

Следствие 1. Если цепь $\{X_n\}$ эргодична, то в нерешетчатом случае

$$P(\xi(t) = (i, j), \chi(t) > v) = \frac{\pi_i p_{ij}}{\sum_k a_k \pi_k p_{kl}} \int_v^{\infty} P(\zeta^{(ij)} > u) du, \quad a_{kl} = E \zeta^{(kl)}.$$

В решетчатом случае v должно быть кратно шагу решетки.

Сделаем еще одно замечание, которое может быть полезным при изучении полумарковских процессов. Речь будет идти о так называемых полумарковских функциях восстановления $H_{ij}(t)$. Обозначим $T_{ij}(n)$ момент (в «реальном» времени) n -го скачка процесса $\xi(t)$ из состояния i в состояние j . Положим

$$H_{ij}(t) = \sum_{n=1}^{\infty} P(T_{ij}(n) < t).$$

Если $\nu_{ij}(t)$ число скачков из состояния i в состояние j на интервале времени $[0, t)$, то, очевидно, $H_{ij}(t) = E \nu_{ij}(t)$.

Обозначим $\Delta f(t) = f(t + \Delta) - f(t)$, $\Delta > 0$.

Следствие 2. В нерешетчатом случае

$$\lim_{t \rightarrow \infty} \Delta H_{ij}(t) = \frac{\pi_i p_{ij} \Delta}{\sum_l a_l \pi_l}. \quad (40)$$

В решетчатом случае t и Δ должны быть кратны шагу решетки.

Доказательство. Обозначим $\nu_{ij}^{(k)}(u)$ число переходов процесса $\xi(t)$ за интервал времени $(0, u)$ из i в j при начальном условии $(k, 0)$. Тогда по формуле полной вероятности

$$E \Delta \nu_{ij}(t) = \int_0^{\Delta} \sum_{k=0}^{\infty} P(\xi(t) = k, \chi(t) \in du) E \nu_{ij}^{(k)}(\Delta - u).$$

Так как $\nu_{ij}^{(k)}(u) \leq \nu_{ij}^{(i)}(u)$, то по теореме 5

$$\lim_{t \rightarrow \infty} E \Delta \nu_{ij}(t) \equiv h_{ij}(\Delta) = \frac{1}{\sum_l a_l \pi_l} \sum_{k=0}^{\infty} \pi_k \int_0^{\Delta} P(\zeta^{(k)} > u) du E \nu_{ij}^{(k)}(\Delta - u). \quad (41)$$

Далее

$$P(\zeta^{(i)} < \Delta - u) \leq F_i(\Delta) \rightarrow 0$$

при $\Delta \rightarrow 0$,

$$P(\nu_{ij}^{(k)}(\Delta - u) = s) \leq (p_{ij} F_i(\Delta))^s, \quad k \neq i,$$

$$P(\nu_{ij}^{(i)}(\Delta - u) = s + 1) \leq (p_{ij} F_i(\Delta))^s, \quad s \geq 1,$$

$$P(\nu_{ij}^{(i)}(\Delta - u) = 1) = p_{ij} + o(F_i(\Delta)).$$

Из сказанного следует, что

$$E \nu_{ij}^{(k)}(\Delta - u) = o(F_i(\Delta)), \quad E \nu_{ij}^{(i)}(\Delta - u) = p_{ij} + o(F_i(\Delta)).$$

Поэтому

$$h_{ij}(\Delta) = \frac{\pi_i p_{ij} \Delta}{\sum_l a_l \pi_l} + o(\Delta). \quad (42)$$

Далее в силу равенства

$$H_{ij}(t + 2\Delta) - H_{ij}(t) = \Delta H_{ij}(t) + \Delta H_{ij}(t + \Delta)$$

получаем $h_{ij}(2\Delta) = 2h_{ij}(\Delta)$, что означает линейность $h_{ij}(\Delta)$. Вместе с (42) это доказывает (40). ■

Класс процессов, для которых удастся доказать эргодичность теми же методами, что использовались для полумарковских процессов и в гл. 12, может быть несколько расширен.

2. Регенерирующие процессы. Пусть $X(t)$, $X_0(t)$; $t \geq 0$ — процессы, заданные в пространстве $D(0, \infty)$ функций без разрывов второго рода (фазовым пространством этих процессов может быть не обязательно вещественная прямая, это может быть и любое метрическое пространство). Процесс $X(t)$ мы назовем *регенерирующим*, если он обладает следующими свойствами:

1) Существует состояние x_0 , в которое процесс X попадает с вероятностью 1. После каждого такого попадания его эволюция совпадает с эволюцией исходного процесса $X(t)$ при начальном значении $X(0) = x_0$. Этот новый процесс мы обозначим $X_0(t)$, $X_0(0) = x_0$. Чтобы сформулировать это свойство более точно, введем в рассмотрение момент t_0 первого попадания X в x_0 :

$$t_0 = \min\{t \geq 0: X(t) = x_0\}.$$

Из такого определения не ясно, однако, будет ли t_0 случайной величиной. Для определенности предположим процесс X таким, что t_0 есть величина, для которой

$$\{t_0 > t\} = \bigcup_n \bigcap_{t_k \in S} \left\{ |X(t_k) - x_0| > \frac{1}{n} \right\},$$

где S — счетное, всюду плотное на $[0, t]$ множество. В этом случае $\{t_0 > t\}$, очевидно, есть событие, а t_0 — случайная величина. Описанное выше свойство означает, что t_0 — собственная случайная величина: $P(t_0 < \infty) = 1$ и что распределение $X(t_0 + u)$, $u \geq 0$, совпадает с распределением $X_0(u)$, $u \geq 0$, при любой «предыстории» процесса $X(t)$, $t \leq t_0$.

2) Время возвращения τ в состояние x_0 имеет конечное математическое ожидание $E\tau < \infty$, $\tau = \min\{t: X_0(t) = x_0\}$.

Сказанное означает, что эволюция регенерирующего процесса разбивается на независимые одинаково устроенные циклы по попаданию в состояние x_0 . Моменты попадания в x_0 называются моментами регенерации. Поведение процесса X внутри циклов может быть произвольным, никаких условий, в том числе условий марковости, не накладывается.

Введем в рассмотрение так называемую «переходную табу-вероятность»

$$P(t, B) = P(X_0(t) \in B, \tau > t).$$

Мы будем предполагать, что эта функция измерима по t и интегрируема по Риману.

Теорема 6. Пусть $X(t)$ регенерирующий процесс, случайная величина τ нерешетчата. Тогда для любого борелевского множества B

$$P(X(t) \in B) \rightarrow \pi(B) = \frac{1}{E\tau} \int_0^\infty P(u, B) du.$$

Если τ решетчата (это имеет место для процессов $X(t)$ с дискретным временем), то утверждение теоремы сохраняется при очевидных изменениях: $t \rightarrow \infty$ по элементам решетки, интеграл заменится на сумму.

Доказательство. Обозначим $T_0 = 0$,

$$T_k = \tau_1 + \dots + \tau_k$$

момент k -ой регенерации процесса $X_0(t)$, $H(u) = \sum_{k=0}^{\infty} P(\tau_k < u)$. Тогда, используя формулу полной вероятности и основную теорему восстановления, мы получим

$$\begin{aligned} P(X_0(t) \in B) &= \sum_{k=0}^{\infty} \int_0^t P(T_k \in du) P(t-u, B) = \\ &= \int_0^t dH(u) P(t-u, B) \rightarrow \frac{1}{E\tau} \int_0^{\infty} P(u, B) du = \pi(B). \end{aligned}$$

Для процесса $X(t)$ выполняется

$$P(X(t) \in B) = \int_0^t P(t_0 \in du) P(X_0(t-u) \in B) \rightarrow \pi(B).$$

Теорема доказана. ■

§ 5. Диффузионные процессы

Рассмотрим еще один важный класс марковских процессов, на этот раз с непрерывными траекториями.

Определение 6. Однородный марковский процесс $\xi(t)$ в фазовом пространстве $(\mathbb{R}, \mathfrak{B})$ с переходной функцией $P(t, x, B)$ называется *диффузионным*, если

- 1) $\lim_{\Delta \rightarrow 0} \frac{1}{\Delta} \int (y-x) P(\Delta, x, dy) = a(x),$
- 2) $\lim_{\Delta \rightarrow 0} \frac{1}{\Delta} \int (y-x)^2 P(\Delta, x, dy) = b^2(x),$
- 3) при каких-нибудь $\delta > 0, c < \infty$

$$\int (y-x)^{2+\delta} P(\Delta, x, dy) < c\Delta^{1+\delta/2}.$$

Обозначим $\Delta\xi(t) = \xi(t+\Delta) - \xi(t)$. Тогда приведенные условия можно записать в виде

$$\begin{aligned} E[\Delta\xi(t) | \xi(t) = x] &\sim a(x)\Delta, \\ E[(\Delta\xi(t))^2 | \xi(t) = x] &\sim b^2(x)\Delta, \\ E[|\Delta\xi(t)|^{2+\delta} | \xi(t) = x] &< c\Delta^{1+\delta/2}. \end{aligned}$$

Коэффициенты $a(x)$, $b(x)$ называют соответственно *коэффициентами сноса и диффузии*. Условие 3) есть некоторый аналог условия Ляпунова. Его можно заменить на условие типа Линдеберга:

$$3a) E[(\Delta\xi(t))^2; |\Delta\xi(t)| > \varepsilon] = o(\Delta) \text{ при любом } \varepsilon > 0.$$

Из условия 3) и теоремы Колмогорова немедленно вытекает, что диффузионный процесс $\xi(t)$ можно рассматривать как процесс с непрерывными траекториями.

Стандартный винеровский процесс $w(t)$ является диффузионным процессом, так как для него

$$P(t, x, B) = \frac{1}{\sqrt{2\pi t}} \int_B e^{-(x-y)^2/(2t)} dy,$$

$$E\Delta w(t) = 0, \quad E[\Delta w(t)]^2 = \Delta, \quad E[\Delta w(t)]^4 = 3\Delta^2.$$

Значит, винеровский процесс имеет нулевой снос и постоянный коэффициент диффузии. Ясно, что процесс $w(t) + at$ будет иметь снос a и тот же коэффициент диффузии.

Мы видели в § 2, что «локальная» характеристика Q марковского процесса $\xi(t)$ с дискретным \mathcal{X} однозначно определяет эволюцию процесса. Аналогичная картина наблюдается и здесь: *распределение процесса однозначно определяется коэффициентами $a(x)$ и $b(x)$* . Путь для установления этого факта опять лежит через уравнение Колмогорова—Чепмэна.

Теорема 7. *Если переходные вероятности диффузионного процесса $P(t, x, B)$ дважды непрерывно дифференцируемы по x , то $P(t, x, B)$ дифференцируема по t и удовлетворяет уравнению*

$$\frac{\partial P}{\partial t} = a \frac{\partial P}{\partial x} + \frac{b^2}{2} \frac{\partial^2 P}{\partial x^2} \quad (43)$$

и начальному условию

$$P(0, x, B) = I_B(x). \quad (44)$$

Замечание 1. Условия теоремы на гладкость переходной функции P на самом деле могут быть доказаны в предположении непрерывности a и b и условий $b \geq b_0 > 0$, $|a| \leq c(|x| + 1)$, $b^2 \leq c(|x| + 1)$.

Доказательство теоремы 7. Обозначим для краткости через P'_t , P'_x , P''_x соответственно частные производные $\partial P/\partial t$, $\partial P/\partial x$, $\partial^2 P/\partial x^2$, и воспользуемся разложением

$$\begin{aligned} P(t, y, B) - P(t, x, B) &= (y - x)P'_x + \frac{(y - x)^2}{2}P''_x + \\ &+ \frac{(y - x)^2}{2}[P''_x(t, y_x, B) - P''_x(t, x, B)] \quad y_x \in (x, y). \end{aligned} \quad (45)$$

Тогда в силу уравнения Колмогорова—Чепмэна

$$\begin{aligned} P(t + \Delta, x, B) - P(t, x, B) &= \int P(\Delta, x, dy)[P(t, y, B) - P(t, x, B)] = \\ &= a(x)P'_x\Delta + \frac{b^2(x)}{2}P''_x\Delta + o(\Delta) + R, \end{aligned} \quad (46)$$

где

$$R = \int \frac{(y - x)^2}{2} [P''_x(t, y_x, B) - P''_x(t, x, B)] = \int_{|y-x| \leq \varepsilon} + \int_{|y-x| > \varepsilon}.$$

Первый интеграл здесь в силу непрерывности P''_x не превосходит $\delta(\varepsilon)[b^2(x)/2\Delta + o(\Delta)]$, $\delta(\varepsilon) \rightarrow 0$ при $\varepsilon \rightarrow 0$; второй интеграл по условию 3а) есть $o(\Delta)$. Так как ε произвольно, то $R = o(\Delta)$ и из сказанного выше следует, что существует

$$\lim_{\Delta \rightarrow 0} \frac{P(t + \Delta, x, B) - P(t, x, B)}{\Delta} = a(x)P'_x + \frac{b^2(x)}{2}P''_x.$$

Из теории дифференциальных уравнений известно, что при широких предположениях о коэффициентах a , b и при $B = (-\infty, z)$ задача Коши (43), (44) для уравнения (43) имеет единственное решение P , бесконечное число раз дифференцируемое по t , x и z . Отсюда следует, в частности, что $P(t, x, B)$ имеет плотность $p(t, x, z)$, являющуюся фундаментальным решением (43).

Из теоремы 7 нетрудно получить также, что вместе с $P(t, x, B)$ уравнению (43) будут удовлетворять функции $u(t, x) = \int \varphi(z)P(t, x, dz) = E[\varphi(\xi^{(x)}(t))]$ для любой гладкой функции φ с конечным носителем.

Мы рассматривали в доказательстве теоремы 7 (см. (46)) приращение времени Δ , предшествующее основному интервалу. В связи с этим уравнения (43) называются *обратными* уравнениями Колмогорова. Аналогичным образом могут быть получены и *прямые уравнения*.

Теорема 8 (прямые уравнения Колмогорова). Пусть переходная плотность $p(t, x, y)$ такова, что существуют непрерывные производные

$$\frac{\partial}{\partial y} [a(y)p(t, x, y)] \quad \text{и} \quad \frac{\partial^2}{\partial y^2} [b^2(y)p(t, x, y)].$$

Тогда $p(t, x, y)$ удовлетворяет уравнению

$$Dp \equiv \frac{\partial p}{\partial t} + \frac{\partial}{\partial y} [a(y)p(t, x, y)] - \frac{1}{2} \frac{\partial^2}{\partial y^2} [b^2(y)p(t, x, y)] = 0. \quad (47)$$

Доказательство. Пусть $\varphi(y)$ — гладкая функция с ограниченным носителем, $u(t, x) = E\varphi(\xi^{(x)}(t)) = \int \varphi(y)p(t, x, y) dy$. Тогда

$$u(t + \Delta, x) - u(t, x) = \int p(t, x, z) \left[\int p(\Delta, z, y) \varphi(y) dy - \int p(\Delta, z, y) \varphi(z) dy \right] dz. \quad (48)$$

Если воспользоваться разложением в ряд $\varphi(y) - \varphi(z)$, то для выражения в квадратных скобках мы аналогично тому, как это было в доказательстве теоремы 5, в силу свойств 1)–3) получим

$$\left[a(z)\varphi'(z) + \frac{b^2(z)}{2}\varphi''(z) \right] \Delta + o(\Delta).$$

Отсюда следует, что существует

$$\frac{\partial u}{\partial t} = \int p(t, x, z) \left[a(z)\varphi'(z) + \frac{1}{2} \frac{b^2(z)}{2} \varphi''(z) \right] dz.$$

Используя интегрирование по частям, получим

$$\frac{\partial u}{\partial t} = \int \left\{ -\frac{\partial}{\partial z} [a(z)p(t, x, z)] + \frac{1}{2} \frac{\partial^2}{\partial z^2} [b^2(z)p(t, x, z)] \right\} \varphi(z) dz = 0,$$

или, что то же,

$$\int Dp(t, x, z) \varphi(z) dz = 0.$$

Так как φ произвольна, то отсюда следует (47). ■

Как и в случае дискретного \mathcal{X} , различие между прямым и обратным уравнениями Колмогорова становится более наглядным в случае неоднородных процессов диффузии, когда переходные вероятности $P(s, x; t, B)$ зависят от двух временных аргументов, а функции a и b в условиях 1)–3) будут функциями от s и x . Тогда обратное уравнение Колмогорова (для плотностей) будет связывать производные переходных плотностей $p(s, x; t, y)$ по первым двум переменным, а прямое — по вторым.

Вернемся к однородным процессам диффузии. Исследование условий, при которых существует стационарное распределение $\xi^{(x)}(t)$ при $t \rightarrow \infty$, не зависящее от x , может быть проведено с помощью того же подхода, что и в § 2. Теорема 3 при этом полностью сохранится (надо лишь обозначение i_0 заменить на x_0 для единообразия обозначений в условиях этого параграфа). Доказательство теоремы 3 также сохранится, однако потребует несколько более аккуратных рассуждений (в новой ситуации на событии B_{dv} выполняется $\xi(t-v) \in dx_0$ (а не $= x_0$)).

Если стационарное распределение

$$\lim_{t \rightarrow \infty} p(t, x, y) = p(y) \quad (49)$$

существует, то каким образом оно может быть найдено? Так как зависимость $p(t, x, y)$ при $t \rightarrow \infty$ от t и x становится исчезающей, то обратные уравнения Колмогорова превращаются при $t \rightarrow \infty$ в тождество $0 = 0$. Если обратиться к прямым уравнениям и перейти в (48) к пределу сначала по $t \rightarrow \infty$, а затем по $\Delta \rightarrow 0$, мы придем после тех же, что в теореме 3, рассуждений к следующему заключению.

Следствие 3. Если выполнено (49) и условия теоремы 8, то стационарная плотность $p(y)$ удовлетворяет уравнению

$$- [a(y)p(y)]' + \frac{1}{2} [b^2(y)p(y)]'' = 0$$

(которое получается из (47), если положить $\partial p / \partial t = 0$).

Пример 4. Процесс Орнштейна—Уленбека

$$\xi^{(x)}(t) = xe^{at} + \sigma e^{at} w \left(\frac{1 - e^{-2at}}{2a} \right),$$

где $w(u)$ — стандартный винеровский процесс, является однородным диффузионным процессом с переходной плотностью

$$p(t, x, y) = \frac{1}{\sqrt{2\pi}\sigma(t)} \exp \left\{ -\frac{(y - xe^{at})^2}{2\sigma^2(t)} \right\}, \quad \sigma^2(t) = \frac{\sigma^2}{2a} (e^{2at} - 1). \quad (50)$$

Читателю предоставляется проверить, что этот процесс имеет коэффициенты $a(x) = ax$, $b(x) = \sigma = \text{const}$, а функция (50) удовлетворяет прямому и обратному уравнениям. При $a < 0$ существует стационарный процесс (определение см. в следующей главе)

$$\xi(t) = \sigma e^{at} w \left(\frac{e^{-2at}}{2a} \right),$$

плотность распределения которого (она от t не зависит) равна

$$p(y) = \lim_{t \rightarrow \infty} p(x, t, y) = \frac{1}{\sqrt{2\pi}\sigma(\infty)} \exp \left\{ -\frac{y^2}{2\sigma^2(\infty)} \right\}, \quad \sigma^2(\infty) = -\frac{\sigma^2}{2a}.$$

В заключение этого параграфа рассмотрим важную для разного рода приложений задачу об отыскании вероятности того, что траектория диффузионного процесса не выйдет за пределы заданной полосы. Для простоты мы ограничимся рассмотрением этой задачи для винеровского процесса. Пусть $c > 0$, $d < 0$.

Обозначим

$$U(t, x, B) = P\left(w^{(x)}(u) \in (d, c) \text{ при всех } u \in [0, t]; w^{(x)}(t) \in B\right) = \\ = P\left(\sup_{u \leq t} w^{(x)}(u) < c, \inf_{u \leq t} w^{(x)}(u) > d, x^{(x)}(t) \in B\right).$$

Опустим проверку того факта, что функция U дважды непрерывно дифференцируема по x , и ограничимся тем, что докажем следующее предложение:

Теорема 9. Функция U удовлетворяет уравнению (43) с начальным условием

$$U(0, x, B) = I_B(x) \quad (51)$$

и с граничными условиями

$$U(t, c, B) = U(t, d, B) = 0. \quad (52)$$

Доказательство. Отметим прежде всего, что функция $U(t, x, B)$ при $x \in (d, c)$ удовлетворяет условиям 1)–3), налагаемым на переходную функцию $P(t, x, B)$. Действительно, рассмотрим, например, свойство 1).

Надо убедиться, что

$$\int_d^c (y - x) U(\Delta, x, dy) \sim \Delta a(x) + o(\Delta) \quad (53)$$

(при $a(x) = 0$ в нашем случае). Но $U(t, x, B) = P(t, x, B) - V(t, x, B)$, где

$$V(t, x, B) = P\left(\left\{\sup_{u \leq t} w^{(x)}(u) \geq c \text{ или } \inf_{u \leq t} w^{(x)}(u) \leq d\right\} \cap \left\{w^{(x)}(t) \in B\right\}\right), \\ \int_d^c (y - x) V(\Delta, x, dy) \leq \max(c, -d) \left[P\left(\sup_{u \leq \Delta} w^{(x)}(u) \geq c\right) + \left(\inf_{u \leq \Delta} w^{(x)}(u) \leq d\right)\right].$$

Например, первая вероятность в квадратных скобках оценивается, как мы знаем (см. (16.2) и лемму (16.1)), значением

$$2P\left(w^{(x)}(\Delta) > c\right) = 2P\left(w(1) > \frac{c - x}{\sqrt{\Delta}}\right) \sim \frac{2}{\sqrt{2\pi z}} e^{-z^2/2}, \quad z = \frac{c - x}{\sqrt{\Delta}}.$$

При любом $x < c$ и любом $k > 0$ это есть $o(\Delta^k)$. Поэтому (53) доказано. Точно так же проверяются свойства 2), 3).

Далее, т. к. по формуле полной вероятности при $x \in (d, c)$

$$U(t + \Delta, x, B) = \int_d^c U(\Delta, x, dy) U(t, y, B),$$

то, пользуясь разложением вида (45) для функции U , точно так же, как в (46), получим

$$U(t + \Delta, x, B) - U(t, x, B) = \int U(\Delta, x, dy) [U(t, y, B) - U(t, x, B)] = \\ = a(x) \frac{\partial U}{\partial x} \Delta + \frac{b^2(x)}{2} \frac{\partial^2 U}{\partial x^2} \Delta + o(\Delta).$$

Отсюда следует существование $\partial U / \partial t$ и уравнение (43) для функции U .

Выполнение граничных и начальных условий очевидно. ■

Читатель может убедиться, что функция

$$u(t, x, y) = \frac{\partial}{\partial y} U(t, x, (-\infty, y)), \quad y \in (d, c),$$

играющая роль фундаментального решения для граничной задачи (51), (52) (функция u удовлетворяет (43) при граничных условиях (52) и начальном значении, вырождающемся в δ -функцию), равна

$$u(t, x, y) = \frac{1}{\sqrt{2\pi t}} \left[\sum_{k=-\infty}^{\infty} \exp \left\{ -\frac{[y + 2k(c-d)]^2}{2t} \right\} - \sum_{k=0}^{\infty} \exp \left\{ -\frac{[y - 2c - 2k(c-d)]^2}{2t} \right\} - \sum_{k=0}^{\infty} \exp \left\{ -\frac{[y - 2d - 2k(c-d)]^2}{2t} \right\} \right].$$

Этот ответ можно получить и непосредственно из вероятностных соображений (см., например, [20]).

Процессы с конечными моментами второго порядка, гауссовские процессы

§ 1. Процессы с конечными моментами второго порядка

Пусть $\{\xi(t), -\infty < t < \infty\}$ — случайный процесс, для которого существуют $a(t) = E\xi(t)$, $R(t, u) = E\xi(t)\xi(u)$. Так как от $\xi(t)$ всегда можно перейти к изучению процесса $\xi(t) - a(t)$ то, не ограничивая общности, можно считать, что $a(t) \equiv 0$.

Определение 1. Функция $R(t, u)$ называется *ковариационной функцией* процесса $\xi(t)$.

Определение 2. Функция $R(t, u)$ называется *неотрицательно (положительно) определенной*, если для любых k ; u_1, \dots, u_k ; $a_1, \dots, a_k \neq 0$

$$\sum a_i a_j R(u_i, u_j) \geq 0 \quad (> 0).$$

Очевидно, что ковариационная функция $R(t, u)$ является неотрицательно определенной, так как

$$\sum a_i a_j R(u_i, u_j) = E \left(\sum a_j \xi(u_j) \right)^2 \geq 0.$$

Определение 3. Процесс $\xi(t)$ называется *непредсказуемым*, если никакая линейная комбинация величин $\xi(u_1), \dots, \xi(u_k)$ не равна нулю с вероятностью 1, т. е. если не существует таких u_1, \dots, u_k ; a_1, \dots, a_k , что $P \left(\sum a_i \xi(u_i) = 0 \right) = 1$.

Если $R(t, u)$ — ковариационная функция непредсказуемого процесса, то $R(t, u)$ — положительно определенная функция. Ниже мы увидим, что в известном смысле верно и обратное утверждение.

Непредсказуемость означает, что мы не можем определить $\xi(t_k)$ как линейную комбинацию $\xi(t_j)$, $j < k$.

Пример 1. Процесс $\xi(t) = \sum_{k=1}^N \xi_k \varphi_k(t)$, где $\varphi_k(t)$ линейно независимы, ξ_k независимы, не является непредсказуемым, так как по значениям $\xi(t_1), \dots, \xi(t_N)$ можно определить $\xi(t)$ при всех остальных t .

Рассмотрим Гильбертово пространство L_2 всех случайных величин η на (Ω, \mathcal{F}, P) с конечными вторыми моментами, $E\eta = 0$, и со скалярным произведением $(\eta_1, \eta_2) = E\eta_1\eta_2$, соответствующим расстоянию $\|\eta_1 - \eta_2\| = [E(\eta_1 - \eta_2)^2]^{1/2}$. Сходимость в L_2 есть, очевидно, сходимость в среднеквадратичном.

Случайный процесс $\xi(t)$ можно представлять себе как кривую в L_2 .

Определение 4. Случайный процесс $\xi(t)$ называется *стационарным в широком смысле*, если функция $R(t, u) = R(t - u)$ зависит лишь от разности $t - u$. Функция $R(s)$ называется неотрицательно (положительно) определенной, если таковой оказывается функция $R(t, t + s)$. Для краткости процессы, стационарные в широком смысле, мы часто будем называть просто *стационарными*.

Для винеровского процесса $R(t, u) = Ew(t)w(u) = \min(t, u)$, так что процесс $w(t)$ не может быть стационарным. Однако процесс $\xi(t) = w(t + 1) - w(t)$ уже будет стационарным.

Очевидно, что для стационарного процесса функция $R(s)$ является четной, $E\xi^2(t) = R(0) = \text{const}$. Положим для простоты $R(0) = 1$. Тогда в силу неравенства Коши—Буняковского

$$|R(s)| = |E\xi(t)\xi(t+s)| \leq [E\xi^2(t)E\xi^2(t+s)]^{1/2} = R(0) = 1.$$

Теорема 1. 1) Процесс $\xi(t)$ непрерывен в среднеквадратичном ($\xi(t + \Delta) \xrightarrow{(2)} \xi(t)$ при $\Delta \rightarrow 0$) тогда и только тогда, когда функция $R(u)$ непрерывна в нуле.

2) Если функция $R(u)$ непрерывна в нуле, то она непрерывна всюду.

Доказательство.

$$1) \quad \|\xi(t + \Delta) - \xi(t)\|^2 = E(\xi(t + \Delta) - \xi(t))^2 = 2R(0) - 2R(\Delta).$$

$$\begin{aligned} 2) \quad R(t + \Delta) - R(t) &= E(\xi(t + \Delta)\xi(0) - \xi(t)\xi(0)) = \\ &= (\xi(0), \xi(t + \Delta) - \xi(t)) \leq \|\xi(t + \Delta) - \xi(t)\| = \\ &= \sqrt{2(R(0) - R(\Delta))}. \end{aligned} \quad (1)$$

В соответствии с содержанием гл. 6.17 процесс $\xi(t)$, непрерывный в среднеквадратичном, будет стохастически непрерывным. Однако непрерывность в среднеквадратичном не влечет за собой непрерывность траектории. Читатель может убедиться в этом, рассмотрев в качестве примера процесс

$$\xi(t) = \eta(t + 1) - \eta(t) - 1,$$

где $\eta(t)$ — пуассоновский процесс с параметром 1. Для него функция

$$R(t) = \begin{cases} 0 & \text{при } t \geq 1, \\ 1 - t & \text{при } 0 \leq t \leq 1 \end{cases}$$

непрерывна, хотя траектории $\xi(t)$ разрывны.

Если же

$$|R(\Delta) - R(0)| < c\Delta^{1+\varepsilon} \quad (2)$$

при некотором $\varepsilon > 0$, то по теореме Колмогорова для $\xi(t)$ существует непрерывная модификация. Отсюда следует, в частности, что если $R(t)$ дважды дифференцируема в точке $t = 0$, то траектории $\xi(t)$ можно считать непрерывными. Действительно, в этом случае в силу четности $R(t)$ выполняется $R'(0) = 0$, $R(\Delta) - R(0) \sim 1/2 R''(0)\Delta^2$.

В целом, чем выше гладкость ковариационной функции в нуле, тем более высокой гладкостью обладают траектории $\xi(t)$.

Предположим, что траектории процесса $\xi(t)$ измеримы (например, принадлежат пространству D).

Теорема 2 (простейшая эргодическая теорема). Если

$$R(s) \rightarrow 0 \quad \text{при} \quad s \rightarrow \infty, \quad (3)$$

то

$$\zeta_T = \frac{1}{T} \int_0^T \xi(t) dt \xrightarrow{(2)} 0.$$

Доказательство.

$$\|\zeta_T\|^2 = \frac{1}{T^2} \int_0^T \int_0^T R(t-u) dt du.$$

В силу четности $R(s)$

$$J \equiv \int_0^T \int_0^T R(t-u) dt du = 2 \int_0^T \int_u^T R(t-u) dt du.$$

Сделаем ортогональную замену переменных, положив $v = 1/\sqrt{2}(t-u)$, $s = 1/\sqrt{2} \times (t+u)$. Тогда

$$J \leq 2 \int_{s=0}^{T/\sqrt{2}} \int_{v=0}^{T/\sqrt{2}} R(v\sqrt{2}) dv ds \leq 2T \int_0^T R(v) dv,$$

$$\|\zeta_T\|^2 \leq \frac{2}{T} \int_0^T R(v) dv \rightarrow 0.$$

Пример 2. Стационарный процесс белого шума $\xi(t)$ определяется как процесс с независимыми значениями, т.е. как процесс, для которого при любых t_1, \dots, t_n значения $\xi(t_1), \dots, \xi(t_n)$ независимы. Для него

$$R(t) = \begin{cases} 1 & \text{при } t = 0, \\ 0 & \text{при } t \neq 0, \end{cases}$$

и условие (3) выполнено. Однако применять здесь теорему 2 нельзя, так как траектории $\xi(t)$ будут с вероятностью 1 неизмеримыми (неизмеримым с вероятностью 1 будет, например, множество $B = \{t: \xi(t) > 0\}$).

Определение 5. Процесс $\xi(t)$ называется *стационарным в узком смысле*, если при любых t_1, \dots, t_k распределение $(\xi(t_1+u), \xi(t_2+u), \dots, \xi(t_k+u))$ не зависит от u .

Очевидно, что если $\xi(t)$ — процесс, стационарный в узком смысле, то

$$E\xi(t)\xi(u) = E\xi(t-u)\xi(0) = R(t-u),$$

и $\xi(t)$ будет стационарным в широком смысле. Обратное, конечно, не верно. Однако существует класс процессов, для которых оба понятия стационарности совпадают.

§ 2. Гауссовские процессы

Определение 6. Процесс $\xi(t)$ называется *гауссовским*, если его конечномерные распределения нормальны.

Будем считать снова, что $E\xi(t) = 0$, $R(t, u) = E\xi(t)\xi(u)$.

Конечномерные распределения полностью определяются х. ф. $(\lambda = (\lambda_1, \dots, \lambda_k), \xi = (\xi(t_1), \dots, \xi(t_k)))$.

$$Ee^{i(\lambda, \xi)} = Ee^{i \sum \lambda_j \xi(t_j)} = e^{-\frac{1}{2} \lambda R \lambda^T}$$

где $R = \|R(t_i, t_j)\|$, индекс T означает транспонирование, так что

$$\lambda R \lambda^T = \sum \lambda_i \lambda_j R(t_i, t_j).$$

Таким образом, для гауссовского процесса конечномерные распределения полностью определяются ковариационной функцией $R(t, u)$.

Мы видели, что для непредсказуемого процесса $\xi(t)$ функция $R(t, u)$ положительно определена. Обратному утверждению можно придать следующую форму.

Теорема 3. Если функция $R(t, u)$ является положительно определенной, то существует непредсказуемый гауссовский процесс с ковариационной функцией $R(t, u)$.

Доказательство. Определим для произвольных t_1, \dots, t_k конечномерные распределения $\xi(t_1), \dots, \xi(t_k)$ через плотность

$$p_{t_1, \dots, t_k}(x_1, \dots, x_k) = \frac{\sqrt{|A|}}{(2\pi)^{k/2}} \exp \left\{ -\frac{1}{2} x A x^T \right\},$$

где A есть матрица, обратная к матрице вторых моментов $R = \|R(t_i, y_j)\|$ (см: § 7.6). Эти распределения будут, очевидно, согласованы, поскольку согласованными являются матрицы вторых моментов (матрица для $\xi(t_1), \dots, \xi(t_{k-1})$ является подматрицей матрицы R). Остается воспользоваться теоремой Колмогорова. ■

Пример 3. Пусть $w(t)$ — стандартный винеровский процесс. Процесс $w^0(t) = w(t) - tw(1)$, $t \in [0, 1]$ называется *броуновским мостом* (его «концы закреплены»: $w^0(0) = w^0(1) = 0$). Корреляционная функция для $w^0(t)$ равна

$$R(t, u) = E(w(t) - tw(1))(w(u) - uw(1)) = t(1 - u)$$

при $u \geq t$.

Гауссовский процесс $\xi(t)$, стационарный в широком смысле, является стационарным в узком смысле. Это немедленно вытекает из того, что для $R(t, u) = R(t - u)$ конечномерные распределения $\xi(t)$ становятся инвариантными относительно сдвига времени:

$$p_{t_1, \dots, t_k}(x_1, \dots, x_k) = p_{t_1 + u, \dots, t_k + u}(x_1, \dots, x_k),$$

поскольку $\|R(t_i + u, t_j + u)\| = \|R(t_i, t_j)\|$.

Если $\xi(t)$ — гауссовский процесс, то условия гладкости траекторий могут быть существенно ослаблены по сравнению с (2).

Пусть для простоты гауссовский процесс $\xi(t)$ стационарен.

Теорема 4. Если

$$|R(h) - R(0)| < c \left(\log \frac{1}{h} \right)^{-\alpha}, \quad \alpha > 3, \quad c < \infty,$$

то траектории $\xi(t)$ можно считать непрерывными.

Доказательство. Воспользуемся теоремой 17.2 и положим $\varepsilon(h) = (\log \frac{1}{h})^{-\beta}$ при $1 < \beta < (\alpha - 1)/2$ (имеется в виду \log по основанию 2). Тогда $\sum_{n=1}^{\infty} \varepsilon(2^{-n}) = \sum_{n=1}^{\infty} n^{-\beta} < \infty$, и в силу (1)

$$\begin{aligned} P(|\xi(t+h) - \xi(t)| > \varepsilon(h)) &= 2 \left[1 - \Phi \left(\frac{\varepsilon(h)}{\sqrt{2(1-R(h))}} \right) \right] \leq \\ &\leq 2 \left[1 - \Phi \left(c\varepsilon(h) \left(\log \frac{1}{h} \right)^{\alpha/2} \right) \right] = 2 \left[1 - \Phi \left(c \left(\log \frac{1}{h} \right)^{\alpha/2-\beta} \right) \right]. \end{aligned} \quad (4)$$

Так как аргумент под знаком Φ неограниченно растет при $h \rightarrow 0$, $\gamma = \alpha - 2\beta > 1$, то в силу (18.2)

$$1 - \Phi(x) \sim \frac{1}{\sqrt{2\pi}x} e^{-x^2/2} \quad \text{при } x \rightarrow \infty,$$

и правая часть в (4) не превосходит

$$q(h) = c_1 \left(\log \frac{1}{h} \right)^{\beta-\alpha/2} \exp \left\{ -c_2 \left(\log \frac{1}{h} \right)^{\alpha-2\beta} \right\},$$

так что

$$\sum_{n=1}^{\infty} 2^n q(2^{-n}) = c_1 \sum_{n=1}^{\infty} n^{-\gamma/2} \exp\{-c_2 n^\gamma + n \ln 2\} < \infty,$$

так как $c_2 > 0$, $\gamma > 1$. Условия теоремы 15.2 выполнены, теорема 4 доказана. ■

§ 3. Задача о прогнозе

Пусть известно распределение процесса $\xi(t)$ и известна траектория $\xi(t)$ на множестве $B \subset (-\infty, t]$ (B — либо интервал, либо конечный набор точек). Что можно сказать о значении $\xi(t+u)$? Нашей целью будет отыскание случайной величины ζ , измеримой относительно $\mathfrak{F}_B = \sigma(\xi(v), v \in B)$ (и называемой *прогнозом*) такой, что $E(\xi(t+u) - \zeta)^2$ принимает наименьшее возможное значение. Но ответ на этот вопрос известен (см. § 4.8)

$$\zeta = E(\xi(t+u) | \mathfrak{F}_B).$$

Пусть $\xi(t)$ — гауссовский процесс, $B = \{t_1, \dots, t_k\}$, $t_1 < t_2 < \dots < t_k < t_0 = t+u$, $A = (\sigma^2)^{-1} = \|a_{ij}\|$, $\sigma^2 = \|E\xi(t_i)\xi(t_j)\|_{i,j=1,\dots,k,0}$. Тогда распределение вектора $(\xi(t_1), \dots, \xi(t_0))$ имеет плотность

$$f(x_1, \dots, x_k, x_0) = \frac{\sqrt{|A|}}{(2\pi)^{(k+1)/2}} \exp \left\{ -\frac{1}{2} \sum x_i x_j a_{ij} \right\},$$

а условное распределение $\xi(t_0)$ относительно $\xi(t_1), \dots, \xi(t_k)$ имеет плотность, равную отношению

$$\frac{f(x_1, \dots, x_0)}{\int_{-\infty}^{\infty} f(x_1, \dots, x_0) dx_0}.$$

Экспоненциальная часть этого отношения имеет вид

$$\exp \left\{ -\frac{a_{00} x_0^2}{2} - \sum_{j=1}^k x_0 x_j a_{j0} \right\}.$$

Это означает, что рассматриваемое условное распределение есть Φ_{α, d^2} , где $\alpha = -\sum \frac{x_j a_{j0}}{a_{00}}$, $d^2 = 1/a_{00}$. Таким образом, в рассматриваемом случае наилучший прогноз ζ равен

$$\zeta = -\sum_{j=1}^k \frac{\xi(t_j) a_{0j}}{a_{00}}.$$

Среднеквадратическая ошибка этого прогноза равна $\sqrt{1/a_{00}}$.

Мы получили *линейный* прогноз. В общем случае это свойство линейности обычно нарушается.

Рассмотрим теперь задачу о наилучшем линейном прогнозе в случае произвольного процесса $\xi(t)$ с конечными моментами второго порядка. Для простоты будем считать опять, что $B = \{t_1, \dots, t_k\}$.

Обозначим $H(\xi)$ подпространство L_2 , порожденное случайными величинами $\xi(t)$, $-\infty < t < \infty$, и через $H_B(\xi)$ — подпространство $H(\xi)$, порожденное (натянутое на) $\xi(t_1), \dots, \xi(t_k)$. Элементы $H_B(\xi)$ имеют вид $\sum_{j=1}^k a_j \xi(t_j)$.

Существование и вид наилучшего линейного прогноза в этом случае устанавливаются следующим утверждением.

Теорема 5. Существует единственная точка $\zeta \in H_B(\xi)$ (проекция $\xi(t+u)$ на $H_B(\xi)$, рис. 20) такая, что

$$\xi(t+u) - \zeta \perp H_B(\xi). \quad (5)$$

Это условие эквивалентно тому, что

$$\|\xi(t+u) - \zeta\| = \min_{\theta \in H_B(\xi)} \|\xi(t+u) - \theta\|. \quad (6)$$

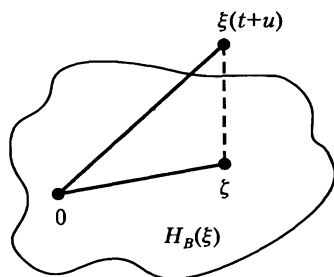


Рис. 20.

Явные формулы для коэффициентов a_j в представлении $\zeta = \sum a_j \xi(t_j)$ содержатся в доказательстве.

Доказательство. Соотношение (5) эквивалентно уравнениям

$$(\xi(t+u) - \zeta, \xi(t_j)) = 0, \quad j = 1, \dots, k.$$

Подставляя сюда $\zeta = \sum_{l=1}^k a_l \xi(t_l) \in H_B(\xi)$, получим

$$R(t+u, t_j) = \sum_{l=1}^k a_l R(t_j, t_l), \quad j = 1, \dots, k, \quad (7)$$

или, в векторной форме, $R_{t+u} = aR$, где $a = (a_1, \dots, a_k)$, $R_{t+u} = (R(t+u, t_1), \dots, R(t+u, t_k))$, $R = \|R(t_i, R_j)\|$.

Если процесс $\xi(t)$ непредсказуем, то матрица R невырождена и уравнение (7) однозначно разрешимо:

$$a = R_{t+u} R^{-1}. \quad (8)$$

Если $\xi(t)$ не является непредсказуемым, то либо R^{-1} все же существует, и тогда справедливо (8), либо R вырождена. В этом случае из совокупности $\xi(t_1), \dots, \xi(t_k)$ надо выделить лишь $l < k$ линейно независимых элементов, для которых все сказанное выше останется справедливым с заменой k на l .

Эквивалентность (5) и (6) вытекает из следующих соображений. Пусть θ — любой другой элемент $H_B(\xi)$. Тогда $\gamma = \theta - \zeta \in H_B(\xi)$, $\gamma \perp \xi(t+u) - \zeta$, так что

$$\|\xi(t+u) - \theta\| = \|\xi(t+u) - \zeta\| + \|\gamma\| \geq \|\xi(t+u) - \zeta\|.$$

Замечание 1. Может случиться (если процесс $\xi(t)$ не является непредсказуемым), что $\xi(t+u) \in H_B(\xi)$. Тогда погрешность прогноза ζ будет равна нулю. ■

Приложения

Приложение 1

Теорема о продолжении вероятностной меры

Пусть \mathcal{A} есть некоторая алгебра множеств из Ω , на которой задана вероятностная мера P , т.е. числовая функция, удовлетворяющая условиям P1–P3 гл. 2. Пусть \mathcal{P} означает класс всех подмножеств Ω . Для любого $A \in \mathcal{P}$ всегда существует последовательность $\{A_n\}_{n=1}^{\infty}$ непересекающихся множеств из \mathcal{A} такая, что $\bigcup_{n=1}^{\infty} A_n \supset A$ (достаточно взять $A_1 = \Omega$, $A_n = \emptyset$; $n \geq 2$). Обозначим $\gamma(A)$ класс всех таких последовательностей и введем на \mathcal{P} числовую функцию

$$P^*(A) = \inf \left\{ \sum_{n=1}^{\infty} P(A_n); \{A_n\} \in \gamma(A) \right\}.$$

Эта функция (внешняя мера на \mathcal{P} , индуцированная мерой P на \mathcal{A}) обладает свойствами:

- 1) $P^*(A) \leq P^*(B) \leq 1$, если $A \subset B$.
- 2) $P^*\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n)$, если множества $A_n \in \mathcal{A}$, $n = 1, 2, \dots$, и не пересекаются.
- 3) $P^*\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} P^*(A_n)$ для любых $A_1, A_2, \dots \in \mathcal{P}$.

Свойство 1) очевидно. Свойство 2) вытекает из следующих рассуждений. Пусть $\{B_n\}$ — любая последовательность из $\gamma(A)$, где $A = \bigcup_{n=1}^{\infty} A_n$. Так

как $\bigcup_{m=1}^{\infty} A_n B_m = A_n \in \mathcal{A}$, то $P(A_n) = \sum_{m=1}^{\infty} P(A_n B_m)$.

Следовательно,

$$\sum_{n=1}^{\infty} P(A_n) = \sum_n \sum_m P(A_n B_m) = \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} P(A_n B_m).$$

Но при каждом $N < \infty$

$$\sum_{n=1}^N P(A_n B_m) \leq P(B_m).$$

Значит, это неравенство верно и при $N = \infty$, и для любой $\{B_m\}_{m=1}^{\infty} \in \gamma(A)$

$$\sum_{n=1}^N P(A_n) \leq P(B_m).$$

Отсюда следует, что $P^*(A) \geq \sum_{n=1}^{\infty} P(A_n)$. Так как обратное неравенство очевидно, то $P^*(A) = \sum_{n=1}^{\infty} P(A_n)$.

Для доказательства свойства 3) рассмотрим при некотором $\varepsilon > 0$ последовательности $\{A_{nk}\}_{k=1}^{\infty} \in \gamma(A_n)$ такие, что

$$\sum_{k=1}^{\infty} P(A_{nk}) \leq P^*(A_n) + \frac{\varepsilon}{2^n}.$$

Последовательность множеств $\{A_{nk}\}_{n,k=1}^{\infty}$, очевидно, содержит $\bigcup A_n$, и, значит,

$$P^*\left(\bigcup A_n\right) \leq \sum_n \sum_k P(A_{nk}) \leq \sum_{n=1}^{\infty} P^*(A_n) + \varepsilon.$$

В силу произвольности ε свойство 3) доказано. ■

Введем теперь операцию симметрической разности \oplus над произвольными множествами A, B из \mathcal{P} с помощью равенства

$$A \oplus B = A\bar{B} \cup \bar{A}B.$$

Нетрудно заметить, что

$$\begin{aligned} A \oplus B &= B \oplus A = \bar{A} \oplus \bar{B} \subset A \cup B, \quad A \oplus A = \emptyset, \\ A \oplus \emptyset &= A, \quad (A \oplus B) \oplus C = A \oplus (B \oplus C). \end{aligned}$$

С помощью этой операции и функции P^* введем на \mathcal{P} расстояние ρ , положив для любых $A, B \in \mathcal{P}$

$$\rho(A, B) = P^*(A \oplus B).$$

Эта конструкция вполне аналогична той, которая использовалась в § 3.4 (мы рассматривали там расстояние $d(A, B) = P(A \oplus B)$ между измеримыми A и B). Свойства расстояния ρ те же, что и в (3.10). Нам понадобятся следующие из них:

- 1) $\rho(A, B) = \rho(B, A) \geq 0$, $\rho(A, A) = 0$,
- 2) $\rho(\bar{A}, \bar{B}) = \rho(A, B)$,
- 3) $\rho(AB, CD) \leq \rho(A, C) + \rho(B, D)$,
- 4) $\rho\left(\bigcup A_k, \bigcup B_k\right) \leq \sum_k \rho(A_k, B_k)$.

Можно отметить также, что

- 5) $|P^*(A) - P^*(B)| \leq \rho(A, B)$, и, следовательно, $P^*(\cdot)$ есть функция, равномерно непрерывная относительно ρ .

Свойства 1)–3) имеют место в (3.10) и доказываются совершенно аналогично на основании свойств меры P^* . Свойство 4) вытекает из свойства 3) меры P^* и соотношения $(A = \bigcup A_n, B = \bigcup B_n)$

$$A \oplus B \subset \bigcup (A_n \oplus B_n),$$

поскольку

$$\begin{aligned} A \oplus B &= \left[(\bigcup A_n) \cap (\bigcap \bar{B}_n)\right] \cup \left[(\bigcap \bar{A}_n) \cap (\bigcup B_n)\right] \subset \left[\bigcup A_n \bar{B}_n\right] \cup \left[\bigcup B_n \bar{A}_n\right] = \\ &= \bigcup (A_n \bar{B}_n \cup \bar{A}_n B_n) = \bigcup (A_n \oplus B_n). \end{aligned}$$

Свойство 5) вытекает из того, что

$$A \subset B \cup (A \oplus B), \quad B \subset A \cup (A \oplus B) \quad (1)$$

и, значит,

$$P^*(A) - P^*(B) \leq P^*(A \oplus B) = \rho(A, B),$$

$$P^*(B) - P^*(A) \leq P^*(A \oplus B) = \rho(A, B).$$

Аналогично принятому в § 3.4 назовем множество $A \in \mathcal{P}$ аппроксимируемым, если для него существует последовательность $A_n \in \mathcal{A}$, для которой $\rho(A, A_n) \rightarrow 0$. Совокупность всех аппроксимируемых множеств обозначим \mathfrak{A} . Это есть, очевидно, замыкание \mathcal{A} относительно ρ .

Лемма. \mathfrak{A} есть σ -алгебра.

Доказательство. Проверим, что \mathfrak{A} удовлетворяет свойствам σ -алгебры $A1, A2', A3$ гл. 2. Свойство $A1$: $\Omega \in \mathfrak{A}$ очевидно выполняется, так как $\mathcal{A} \subset \mathfrak{A}$. Свойство $A3$: $\bar{A} \in \mathfrak{A}$, если $A \in \mathfrak{A}$, следует из того, что для $A \in \mathfrak{A}$ существуют $A_n \in \mathcal{A}$ такие, что при $n \rightarrow \infty$

$$\rho(A, A_n) \rightarrow 0, \quad \rho(\bar{A}, \bar{A}_n) = \rho(A, A_n) \rightarrow 0.$$

Рассмотрим, наконец, свойство $A2'$. Покажем сначала, что если $A_n \in \mathcal{A}$, то $A = \bigcup A_n \in \mathfrak{A}$.

Действительно, не ограничивая общности, можно считать, что A_n не пересекаются. Тогда в силу свойств меры P^* для любого $\varepsilon > 0$ имеем

$$\sum P(A_k) \leq P^*(\Omega) = 1, \\ \rho\left(A, \bigcup_{k=1}^n A_k\right) = P^*\left(\bigcup_{k=n+1}^{\infty} A_k\right) = \sum_{k=n+1}^{\infty} P(A_k) < \varepsilon$$

при достаточно большом n .

Пусть теперь $A_n \in \mathfrak{A}$. Требуется показать, что $A = \bigcup_{n=1}^{\infty} A_n \in \mathfrak{A}$. Пусть $\{B_n\}$ — последовательность множеств из \mathcal{A} таких, что $\rho(A_n, B_n) < \varepsilon/2^n$. Тогда $B = \bigcup B_n \in \mathfrak{A}$ и в силу свойства 4) расстояния ρ

$$\rho(A, B) \leq \sum_{n=1}^{\infty} \rho(A_n, B_n) < \varepsilon.$$

Лемма доказана. ■

Теперь можно доказать основное утверждение.

Теорема 1*. Вероятность P может быть продолжена с алгебры \mathcal{A} на σ -алгебру \mathfrak{A} до некоторой вероятности \bar{P} .

* Теорема о продолжении меры на минимальную σ -алгебру, содержащую \mathcal{A} , была получена Каратеодори. Метризация нормированных алгебр Буля \mathcal{A} с помощью расстояния $\rho(A, B) = P(A \oplus B)$ использовалась многими авторами (см., например, доклад А. Н. Колмогорова на VI съезде польских математиков в 1948 г.; П. Халмош [23]). Использовать для продолжения меры свойства ее непрерывности относительно расстояния $\rho(A, B) = P^*(A \oplus B)$ предложил Л. Я. Савельев.

Доказательство. Положим для $A \in \mathfrak{A}$

$$\bar{P}(A) = P^*(A).$$

Очевидно, что $\bar{P}(A) = P(A)$ при $A \in \mathcal{A}$, $\bar{P}(\Omega) = 1$. Чтобы убедиться, что \bar{P} есть вероятность, надо доказать лишь счетную аддитивность \bar{P} . Докажем сначала конечную аддитивность. Ее достаточно доказать для двух множеств:

$$P^*(A + B) = P^*(A) + P^*(B), \quad (2)$$

где $A, B \in \mathfrak{A}$, $A \cap B = \emptyset$. Пусть $A_n \in \mathcal{A}$, $B_n \in \mathcal{A}$ таковы, что $\rho(A, A_n) \rightarrow 0$, $\rho(B, B_n) \rightarrow 0$ при $n \rightarrow \infty$. Тогда

$$|P^*(A + B) - P^*(A_n + B_n)| \leq \rho(A + B, A_n + B_n) \leq \rho(A, A_n) + \rho(B, B_n) \rightarrow 0, \\ P^*(A_n + B_n) = P(A_n + B_n) = P(A_n) + P(B_n) - P(A_n B_n). \quad (3)$$

Здесь

$$P(A_n) \rightarrow P^*(A), \quad P(B_n) \rightarrow P^*(B), \\ P(A_n B_n) \leq P^*(A_n B) + P^*(B_n \bar{B}) \leq P^*(A_n \bar{A}) + P^*(B_n \bar{B}) \leq \rho(A, A_n) + \rho(B, B_n) \rightarrow 0.$$

Значит, из (3) следует (2).

Докажем теперь счетную аддитивность. Пусть $A_n \in \mathfrak{A}$ не пересекаются. Тогда, полагая $A_n = \bigcup_{k=1}^n A_k$, из конечной аддитивности \bar{P} получим

$$\bar{P}(A) = \sum_{k=1}^n \bar{P}(A_k) + \bar{P}\left(\bigcup_{k=n+1}^{\infty} A_k\right).$$

Следовательно,

$$\bar{P}(A) \geq \sum_{k=1}^{\infty} \bar{P}(A_k).$$

С другой стороны,

$$\bar{P}(A) = P^*(A) \leq \sum_{k=1}^{\infty} P^*(A_k) = \sum_{k=1}^{\infty} \bar{P}(A_k).$$

Теорема 2. Продолжение вероятности P с алгебры \mathcal{A} на σ -алгебру \mathfrak{A} единственно.

Доказательство. Предположим, что на \mathfrak{A} существует еще одна вероятность P_1 , совпадающая с P на \mathcal{A} и такая, что для некоторого $A \in \mathfrak{A}$

$$P_1(A) \neq \bar{P}(A).$$

Допустим сначала, что $\varepsilon = P_1(A) - \bar{P}(A) > 0$. Рассмотрим последовательность $\{B_n\} \subset \gamma(A)$ такую, что

$$\sum_{n=1}^{\infty} P(B_n) - \bar{P}(A) < \frac{\varepsilon}{2}.$$

Тогда $P_1(A) = \bar{P}(A) + \varepsilon \geq \sum_{n=1}^{\infty} P(B_n) + \varepsilon/2$, что противоречит принадлежности $A \subset \bigcup_{n=1}^{\infty} B_n$. Значит,

$$P_1(A) \leq \bar{P}(A), \quad A \in \mathfrak{A}.$$

Так как \bar{P} ρ -непрерывна в точке \emptyset , то отсюда следует, что и P_1 ρ -непрерывна в точке \emptyset , а вместе с этим и в любой «точке» $A \in \mathfrak{A}$. Действительно, в силу (1)

$$|P_1(A) - P_1(B)| \leq P_1(A \oplus B) \leq \bar{P}(A \oplus B) \rightarrow 0,$$

если только $\rho(A, B) = \bar{P}(A \oplus B) \rightarrow 0$. Поэтому для $A \in \mathfrak{A}$

$$\bar{P}(A) = \lim_{\substack{B \rightarrow A \\ B \in \mathfrak{A}}} P(B) = \lim_{\substack{B \rightarrow A \\ B \in \mathfrak{A}}} P_1(B) = P_1(A).$$

Пусть $\mathfrak{A}^* = \sigma(\mathfrak{A})$ есть σ -алгебра, порожденная \mathfrak{A} . Так как $\mathfrak{A} \subset \mathfrak{A}^*$, то $\mathfrak{A}^* \subset \mathfrak{A}$, и из полученных утверждений очевидным образом вытекает

Следствие 1. Вероятность P может быть единственным образом продолжена с алгебры \mathfrak{A} на минимальную σ -алгебру \mathfrak{A}^* , порожденную \mathfrak{A} .

Замечание 1. σ -алгебра \mathfrak{A} , определенная выше как замыкание алгебры \mathfrak{A} относительно введенного расстояния ρ , оказывается во многих случаях шире σ -алгебры $\mathfrak{A}^* = \sigma(\mathfrak{A})$, порожденной \mathfrak{A} . Это обстоятельство тесно связано с понятием *пополнения меры*. Речь идет вот о чем. Пусть $\mathfrak{A} = \mathfrak{F}$ с самого начала есть σ -алгебра. Тогда меру \bar{P} в теореме 1 можно построить весьма просто. Для этого мы продолжим меру P с $\langle \Omega, \mathfrak{F} \rangle$ на более широкую, чем \mathfrak{F} , σ -алгебру, построенную следующим образом. Мы будем говорить, что множество N из Ω принадлежит классу \mathcal{N} , если для N найдется такое $A = A(N) \in \mathfrak{F}$, что $N \subset A$, $P(A) = 0$. Нетрудно видеть, что класс множеств вида $B + N$, где $B \in \mathfrak{F}$, $N \in \mathcal{N}$, снова образует σ -алгебру. Обозначим ее $\mathfrak{F}_{\mathcal{N}}$. Положив $P(B + N) = P(B)$, мы получим продолжение меры P на $\langle \Omega, \mathfrak{F}_{\mathcal{N}} \rangle$. Такая мера называется *полной*, а проведенная операция — *пополнением меры* P .

Теперь можно установить, что построенная в теореме 1 мера \bar{P} является *полной*, а σ -алгебра \mathfrak{A} совпадает с $\mathfrak{F}_{\mathcal{N}}$.

Если, например, $\Omega = [0, 1]$, а \mathfrak{A} есть алгебра, порожденная интервалами, то $\mathfrak{A}^* = \sigma(\mathfrak{A})$ будет, как мы уже знаем, борелевской σ -алгеброй, а \mathfrak{A} будет лебеговским расширением \mathfrak{A}^* , состоящим из множеств, «измеримых по Лебегу».

Теорема Колмогорова о согласованных распределениях

Пусть T — некоторое множество индексов t , \mathbb{R}_t при каждом $t \in T$ есть вещественная прямая $(-\infty, \infty)$. Пусть $N \in T$ — некоторое конечное подмножество T . Тогда произведение пространств $\prod_{t \in N} \mathbb{R}_t = \mathbb{R}^N$ есть евклидово пространство размерности, равной числу n элементов в N , построенное на n осях пространства $\mathbb{R}^T = \prod_{t \in N} \mathbb{R}_t$.

Предположим, что для любого конечного подмножества $N \subset T$ на $(\mathbb{R}^N, \mathfrak{B}^N)$, где \mathfrak{B} — σ -алгебра борелевских множеств в \mathbb{R}^N , задана вероятностная мера P_N . Тем самым на \mathbb{R}^T задано семейство мер. Это семейство называется *согласованным*, если для любых $L \subset N$ и борелевских множеств B из \mathbb{R}^L

$$P_L(B) = P_N(B \times \mathbb{R}^{N-L}),$$

Мера P_L называется *проекцией* P_N на \mathbb{R}^L . Множество из \mathbb{R}^T , которое можно представить в виде $B \times \mathbb{R}^{T-N}$, где $B \in \mathfrak{B}^N$, а N — конечное множество, называется *цилиндрическим* в \mathbb{R}^T . Множество B называется *основанием* цилиндра.

Обозначим \mathfrak{B}^T σ -алгебру множеств из \mathbb{R}^T , порожденную цилиндрическими множествами.

Теорема. Если на \mathbb{R}^T задано согласованное семейство вероятностных мер, то на $(\mathbb{R}^T, \mathfrak{B}^T)$ существует единственная вероятностная мера P такая, что при любом N P_N совпадает с проекцией P на \mathbb{R}^N .

Доказательство. Цилиндрические множества в \mathbb{R}^T образуют алгебру. Покажем, что при $B \in \mathfrak{B}^N$ соотношения

$$P(B \times \mathbb{R}^{T-N}) = P_N(B)$$

определяют меру на этой алгебре. Прежде всего, в силу согласованности мер P_N это определение вероятности цилиндрических множеств корректно (имеются в виду случаи, когда $B = B_1 \times \mathbb{R}^{N-L}$, $B_1 \in \mathfrak{B}^L$. Тогда левая часть будет равна также $P(B_1 \times \mathbb{R}^{T-L})$). Далее, введенная вероятность аддитивна. Действительно, пусть $B_1 \times \mathbb{R}^{T-N_1}$ и $B_2 \times \mathbb{R}^{T-N_2}$ — два непересекающихся цилиндрических множества. Тогда, обозначив $N = N_1 \cup N_2$, будем иметь

$$\begin{aligned} P(B_1 \times \mathbb{R}^{T-N_1} \cup B_2 \times \mathbb{R}^{T-N_2}) &= \\ &= P(\{B_1 \times \mathbb{R}^{N-N_1} \cup B_2 \times \mathbb{R}^{N-N_2}\} \times \mathbb{R}^{T-N}) = P_N(\{B_1 \times \mathbb{R}^{N-N_1} \cup B_2 \times \mathbb{R}^{N-N_2}\}) = \\ &= P_N(B_1 \times \mathbb{R}^{N-N_1}) + P_N(B_2 \times \mathbb{R}^{N-N_2}) = P(B_1 \times \mathbb{R}^{T-N_1}) + P(B_2 \times \mathbb{R}^{T-N_2}). \end{aligned}$$

Чтобы проверить счетную аддитивность P , воспользуемся эквивалентностью свойств $P3$ и $P3'$ (см. гл. 2). В силу этой эквивалентности достаточно показать, что если \mathcal{B}_n , $n = 1, 2, \dots$ — убывающая последовательность цилиндрических множеств, то соотношения $P(\mathcal{B}_n) > \varepsilon$, $n = 1, 2, \dots$, при некотором $\varepsilon > 0$ означают, что $\mathcal{B} = \bigcap_{n=1}^{\infty} \mathcal{B}_n$ непусто. Вложенность множеств \mathcal{B}_n означает, что в представлениях $\mathcal{B}_n = B_n \times \mathbb{R}^{T-N_n}$ выполняется $N_n \subset N_{n+1}$, $B_{n+1} \cap \mathbb{R}^{N_n} \subset B_n$. Не ограничивая общности, мы будем считать в дальнейшем, что число элементов множества $N_n = \{t_1, \dots, t_n\}$ равно n , и через x_i (с различными верхними индексами) будем обозначать координаты пространства \mathbb{R}_i .

Итак, пусть $P(\mathcal{B}_n) = P_{N_n}(B_n) \geq \varepsilon > 0$. Докажем, что $\mathcal{B} = \bigcap_{n=1}^{\infty} \mathcal{B}_k$ непусто.

Для каждого борелевского множества $B_n \subset \mathbb{R}^{N_n}$ существует компакт K_n такой, что

$$K_n \subset B_n, \quad P_{N_n}(B_n - K_n) < \frac{\varepsilon}{2^{n+1}}.$$

Если обозначить $\mathcal{K}_n = K_n \times \mathbb{R}^{T-N_n}$, то мы получим

$$P(\mathcal{B}_n - \mathcal{K}_n) = P_{N_n}(B_n - K_n) < \frac{\varepsilon}{2^{n+1}}.$$

Введем в рассмотрение множества $\mathcal{D}_n = \bigcap_{k=1}^n \mathcal{K}_k$. Легко видеть, что $\mathcal{D}_n \subset \mathcal{B}_n$ —

тоже цилиндры. Так как $\mathcal{B}_n - \bigcap_{k=1}^n \mathcal{K}_k \subset \bigcap_{k=1}^n (\mathcal{B}_k - \mathcal{K}_k)$, то для них

$$P(\mathcal{B}_n - \mathcal{D}_n) \leq P\left(\bigcap_{k=1}^n (\mathcal{B}_k - \mathcal{K}_k)\right) \leq \sum_{k=1}^n P(\mathcal{B}_k - \mathcal{K}_k) \leq \frac{\varepsilon}{2};$$

$$P(\mathcal{D}_n) \geq P(\mathcal{B}_n) - \frac{\varepsilon}{2} \geq \frac{\varepsilon}{2}.$$

Из последнего соотношения следует, что \mathcal{D}_n есть убывающая последовательность непустых цилиндрических множеств. Обозначим $X^n = (x_1^n, x_2^n, \dots, x_n^n)$ некоторую точку основания $D_n = \bigcap_{k=1}^n K_k \times \mathbb{R}^{N_n - N_k}$ цилиндра \mathcal{D}_n , определяющую цилиндрическое множество \mathcal{D}_n в \mathbb{R}^T . Так как \mathcal{D}_n убывают, то $(x_1^{n+p}, x_2^{n+p}, \dots, x_n^{n+p}) \in K_n$ при любом $p \geq 0$. В силу компактности K_n можно выбрать последовательность n_{1k} такую, что $x_1^{n_{1k}} \rightarrow x_1$ при $k \rightarrow \infty$. Из этой последовательности можно выбрать подпоследовательность n_{2k} такую, что $x_2^{n_{2k}} \rightarrow x_2$ и т.д.

Рассмотрим теперь диагональную последовательность точек (точнее, цилиндрических множеств) $X^{n_{kk}} = (x_1^{n_{kk}}, x_2^{n_{kk}}, \dots, x_{n_{kk}}^{n_{kk}})$. Ясно, что при $k \rightarrow \infty$ $X^{n_{kk}} \rightarrow X = (x_1, x_2, \dots)$ (имеется в виду покоординатная сходимость). При этом

$$(x_1^{n_{kk}}, x_2^{n_{kk}}, \dots, x_m^{n_{kk}}) \rightarrow (x_1, \dots, x_m) \in K_m$$

при любом m . Это означает, что соответствующее точке X множество $\mathcal{X} = \{y(t) \in \mathbb{R}^T : y(t_1) = x_1, y(t_2) = x_2, \dots\} \in K_m \in \mathcal{B}_m$ при любом m и, значит, $\mathcal{X} \in \bigcap_{m=1}^{\infty} \mathcal{B}_m$.

Таким образом, \mathcal{B} не пусто, и счетная аддитивность P на алгебре цилиндрических множеств доказана. Следовательно, P есть мера, и остается воспользоваться теоремой о продолжении меры с алгебры на σ -алгебру, порожденную этой алгеброй. ■

Приложение 3

Интегрирование

§ 1. Пространство с мерой

Пусть $\langle \Omega, \mathfrak{F} \rangle$ — измеримое пространство. Мы будем говорить, что задано *пространство с мерой* $\langle \Omega, \mathfrak{F}, \mu \rangle$, если на \mathfrak{F} задана неотрицательная счетно-аддитивная функция μ , т. е. функция, обладающая следующими свойствами:

1) $\mu(\bigcup_j A_j) = \sum_j \mu(A_j)$ для любого счетного набора непересекающихся множеств $A_j \in \mathfrak{F}$ (σ -аддитивность).

2) $\mu(A) \geq 0$ для каждого $A \in \mathfrak{F}$.

3) $\mu(\emptyset) = 0$ где \emptyset — пустое множество.

Значение $\mu(A)$ называется *мерой* множества A . Мы будем рассматривать только *конечные* и *σ -конечные* меры. В первом случае предполагается, что $\mu(\Omega) < \infty$. Во втором — что существует разбиение Ω на счетное число подмножеств A_j таких, что $\mu(A_j) < \infty$.

Вероятностное пространство является примером пространства с конечной (единичной) мерой. Пространство $\langle \mathbb{R}, \mathfrak{B}, \mu \rangle$, где \mathbb{R} — вещественная прямая, \mathfrak{B} — σ -алгебра борелевских множеств, а μ — мера Лебега, представляет собой пример пространства с σ -конечной мерой.

Можно рассматривать также функции множеств $\mu(A)$, удовлетворяющие лишь условиям 1), 3) и не обязательно неотрицательные. Такие функции называются *обобщенными мерами*. Любая конечная обобщенная мера (т. е. такая, что $\sup_A \mu(A) < \infty$, $\inf_A \mu(A) > -\infty$) может быть представлена в виде разности двух конечных неотрицательных мер (теорема Хана о разложении, см § 5 этого Приложения). Обобщенные меры нам понадобятся лишь в § 5. В остальных местах и при отсутствии специальных оговорок мы везде под мерами будем понимать функции множеств, удовлетворяющие условиям 1)–3).

Так же, как при изучении простейших свойств вероятности (§ 2.2), легко устанавливаются следующие свойства меры:

1) $\mu(A) \leq \mu(B)$, если $A \subset B$.

2) $\mu(\bigcup_j A_j) \leq \sum \mu(A_j)$ любых A_j .

3) Если $A_n \subset A_{n+1}$, $\bigcup A_n = A$, то $\mu(A_n) \rightarrow \mu(A)$, или, что то же самое,

3') если $A_n \supset A_{n+1}$, $\bigcap A_n = A$, $\mu(A_1) < \infty$, то $\mu(A_n) \rightarrow \mu(A)$.

Рассмотрим далее на $\langle \Omega, \mathfrak{F} \rangle$ измеримые функции, т. е. функции $\xi(\omega)$, обладающие свойством $\{\omega: \xi(\omega) \in B\} \in \mathfrak{F}$ для любого борелевского множества B на прямой.

Аналогично тому, как это делалось для вероятностной меры, вводятся понятия *сходимости по мере* и *сходимости почти всюду*.

Будем говорить, что последовательность ξ_n сходится к ξ почти всюду (п. в.) $\xi_n \xrightarrow{\text{п. в.}} \xi$, если $\xi_n(\omega) \rightarrow \xi(\omega)$ для всех ω , за исключением множества меры 0.

Будем говорить, что ξ_n сходится к ξ по мере: $\xi_n \xrightarrow{\mu} \xi$, если для каждого $\varepsilon > 0$ при $n \rightarrow \infty$

$$\mu(\{|\xi_n - \xi| > \varepsilon\}) \rightarrow 0.$$

Займемся теперь построением и изучением свойств интегралов. Рассмотрим сначала конечные меры и будем считать их, не ограничивая общности, вероятностными. В этом случае мы будем вместо $\mu(A)$ писать $P(A)$.

К интегралам по произвольным σ -конечным мерам вернемся в § 4 этого раздела.

§ 2. Интеграл по вероятностной мере

1. Интегралы от простых функций.

Измеримая функция $\xi(\omega)$ называется *простой*, если множество ее значений конечно. Индикатором множества $F \in \mathfrak{F}$ называется простая функция

$$I_F(\omega) = \begin{cases} 1, & \text{если } \omega \in F, \\ 0, & \text{если } \omega \notin F. \end{cases}$$

Очевидно, что любую простую функцию $\xi(\omega)$ можно записать в виде

$$\xi(\omega) = \sum_{k=1}^n x_k I_{F_k}(\omega),$$

где x_k , $k = 1, 2, \dots, n$, — значения, принимаемые ξ , а $F_k = \{\omega: \xi(\omega) = x_k\}$. Множества F_k попарно не пересекаются и $\bigcup_{k=1}^n F_k = \Omega$. Интегралом от простой случайной величины $\xi(\omega)$ называется число

$$\int \xi dP = \int \xi(\omega) dP(\omega) = \sum_{k=1}^n x_k P(F_k) = E\xi.$$

Интегралом по множеству $A \in \mathfrak{F}$ от простой измеримой функции $\xi(\omega)$ называется

$$\int_A \xi dP = \int \xi(\omega) I_A(\omega) dP(\omega).$$

Корректность этих определений (разбиения на множества F_k могут быть различны) проверяется очевидным образом.

2. Определение интегралов от произвольных функций.

Лемма 1. Пусть $\xi(\omega) \geq 0$. Существует последовательность $\xi_n(\omega)$ измеримых простых функций такая, что $\xi_n(\omega) \uparrow \xi(\omega)$ при $n \rightarrow \infty$.

Доказательство. Разобьем отрезок $[0, n]$ на $n \cdot 2^n$ равных отрезков. Пусть $x_0 = 0$, $x_1, \dots, x_{n \cdot 2^n} = n$ означают точки деления, так что $x_{i+1} - x_i = 1/2^n$. Положим $F_i = \{\omega: x_i \leq \xi(\omega) < x_{i+1}\}$, $i = 1, 2, \dots, n \cdot 2^n - 1$, $F_0 = \{0 \leq \xi(\omega) < x_1\} \cup \{\xi(\omega) \geq n\}$, $\xi_n(\omega) = \sum_{i=0}^{n \cdot 2^n - 1} x_i I_{F_i}(\omega) \leq \xi(\omega)$. $\xi_n(\omega)$ есть, очевидно, измеримая простая функция,

обладающая тем свойством, что в точке $\omega \in \Omega$ при $n > \xi(\omega)$ выполняется неравенство

$$0 \leq \xi(\omega) - \xi_n(\omega) \leq \frac{1}{2^n}.$$

Лемма 2. Пусть $\xi_n \uparrow \xi \geq 0$, $\eta_n \uparrow \xi \geq 0$ — последовательности простых случайных величин. Тогда

$$\lim_{n \rightarrow \infty} \int \xi_n dP = \lim_{n \rightarrow \infty} \int \eta_n dP.$$

Доказательство. Убедимся, что для любого m

$$\int \xi_m dP \leq \lim_{n \rightarrow \infty} \int \eta_n dP.$$

Функция ξ_m простая. Следовательно, она с вероятностью 1 ограничена некоторой постоянной: $\xi_m \leq c_m$. Поэтому при любых n , $\varepsilon > 0$

$$\xi_m - \eta_n \leq c_m \cdot I_{\{\xi_m \geq \eta_n + \varepsilon\}} + \varepsilon.$$

Отсюда следует, что

$$E\xi_m \leq c_m P\{\xi_m \geq \eta_n + \varepsilon\} + \varepsilon + E\eta_n.$$

Вероятность в правой части при $n \rightarrow \infty$ мала:

$$P\{\xi_m \geq \eta_n + \varepsilon\} \leq P\{\xi \geq \eta_n + \varepsilon\} \rightarrow 0,$$

так как η_n почти наверное (и, следовательно, по вероятности) сходится к ξ . Поэтому $E\xi_m \leq \varepsilon + \lim_{n \rightarrow \infty} E\eta_n$. Так как ε произвольно, то

$$\lim_{n \rightarrow \infty} E\xi_n \leq \lim_{n \rightarrow \infty} E\eta_n.$$

Меняя местами $\{\xi_n\}$ и $\{\eta_n\}$, получим обратное неравенство. ■

Утверждения лемм 1, 2 делают корректными следующие определения

Интегралом неотрицательной функции $\xi(\omega)$ называется число

$$\int \xi dP = \lim_{n \rightarrow \infty} \int \xi_n dP, \quad (1)$$

где ξ_n — последовательность простых измеримых функций таких, что $\xi_n \uparrow \xi$ при $n \rightarrow \infty$.

Интеграл $\int \xi dP$ будем обозначать также через $E\xi$. Мы будем говорить, что интеграл $\int \xi dP$ существует, а ξ интегрируема, если $E\xi < \infty$.

Интегралом от произвольной (принимającej значения разных знаков) функции $\xi(\omega)$ называется число

$$E\xi = E\xi^+ - E\xi^-, \quad \xi^\pm = \max(0, \pm \xi),$$

если хотя бы одно из значений $E\xi^\pm$ конечно. В противном случае $E\xi$ не определено. $E\xi$ существует тогда и только тогда, когда существует $E|\xi| < \infty$ (ведь $|\xi| = \xi^+ + \xi^-$). Если существует $E\xi$, то существует также $E(\xi; A) = \int_A \xi dP = E\xi I_A$ для любого $A \in \mathfrak{F}$.

Лемма 3. Если $E\xi$ существует и $F_n \in \mathfrak{F}$ есть последовательность множеств такая, что $P(F_n) \rightarrow 0$ при $n \rightarrow \infty$, то

$$E(\xi; F_n) \rightarrow 0.$$

Доказательство. Для любой последовательности $|\xi_m| \uparrow |\xi|$ простых функций справедливо ($A_m = \{|\xi| \leq m\}$)

$$E|\xi| \geq \lim_{m \rightarrow \infty} E|\xi|I_{A_m} \geq \lim_{m \rightarrow \infty} E|\xi_m|I_{A_m} = E|\xi|,$$

поскольку $|\xi_m|I_{A_m} \uparrow |\xi|$. Отсюда следует, что

$$E|\xi| = \lim_{m \rightarrow \infty} E|\xi|I_{A_m} = \lim_{m \rightarrow \infty} E(|\xi|; |\xi| \leq m)$$

и, значит, для любого $\varepsilon > 0$ существует $m(\varepsilon)$ такое, что

$$E|\xi| - E(|\xi|; |\xi| \leq m) < \varepsilon$$

при $m \geq m(\varepsilon)$. Поэтому для таких m

$$\begin{aligned} E(|\xi|; F_n) &= E(|\xi|; \{|\xi| \leq m\} F_n) + E(|\xi|; \{|\xi| > m\} F_n) \leq mP(F_n) + \varepsilon, \\ \lim_{n \rightarrow \infty} E(|\xi|; F_n) &\leq \varepsilon. \end{aligned}$$

Отметим, что в лемме 6.2 содержится некоторое обобщение леммы 3.

3. Свойства интегралов.

11. Если множества $A_j \in \mathfrak{F}$ не пересекаются и $\bigcup_j A_j = \Omega$, то

$$\int \xi dP = \sum_j \int_{A_j} \xi dP. \quad (2)$$

Достаточно доказать это соотношение для $\xi(\omega) \geq 0$. Для простых функций равенство (2) очевидно, так как

$$\int \xi dP = \sum_k x_k P(\xi = x_k) = \sum_j \sum_k x_k P(\xi = x_k; A_j).$$

В общем случае, используя определение (1), получим

$$\int \xi dP = \lim_{n \rightarrow \infty} \int \xi_n dP = \lim_{n \rightarrow \infty} \sum_j \int_{A_j} \xi_n dP = \sum_j \lim_{n \rightarrow \infty} \int_{A_j} \xi_n dP = \sum_j \int_{A_j} \xi dP. \quad (3)$$

Изменение порядка предельного перехода и суммирования здесь законно, так как в силу леммы 3

$$\sum_{j=N}^{\infty} \int_{A_j} \xi_n dP = E\left(\xi_n; \bigcup_{j=N}^{\infty} A_j\right) \leq E\left(\xi; \bigcup_{j=N}^{\infty} A_j\right) \rightarrow 0$$

при $N \rightarrow \infty$ равномерно по n .

$$12. \int (\xi + \eta) dP = \int \xi dP + \int \eta dP.$$

Для простых функций это свойство очевидно. Поэтому для $\xi \geq 0$, $\eta \geq 0$ это свойство следует из аддитивности операции предельного перехода.

В общем случае находим (ξ^{\pm} и η^{\pm} определяются, как и раньше)

$$\begin{aligned} \int (\xi + \eta) dP &= \int (\xi^+ + \eta^+) dP - \int (\xi^- + \eta^-) dP = \\ &= \int \xi^+ dP - \int \xi^- dP + \int \eta^+ dP - \int \eta^- dP = \int \xi dP + \int \eta dP. \end{aligned}$$

13. Если a — произвольная постоянная, то

$$\int a\xi \, dP = a \int \xi \, dP.$$

14. Если $\xi \leq \eta$, то $\int \xi \, dP \leq \int \eta \, dP$.

Доказательство свойств 13, 14 очевидно. Так как $\int \xi \, dP = E\xi$, то свойства 11–14 можно записать с помощью математических ожиданий следующим образом:

11. $E\xi = \sum_j E(\xi; A_j)$, если A_j не пересекаются, $\bigcup A_j = \Omega$.

12. $E(\xi + \eta) = E\xi + E\eta$.

13. $Ea\xi = aE\xi$.

14. $E\xi \leq E\eta$, если $\xi \leq \eta$.

Отметим также следующие свойства интегралов, легко вытекающие из свойств 11–14.

15. $|E\xi| \leq E|\xi|$.

16. Если $c_1 \leq \xi \leq c_2$, то $c_1 \leq E\xi \leq c_2$.

17. Если $\xi \geq 0$ и $E\xi = 0$, то $P(\xi = 0) = 1$.

Это следует из неравенства Чебышева: $P(\xi \geq \varepsilon) \leq E\xi/\varepsilon = 0$ при любом $\varepsilon > 0$.

18. Если $P(\xi = \eta) = 1$ и $E\xi$ существует, то $E\xi = E\eta$.

Действительно,

$$E\eta = \lim_{n \rightarrow \infty} E(\eta; |\eta| \leq n) = \lim_{n \rightarrow \infty} E(\xi; |\xi| \leq n) = E\xi.$$

§ 3. Дальнейшие свойства интегралов

1. Теоремы сходимости.

Ряд теорем сходимости был доказан в § 6.1 и, в частности, теорема о сходимости мажорируемой последовательности:

Если $\xi_n \rightarrow \xi$, $|\xi_n| \leq \eta$, $E\eta < \infty$, то существует $E\xi$ и $E\xi_n \rightarrow E\xi$.

Приведем некоторые другие полезные утверждения о сходимости интегралов.

Теорема 1 (Теорема о монотонной сходимости). Если $0 \leq \xi_n \uparrow \xi$, то $E\xi = \lim_{n \rightarrow \infty} E\xi_n$.

Доказательство. В дополнение к следствию 6.4 нам надо здесь доказать лишь, что $E\xi_n \rightarrow \infty$, если $E\xi = \infty$. Положим $\xi_n^N = \min(\xi_n, N)$, $\xi^N = \min(\xi, N)$. Тогда, очевидно, $\xi_n^N \uparrow \xi^N$ при $n \rightarrow \infty$, $E\xi_n^N \uparrow E\xi^N$. Следовательно, значение $E\xi_n^N \leq E\xi_n$ выбором n и N может быть сколь угодно большим. ■

Эти теоремы можно обобщить следующим образом. Для того, чтобы в § 4 было удобно получить распространение теорем сходимости на интегралы по произвольной мере, мы $E\xi$ будем писать теперь в форме интеграла $\int \xi \, dP$.

Теорема 2 (Фату—Лебега). Пусть η и ζ интегрируемы. Тогда если $\xi_n \leq \eta$, то

$$\limsup \int \xi_n dP \leq \int \limsup \xi_n dP. \quad (4)$$

Если $\xi_n \geq \zeta$, то

$$\liminf \int \xi_n dP \geq \int \liminf \xi_n dP. \quad (5)$$

Если $\xi_n \uparrow \xi$, $\xi_n \geq \zeta$ или $\xi_n \xrightarrow{\text{п.в.}} \xi$, $\zeta \leq \xi_n \leq \eta$, то

$$\lim \int \xi_n dP = \int \xi dP. \quad (6)$$

Доказательство. Докажем, например, (5), положив, не ограничивая общности, $\zeta \equiv 0$. В этом случае

$$\xi \geq \eta_n \equiv \inf_{k \geq n} \xi_k \uparrow \liminf \xi_n, \quad \eta_n \geq 0,$$

и по теореме о монотонной сходимости

$$\liminf \int \xi_n dP \geq \lim \int \eta_n dP = \int \liminf \xi_n dP.$$

Если (5) применить к последовательности $\eta - \xi_n$, то получим (4); (6) следует из предыдущих теорем. ■

2. Связь с интегрированием по мере на прямой.

Пусть $g(x)$ — некоторая борелевская функция, заданная на прямой \mathbb{R} (если \mathfrak{B} — σ -алгебра борелевских множеств на прямой и $B \in \mathfrak{B}$, то $\{x: g(x) \in B\} \in \mathfrak{B}$). Тогда $\eta = g(\xi(\omega))$, очевидно, также будет случайной величиной. Как мы видели в § 3.2, случайная величина ξ индуцирует вероятностное пространство $(\mathbb{R}, \mathfrak{B}, P_\xi)$ с мерой P_ξ на прямой такое, что $P_\xi(B) = P(\xi \in B)$. Следовательно, можно говорить об интегралах по этой мере. Справедлива

Теорема 3. Если $E\eta$ существует, то

$$E\eta = \int_{\Omega} \eta dP = \int_{\mathbb{R}} g(x) P_\xi(dx)$$

(в правой части использована несколько иная запись $\int g(x) dP_\xi(x)$).

Доказательство. Пусть сначала $g(x) = I_B(x)$ есть индикатор множества $B \in \mathfrak{B}$. Тогда $\eta = g(\xi(\omega)) = I_{\{\xi \in B\}}(\omega)$ и $E\eta = P(\xi \in B)$. Поэтому

$$\int g(x) P_\xi(dx) = \int I_B(x) P_\xi(dx) = P_\xi(B) = P(\xi \in B) = E\eta.$$

Используя свойства интеграла, легко установить, что утверждение теоремы верно для простых функций g . Применение предельного перехода распространяет это утверждение на ограниченные функции. Пусть теперь $g \geq 0$. Если функция $g(\xi)I_B(\xi) = \eta(\omega)I_{\{\xi \in B\}}(\omega)$ ограничена, то получаем $\int_B g(x) dP_\xi(x) = E(\eta; \xi \in B)$.

Поэтому

$$\int_{\{g \leq n\}} g dP_\xi = E(\eta; \eta \leq n).$$

Переходя к пределу по n , получим утверждение теоремы. Рассмотрение случая, когда g принимает значения обоих знаков, затруднений не вызывает. ■

Если обозначить

$$F(x) = P(\xi < x),$$

то наряду с интегралом

$$\int_R g(x) P_\xi(dx), \quad (7)$$

который только что рассматривался, можно рассмотреть также интеграл Римана—Стилтьеса

$$\int g(x) dF(x), \quad (8)$$

определение которого дано в §3.6. Там же было показано, что для *непрерывных* функций $g(x)$ эти интегралы совпадают. Кроме того, в §3.6 были обсуждены некоторые другие условия совпадения этих интегралов.

Напомним также, что если $F(x) = \int_{-\infty}^x f(t) dt$, а функции $g(x)$ и $f(t)$ интегрируемы по Риману, то интегралы (7) и (8) совпадают с интегралом Римана

$$\int g(x)f(x) dx.$$

3. Произведения мер и повторные интегралы.

Рассмотрим двумерную случайную величину $\zeta = (\xi, \eta)$, заданную на $(\Omega, \mathfrak{F}, P)$. Случайные величины ξ и η индуцируют выборочное вероятностное пространство $(\mathbb{R}^2, \mathfrak{B}^2, P_{\xi, \eta})$ с мерой $P_{\xi, \eta}$, заданной на элементах σ -алгебры \mathfrak{B}^2 борелевских множеств на плоскости (минимальная σ -алгебра, порожденная прямоугольниками) и такой, что

$$P_{\xi, \eta}(A \times B) = P(\xi \in A, \eta \in B).$$

Здесь $A \times B$ есть множество точек на плоскости (x, y) для которых $x \in A$, $y \in B$. Если $g(x, y)$ есть борелевская функция ($\{(x, y): g(x, y) \in B\} \in \mathfrak{B}^2$ для каждого $B \in \mathfrak{B}$), то из предыдущего легко следует, что

$$Eg(\xi, \eta) = \int_{\mathbb{R}^2} g(x, y) P_{\xi, \eta}(dx dy), \quad (9)$$

так как оба интеграла равны $\int_{\mathbb{R}} \theta P_\theta(dx)$ для $\theta = g(\xi, \eta)$.

Пусть теперь ξ и η независимы, т. е.

$$P(\xi \in A, \eta \in B) = P(\xi \in A)P(\eta \in B)$$

для любых $A, B \in \mathfrak{B}$.

Теорема Фубини (теорема о повторных интегралах). Если $g(x, y) \geq 0$ — борелевская функция, ξ и η независимы, то

$$Eg(\xi, \eta) = E[Eg(x, \eta)|_{x=\xi}].$$

Для произвольных борелевских функций $g(x, y)$ приведенное равенство справедливо, если существует $Eg(\xi, \eta)$.

Это же утверждение в гл. 3 мы записывали с помощью интегралов

$$\int g(x, y) P_{\xi, \eta}(dx dy) = \int \left[\int g(x, y) P_{\eta}(dy) \right] P_{\xi}(dx). \quad (10)$$

Нам будет полезна следующая

Лемма 4.

1. Сечение B_x любого множества $B \in \mathfrak{B}^2$

$$B_x = \{y: (x, y) \in B\}$$

измеримо: $B_x \in \mathfrak{B}$.

2. Сечение $g_x(y) = g(x, y)$ любой борелевской функции g (\mathfrak{B}^2 -измеримой) есть борелевская функция.

3. Интеграл

$$\int g(x, y) P_{\eta}(dy) \quad (11)$$

от борелевской функции g есть борелевская функция от x .

Доказательство.

1. Пусть \mathcal{K}_1 — класс всех множеств из \mathfrak{B}^2 , у которых сечения измеримы. Очевидно, что \mathcal{K}_1 содержит прямоугольники $B = B_{(1)} \times B_{(2)}$, где $B_{(1)} \in \mathfrak{B}$, $B_{(2)} \in \mathfrak{B}$. Кроме того, \mathcal{K}_1 — σ -алгебра. Действительно, рассмотрим, например, множество $B = \bigcup B^{(k)}$, где $B^{(k)} \in \mathcal{K}_1$. Очевидно, что операция \bigcup над множествами $B^{(k)}$ приводит к тем же операциям над сечениями, так что $B_x = \bigcup B_x^{(k)} \in \mathfrak{B}$. Аналогично обстоит дело с другими операциями (\bigcap и взятие дополнений). Таким образом, \mathcal{K}_1 — σ -алгебра, содержащая прямоугольники. Это значит, что $\mathfrak{B}^2 \subset \mathcal{K}_1$.

2. Для $B \in \mathfrak{B}$ имеем

$$g_x^{-1}(B) = \{y: g_x(y) \in B\} = \{y: g(x, y) \in B\} = \{y: (x, y) \in g^{-1}(B)\} = [g^{-1}(B)]_x \in \mathfrak{B}.$$

3. Интеграл (11) есть результат предельного перехода над измеримыми функциями и поэтому сам измерим. ■

Доказательство теоремы Фубини. Докажем сначала (10) в случае $g(x, y) = I_B(x, y)$, когда теорема Фубини превращается в формулу о последовательном вычислении меры множества $B \in \mathfrak{B}^2$:

$$P((\xi, \eta) \in B) = \int P_{\eta}((x, y) \in B) P_{\xi}(dx) = \int P_{\eta}(B_x) P_{\xi}(dx). \quad (12)$$

Введем в рассмотрение функцию множеств

$$Q(B) = \int P_{\eta}(B_x) P_{\xi}(dx).$$

Очевидно, что $Q(B) \geq 0$, $Q(\emptyset) = 0$. Далее, если $B = \bigcup B^{(k)}$, $B^{(k)}$ не пересекаются, то $B_x = \bigcup B_x^{(k)}$, $B_x^{(k)}$ не пересекаются,

$$Q(B) = \int P_{\eta}(\bigcup B_x^{(k)}) P_{\xi}(dx) = \sum \int P_{\eta}(B_x^{(k)}) P_{\xi}(dx) = \sum Q(B^{(k)}).$$

Это означает, что $Q(B)$ — мера.

Мера $Q(B)$ совпадает с $P_{\xi, \eta}(B) = P((\xi, \eta) \in B)$ на прямоугольниках $B = B_{(1)} \times B_{(2)}$. Действительно, для прямоугольников

$$B_x = \begin{cases} B_{(2)} & \text{при } x \in B_{(1)}, \\ \emptyset & \text{при } x \notin B_{(1)}, \end{cases}$$

$$P((\xi, \eta) \in B) = P_{\xi}(B_{(1)})P_{\eta}(B_{(2)}) = \int_{B_{(1)}} P_{\eta}(B_{(2)})P_{\xi}(dx) = \int P_{\eta}(B_x)P_{\xi}(dx) = Q(B).$$

Это означает, что меры Q и $P_{\xi, \eta}$ совпадают на алгебре, порожденной прямоугольниками. По теореме о продолжении меры получаем, что $Q = P_{\xi, \eta}$.

Справедливость (12) доказана. Отсюда следует справедливость теоремы Фубини для простых функций $g_N = \sum_{j=1}^N c_j I_{A_j}$ так как

$$\begin{aligned} E g_N(\xi, \eta) &= \sum_{j=1}^N c_j E I_{A_j}(\xi, \eta) = \\ &= \sum_{j=1}^N c_j \int E I_{A_j}(x, \eta) P_{\xi}(dx) = \int E g_N(x, \eta) P_{\xi}(dx). \end{aligned} \quad (13)$$

Если теперь $g \geq 0$ — произвольная борелевская функция, то существует последовательность простых функций $g_N \uparrow g$, и нам, как и в (13), остается воспользоваться операцией предельного перехода

$$E g(\xi, \eta) = \lim_{N \rightarrow \infty} E g_N(\xi, \eta) = \lim_{N \rightarrow \infty} \int E g_N(\xi, \eta) P_{\xi}(dx) = \int E g_N(\xi, \eta) P_{\xi}(dx).$$

Для произвольной функции g следует воспользоваться представлением $g = g^+ - g^-$, $g^+ \geq 0$, $g^- \geq 0$. ■

Замечание 1. Из доказательства теоремы видно, что скалярность случайных величин ξ и η не играет существенной роли. Утверждение сохранится и в более общей форме (см. свойство 5А § 4.8) и, в частности, для векторных ξ и η .

§ 4. Интеграл по произвольной мере

Если μ — конечная мера на $\langle \Omega, \mathfrak{F} \rangle$, $\mu(\Omega) < \infty$, то определение интеграла $\int \xi d\mu$ по мере μ ничем не отличается от соответствующей конструкции интеграла по вероятностной мере (можно положить просто $\int_A \xi d\mu = \mu(\Omega) \int_A \xi dP$, где $P(B) = \mu(B)/\mu(\Omega)$ есть распределение вероятностей). Если же μ σ -конечна и $\mu(\Omega) = \infty$, то дело обстоит несколько сложнее, хотя снова все сводится к уже использованным построениям. Сделаем сначала несколько предварительных замечаний.

Пусть $\langle \Omega, \mathfrak{F}, P \rangle$ — вероятностное пространство и $f = f(\omega) \geq 0$ — п. в. конечная неотрицательная измеримая функция (т. е. случайная величина). Рассмотрим функцию множества

$$\mu(A) = \int_A f dP. \quad (14)$$

Если f интегрируема ($\mu(\Omega) < \infty$), то $\mu(A)$ — конечная σ -аддитивная функция множества (см. свойство I 1), удовлетворяющая условиям 1–3 § 1 этого Приложения.

Другими словами, μ — конечная мера на $\langle \Omega, \mathfrak{F} \rangle$. Если же f *неинтегрируема*, то μ — σ -конечная мера, что сразу следует из возможности представления

$$\mu(A) = \sum_{k=1}^{\infty} \int_{A \cap \{k-1 \leq f < k\}} f \, dP$$

(интегралы под знаком суммы, равные $\int_A f I_{(k-1 \leq f < k)} \, dP$, являются, очевидно, конечными мерами).

Таким образом, интегралы вида (14) для любых распределения P и функции $f \geq 0$ представляют собой меру. Оказывается, верно и обратное в известном смысле утверждение.

Лемма 5. Для любой меры μ на $\langle \Omega, \mathfrak{F} \rangle$ найдется распределение P на этом пространстве и измеримая функция $f \geq 0$ такие, что справедливо представление (14).

Таким образом, любая мера представима в виде интеграла от вероятностной меры (т. е. в виде $E(f; A)$ при соответствующих функции f и распределении P).

Доказательство. Пусть μ — σ -конечная мера на $\langle \Omega, \mathfrak{F} \rangle$, и множества $B_j \in \mathfrak{F}$, $j = 1, 2, \dots$, обладают свойствами $\bigcup_{j=1}^{\infty} B_j = \Omega$, $B_i B_j = \emptyset$ при $i \neq j$, $\mu(B_j) < \infty$.

Положим

$$P(A) = \sum_{k=1}^{\infty} \frac{\mu(AB_k)}{2^k \mu(B_k)}. \quad (15)$$

Очевидно, что $P(\Omega) = 1$, и P является мерой. Далее, если $A \subset B_k$, то $\mu(A) = 2^k \mu(B_k) P(A)$. Это означает, что мы должны положить $f(\omega) = 2^k \mu(B_k)$ при $\omega \in B_k$.

Тогда функция множества

$$\varphi(A) = \int_A f \, dP = \int_{\Omega} f I_A \, dP$$

будет совпадать с $\mu(A)$:

$$\varphi(A) = \sum_{k=1}^{\infty} 2^k \mu(B_k) P(AB_k) = \sum_{k=1}^{\infty} 2^k \mu(B_k) \sum_{j=1}^{\infty} \frac{\mu(AB_k B_j)}{2^j \mu(B_j)} = \sum_{k=1}^{\infty} \mu(AB_k) = \mu(A).$$

Здесь, помимо требуемого утверждения, мы получили также, что в представлении (14) множество значений функции f , не ограничивая общности, можно считать счетным.

Функция f , для которой возможно равенство (14), называется *плотностью* меры μ относительно P (или *производной Радона—Никодима* меры μ по P) и обозначается $d\mu/dP$. Очевидно, что изменение функции $f = d\mu/dP$ на множестве P -меры нуль оставляет справедливым равенство (14).

Пусть теперь μ и P — две заданные произвольные меры. Весьма важным в теории вероятностей является вопрос, при каких условиях эти две меры μ и P могут быть связаны соотношением (14) и единственным ли образом (с точностью до значений на множестве нулевой P -меры) определяется при этом функция f (подчеркнем, что в предыдущих рассмотрениях мера P строилась специальным образом по мере μ или наоборот). Ответы на эти вопросы содержатся в теореме Радона—Никодима, которая будет изложена в следующем параграфе.

Теперь же, пользуясь доказанным несложным утверждением леммы 5, дадим определение интеграла по произвольной мере μ .

Пусть μ — σ -конечная мера на (Ω, \mathfrak{F}) , и $\xi \geq 0$ есть \mathfrak{F} -измеримая функция.

Интегралом $\int_A \xi d\mu$ по множеству $A \in \mathfrak{F}$ от функции $\xi \geq 0$ по мере μ называется интеграл по распределению P :

$$\int_A \xi d\mu = \int_A \left(\xi \frac{d\mu}{dP} \right) dP, \quad (16)$$

где P — любая вероятностная мера, удовлетворяющая равенству (14) (например, мера (15)).

Это определение *корректно*, поскольку оно не зависит от выбора P . Действительно, для простых функций ξ ($\xi(\omega) = x_k$ при $\omega \in F_k$)

$$\int_A \xi d\mu = \sum x_k \int_A \frac{d\mu}{dP} I_{F_k} dP = \sum x_k \int_{AF_k} \frac{d\mu}{dP} dP = \sum x_k \mu(AF_k).$$

Если теперь $\xi \geq 0$ — произвольная функция, то по теореме о монотонной сходимости $\int_A \xi d\mu$ равен

$$\lim_{n \rightarrow \infty} \int_A \xi^{(n)} \frac{d\mu}{dP} dP = \lim_{n \rightarrow \infty} \int_A \xi^{(n)} d\mu,$$

где $\xi^{(n)} \uparrow \xi$ есть последовательность простых функций, монотонно сходящихся к ξ (см. лемму 1). И в том, и в другом случае полученный результат от выбора P не зависит.

Интеграл $\int_A \xi d\mu$ от произвольной измеримой функции ξ определяется как

$$\int_A \xi d\mu = \int_A \xi^+ d\mu - \int_A \xi^- d\mu,$$

если оба выражения в правой части конечны (в этом случае говорят, что интеграл $\int_A \xi d\mu$ существует). Здесь, как и прежде, $\xi^+ = \max(0, \xi) \geq 0$, $\xi^- = \max(0, -\xi) \geq 0$, так что $\xi = \xi^+ - \xi^-$.

Мы видим, таким образом, что по существу данное определение интеграла по произвольной мере эквивалентно той конструкции, которая использовалась в §2 этого Приложения. Однако определение в форме (16) избавляет от необходимости повторять весь пройденный путь (причем в более сложных условиях) и позволяет сразу перенести на общий случай все свойства интегралов $\int \xi dP$. Перечислим основные из них, сохранив их нумерацию:

11. $\xi d\mu = \sum_j \int_{A_j} \xi d\mu$, если A_j не пересекаются, $\bigcup A_j = \Omega$.

12. $\int (\xi + \eta) d\mu = \int \xi d\mu + \int \eta d\mu$.

13. $\int a\xi d\mu = a \int \xi d\mu$.

14. $\xi d\mu \leq \int \eta d\mu$, если $\xi \leq \eta$.

15. $|\int \xi d\mu| \leq \int |\xi| d\mu$.

16. Если $c_1 \leq \xi(\omega) \leq c_2$ при $\omega \in A$, то $c_1 \mu(A) \leq \int_A \xi d\mu \leq c_2 \mu(A)$.

17. Если $\xi \geq 0$ и $\int \xi d\mu = 0$, то $\mu(\xi > 0) = 0$.

18. Если $\mu(\xi \neq \eta) = 0$, то $\int \xi d\mu = \int \eta d\mu$.

Очевидно, что полностью сохраняются также и теоремы сходимости.

Теорема о сходимости мажорируемой последовательности. Пусть $|\xi| \leq \eta$ и $\int \eta d\mu$ существует. Тогда, если $\xi_n \rightarrow \xi$ или $\xi_n \rightarrow \xi$ п. в., то

$$\int \xi_n d\mu \rightarrow \int \xi d\mu.$$

Теорема о монотонной сходимости. Если $0 \leq \xi_n \uparrow \xi$, то

$$\int \xi_n d\mu \rightarrow \int \xi d\mu.$$

Теорема Фату—Лебега. Ее формулировка и доказательство получаются из теоремы 2 заменой P на μ .

В заключение отметим, что если $\Omega = \mathbb{R} = (-\infty, \infty)$, $\mathfrak{F} = \mathfrak{B}$ есть σ -алгебра борелевских множеств, μ есть мера Лебега, а функция $g(x)$ непрерывна, то интеграл $\int_{[a,b]} g(x) d\mu(x)$ совпадает с интегралом Римана $\int_a^b g(x) dx$. Это следует из соответствующих замечаний в п. 2 § 3 этого Приложения.

§ 5. Теорема Лебега о разложении и теорема Радона—Никодима

Вернемся к вопросу, который был сформулирован в предыдущем параграфе. При каких условиях на меры μ и λ , заданные на $\langle \Omega, \mathfrak{F} \rangle$, мера μ может быть представлена в виде

$$\mu(A) = \int_A f d\lambda?$$

Здесь мы уже не предполагаем, что мера λ обязательно вероятностная.

Определения. Мера μ называется *абсолютно непрерывной* относительно меры λ (обозначается $\mu \prec \lambda$), если для всякого A такого, что $\lambda(A) = 0$, выполняется $\mu(A) = 0$.

Любое множество N_μ называется *носителем меры μ* , если $\mu(\Omega - N_\mu) = 0$.

Говорят, что мера μ *сингулярна* относительно λ , если найдется такой носитель N_λ меры λ , для которого $\mu(N_\lambda) = 0$. Или, что то же, если найдется носитель N_μ меры μ , для которого $\lambda(N_\mu) = 0$.

Таким образом, последнее определение, в отличие от первого, симметрично, и можно говорить о *взаимной сингулярности мер μ и λ* (такое соотношение часто обозначается $\mu \perp \lambda$).

Теорема Радона—Никодима. Для абсолютной непрерывности $\mu \prec \lambda$ необходимо и достаточно, чтобы существовала функция f , определенная с точностью до λ -эквивалентности (т. е. с точностью до значений на множестве λ -меры нуль), такая,

что

$$\mu(A) = \int_A f d\lambda^*.$$

Как уже отмечалось, функция f называется *производной Радона—Никодима* $d\mu/d\lambda$ меры μ по мере λ (или *плотностью* μ относительно λ).

Так как достаточность в утверждении этой теоремы очевидна, то получим теорему Радона—Никодима как следствие следующей теоремы Лебега о разложении меры:

Теорема Лебега. Пусть μ и λ — две σ -конечные меры на $\langle \Omega, \mathfrak{F} \rangle$. Существует единственное разложение меры μ на две компоненты μ_a и μ_c такие, что

$$\mu_a \prec \lambda, \quad \mu_c \perp \lambda.$$

При этом найдется единственная с точностью до λ -эквивалентности функция f , для которой

$$\mu_a(A) = \int_A f d\lambda.$$

Очевидно, что если $\mu \prec \lambda$, то $\mu_c = 0$, и из теоремы Лебега следует теорема Радона—Никодима.

Доказательство. Так как μ и λ σ -конечны, то существуют возрастающие последовательности вложенных множеств Ω_n^μ и Ω_n^λ такие, что

$$\mu(\Omega_n^\mu) < \infty, \quad \lambda(\Omega_n^\lambda) < \infty, \quad \bigcup_n \Omega_n^\mu = \Omega, \quad \bigcup_n \Omega_n^\lambda = \Omega.$$

Положив $\Omega_n = \Omega_n^\mu \cap \Omega_n^\lambda$, получим возрастающую до Ω последовательность множеств, на которых

$$\mu(\Omega_n) < \infty, \quad \lambda(\Omega_n) < \infty.$$

Если мы докажем теорему о разложении для сужения мер μ и λ на $\langle B_n, \mathfrak{F}_n \rangle$, где $B_n = \Omega_{n+1} - \Omega_n$, \mathfrak{F}_n образована множествами $B_n A$, $A \in \mathfrak{F}$, то тем самым мы докажем ее и на всем Ω . Достаточно будет в качестве μ_a и μ_c взять суммы соответствующих компонент по каждому из сужений. Это замечание означает, что можно рассматривать лишь случай конечных мер.

Итак, пусть μ и λ — конечные меры.

а) Пусть Φ — класс функций $f \geq 0$, для которых

$$\int_A f d\lambda \leq \mu(A) \quad \text{при всех } A \in \mathfrak{F} \quad (17)$$

(класс Φ не пуст, так как $f = 0$ принадлежит Φ). Обозначим

$$\alpha = \sup_{f \in \Phi} \int_{\Omega} f d\lambda \leq \mu(\Omega) < \infty$$

и выберем последовательность f_n такую, что

$$\int f_n d\lambda \rightarrow \alpha.$$

* Это равенство иногда принимается за определение абсолютной непрерывности.

Положим $f'_n = \max(f_1, \dots, f_n)$. Тогда, очевидно, $f'_n \uparrow f' = \sup f_n$, и по теореме о монотонной сходимости

$$\int_A f'_n d\lambda \rightarrow \int_A f' d\lambda. \quad (18)$$

Покажем теперь, что $f' \in \Phi$, т.е. что для f' выполнено (17). Для этого в силу (18) достаточно заметить, что $f'_n \in \Phi$. Пусть A_k , $k = 1, \dots, n$, — непересекающиеся множества, на которых $f'_n = f_k$. Тогда $\bigcup_{k=1}^n A_k = \Omega$,

$$\int_A f'_n d\lambda = \sum_{k=1}^n \int_{AA_k} f_k d\lambda \leq \sum_{k=1}^n \mu d\lambda(AA_k) = \mu(A).$$

Таким образом, для «максимального» элемента f' из Φ также выполнено (17).

б) Положим

$$\mu_a(A) = \int_A f' d\lambda, \quad \mu_c = \mu - \mu_a \quad (19)$$

и докажем, что μ_c сингулярна относительно λ . Для этого нам понадобится следующее утверждение о разложении произвольной обобщенной меры (определение см. в § 1 этого Приложения).

Теорема Хана о разложении. Для любой обобщенной конечной меры γ существуют непересекающиеся множества $D^+ \in \mathfrak{F}$ и $D^- \in \mathfrak{F}$ такие, что для любого $A \in \mathfrak{F}$

$$\gamma(AD^+) \geq 0, \quad \gamma(AD^-) \leq 0.$$

Доказательство. Покажем сначала, что существует множество $D \in \mathfrak{F}$, на котором $\gamma(A)$ достигает своей верхней границы.

Пусть $B_n \in \mathfrak{F}$ — такая последовательность, что $\gamma(B_n) \rightarrow \Gamma = \sup_A \gamma(A)$. Положим $B = \bigcup B_k$ и рассмотрим для данного n разбиение B на 2^n множеств $B_{n,m}$, $m = 1, \dots, 2^n$, вида $\bigcap_{k=1}^n B'_k$, где $B'_k = B_k$ или $B - B_k$, $k \leq n$. При $n < N$ каждое $B_{n,m}$ представляет собой конечную сумму множеств $B_{N,M}$, $1 \leq M \leq 2^N$. Обозначим D_n сумму всех тех $B_{n,m}$, на которых $\gamma(B_{n,m}) \geq 0$. Тогда $\gamma(B_n) \leq \gamma(D_n)$.

С другой стороны, при $N > n$ каждое $B_{N,M}$ либо лежит в D_n , либо не пересекается с ним. Поэтому

$$\gamma(D_n) \leq \gamma(D_n + D_{n+1} + \dots + D_N).$$

Отсюда следует, что для множества $D = \lim_{k \rightarrow \infty} \bigcup_{k=n}^{\infty} D_k$ выполняется $\gamma(B_n) \leq \gamma(D)$, $\Gamma \leq \gamma(D)$. Вспоминая определение Γ , получаем $\gamma(D) = \Gamma$.

Таким образом, существование множества D , на котором $\gamma(D)$ максимально, доказано. Покажем теперь, что для любого $A \in \mathfrak{F}$ выполняется $\gamma(AD) \geq 0$ и $\gamma(A\bar{D}) \leq 0$, где $\bar{D} = \Omega - D$. Действительно, допустив, например, что $\gamma(AD) < 0$, мы придем к противоречию, так как тогда

$$\gamma(D - AD) = \gamma(D) - \gamma(AD) > \gamma(D).$$

Аналогично, допустив $\gamma(\overline{AD}) > 0$, мы получили бы

$$\gamma(D + \overline{AD}) = \gamma(D) + \gamma(\overline{AD}) > \gamma(D).$$

Нам остается положить $D^+ = D$, $D^- = \overline{D}$. ■

Следствие 1. Любая обобщенная конечная мера γ представима в виде $\gamma = \gamma^+ - \gamma^-$, где γ^\pm — неотрицательные конечные меры.

Для доказательства достаточно положить

$$\gamma^\pm(A) = \pm \gamma(AD^\pm),$$

где D^\pm — множества из теоремы Хана о разложении.

Вернемся к доказательству того, что мера μ_c в равенстве (19) сингулярна. Пусть D_n^+ — множества в разложении Хана для обобщенной меры

$$\varphi_n = \mu_c - \frac{1}{n}\lambda.$$

Обозначим $N = \bigcap D_n^-$. Тогда $\overline{N} = \bigcup D_n^+$ и при всех n и $A \in \mathfrak{F}$

$$0 \leq \mu_c(AN) \leq \frac{1}{n}\lambda(AN).$$

Отсюда, полагая $n \rightarrow \infty$, получаем $\mu_c(AN) = 0$ и, следовательно, $\mu_c(A) = \mu_c(A\overline{N})$. То есть множество \overline{N} является носителем μ_c .

Далее, так как

$$\mu_a(A) = \mu(A) - \mu_c(A\overline{N}) \leq \mu(A) - \mu_c(AD_n^+),$$

то

$$\int_A \left(f' + \frac{1}{n} I_{D_n^+} \right) d\lambda = \mu_a(A) + \frac{1}{n} \lambda(AD_n^+) \leq \mu(A) - \varphi_n(AD_n^+) \leq \mu(A).$$

Это означает, что $f' + 1/n I_{D_n^+} \in \Phi$ и, следовательно,

$$\alpha \geq \int \left(f' + \frac{1}{n} I_{D_n^+} \right) d\lambda = \alpha + \frac{1}{n} \lambda(D_n^+).$$

Отсюда вытекает, что $\lambda(D_n^+) = 0$, $\lambda(\overline{N}) = 0$ и, значит, μ_c сингулярна относительно λ , поскольку \overline{N} есть носитель μ_c .

Единственность разложения $\mu = \mu_a + \mu_c$ может быть установлена следующим образом. Допустим, что $\mu = \mu'_a + \mu'_c$ — некоторое другое разложение. Тогда $\gamma \equiv \mu'_a - \mu_a = \mu_c - \mu'_c$. В силу свойства сингулярности существуют множества N и N' такие, что $\mu_c(\overline{N}) = 0$, $\lambda(N) = 0$, $\mu'_c(\overline{N}') = 0$, $\lambda(N') = 0$. Очевидно, $\lambda(D) = 0$, где $D = N + N'$. Если допустить, что $\gamma = \mu'_a - \mu_a = \mu_c - \mu'_c \neq 0$, то найдется $A \in \mathfrak{F}$, при котором $\gamma(A) \neq 0$. Следовательно, либо $\gamma(AD) \neq 0$, либо $\gamma(A\overline{D}) \neq 0$. Однако первое невозможно, поскольку $\lambda(D) = 0$ влечет за собой $\mu'_a(D) = \mu_a(D) = 0$. Второе также невозможно, поскольку $\overline{D} = \overline{N} \cup \overline{N}'$ и, следовательно, $\mu'_c(\overline{D}) = \mu'_c(\overline{D}) = 0$.

Единственность функции f (с точностью до λ -эквивалентности), вытекает из того, что равенства $\mu_a(A) = \int_A f d\lambda = \int_A f' d\lambda$, $\int_A (f - f') d\lambda = 0$ для всех A влекут за собой равенство $f - f' = 0$ п. в. Если допустить, например, что $\lambda(A) > 0$ для $A = \{\omega: f - f' > \varepsilon\}$, то для такого A мы получили бы $\int_A (f - f') d\lambda > 0$. ■

Одним из наиболее существенных приложений теоремы Радона—Никодима является доказательство существования и единственности условных математических ожиданий. Пусть \mathfrak{F}_0 есть σ -подалгебра \mathfrak{F} и ξ — случайная величина на $(\Omega, \mathfrak{F}, P)$ такая, что $E\xi$ существует. Мы определили в § 4.8 условное математическое ожидание $E(\xi/\mathfrak{F}_0)$ величины ξ относительно \mathfrak{F}_0 как \mathfrak{F}_0 -измеримую случайную величину γ , для которой

$$E(\gamma; B) = E(\xi; B) \quad (20)$$

при любом $B \in \mathfrak{F}_0$. Можно считать, не ограничивая общности, что $\xi \geq 0$ (произвольная функция ξ представима в виде разности двух положительных). Тогда правая часть в (20) будет мерой на (Ω, \mathfrak{F}_0) . Так как $E(\xi; B) = 0$, если $P(B) = 0$, то она будет абсолютно непрерывной относительно P . Это влечет за собой по теореме Радона—Никодима существование единственной с точностью до P -эквивалентности функции γ на (Ω, \mathfrak{F}_0) такой, что для всякого $B \in \mathfrak{F}_0$

$$E(\xi; B) = \int_B \gamma dP.$$

Полученное соотношение, очевидно, эквивалентно (20) и устанавливает требуемое существование и единственность условного математического ожидания.

Другое следствие доказанных в этом параграфе утверждений упоминалось в § 3.6 и касалось теоремы Лебега о том, что любое распределение P на прямой $\mathbb{R} = (-\infty, \infty)$ (или соответствующая функция распределения) единственным образом представимо в виде суммы трех компонент $P = P_a + P_c + P_d$, где компонента P_a абсолютно непрерывна относительно меры Лебега: $P_a(A) = \int_A f(x) dx$; P_d — дискретная компонента, сосредоточенная на не более чем счетном множестве точек x_1, x_2, \dots таких, что $P(x_k) > 0$; компонента P_c имеет носитель лебеговой меры 0 и непрерывную функцию распределения. Это есть прямое следствие теоремы Лебега о разложении. Надо только из сингулярной компоненты распределения P (относительно лебеговой меры λ) выделить дискретную ее часть, удалив сначала все точки x , для которых $P(\{x\}) \geq 1/2$, затем все точки x , для которых $P(\{x\}) \geq 1/3$ и т.д. Ясно, что мы получим при этом не более чем счетное множество x -в и что такое выделение дискретной компоненты определяет P_d однозначно.

Все сказанное, очевидно, сохранится и для распределений в n -мерных евклидовых пространствах \mathbb{R}^n .

§ 6. Слабая сходимость и сходимость по вариации распределений в произвольных пространствах

1. Слабая сходимость. В § 6.2 и § 7.6 мы изучали слабую сходимость распределений случайных величин и случайных векторов, т.е. слабую сходимость распределений в \mathbb{R}^k , $k \geq 1$. Теперь мы хотим ввести понятие слабой сходимости в более общих пространствах \mathcal{X} . Как показывают определения в § 6.2, нам понадобится понятие непрерывных функций $f(x)$ на \mathcal{X} . Это возможно, если только пространство \mathcal{X} наделено некоторой топологией. Для простоты мы ограничимся ситуацией, когда пространство \mathcal{X} является метрическим с метрикой $\rho(x, y)$. Итак, пусть дано измеримое пространство $(\mathcal{X}, \mathfrak{B})$ с метрикой ρ , которая «согласована» с σ -алгеброй \mathfrak{B} , т.е. все открытия множества из \mathcal{X} (относительно метрики ρ) принадлежат \mathfrak{B} (ср. с § 15.1), так что любой непрерывный (относительно ρ) функционал

будет измеримым относительно \mathfrak{B} . Это означает, что если на $\langle \mathcal{X}, \mathfrak{B} \rangle$ задано распределение \mathbf{Q} (т. е. задано вероятностное пространство $\langle \mathcal{X}, \mathfrak{B}, \mathbf{Q} \rangle$), то $\{x : f(x) < t\} \in \mathfrak{B}$ для любого t и определена вероятность этого множества.

Пусть теперь $\langle \Omega, \mathfrak{F}, \mathbf{P} \rangle$ — исходное вероятностное пространство. Измеримое отображение $\xi = \xi(\omega)$ пространства $\langle \Omega, \mathfrak{F} \rangle$ в $\langle \mathcal{X}, \mathfrak{B} \rangle$ называется \mathcal{X} -значным случайным элементом. Это отображение может быть и тождественным, если $\langle \Omega, \mathfrak{F} \rangle = \langle \mathcal{X}, \mathfrak{B} \rangle$. Пространство $\langle \mathcal{X}, \mathfrak{B} \rangle$ называют *выборочным* или *фазовым* для случайного элемента ξ . В этой ситуации $f(\xi)$ будет случайной величиной в $\langle \mathcal{X}, \mathfrak{B} \rangle$.

Определение 1. Пусть на $\langle \mathcal{X}, \mathfrak{B} \rangle$ заданы распределение \mathbf{P} и последовательность распределений \mathbf{P}_n . Говорят, что *последовательность \mathbf{P}_n слабо сходится* к $\mathbf{P} : \mathbf{P}_n \Rightarrow \mathbf{P}$, если для любого непрерывного, ограниченного функционала f ($f \in C_b(\mathcal{X})$)

$$\int f(x) d\mathbf{P}_n(x) \rightarrow \int f(x) d\mathbf{P}(x). \quad (21)$$

Если случайные элементы, соответствующие распределениям \mathbf{P}_n и \mathbf{P} обозначить, соответственно, ξ_n и ξ , то (21) эквивалентно

$$\mathbf{E}f(\xi_n) \rightarrow \mathbf{E}f(\xi). \quad (22)$$

Это, в свою очередь, эквивалентно тому, что для любого непрерывного функционала $f(f \in C(\mathcal{X}))$

$$f(\xi_n) \Rightarrow f(\xi). \quad (23)$$

Действительно, (23) означает, что для любой непрерывной ограниченной функции g ($g \in C_b(\mathbb{R})$)

$$\mathbf{E}g(f(\xi_n)) \rightarrow \mathbf{E}g(f(\xi)), \quad (24)$$

что эквивалентно (22).

Если $\mathcal{X} = \mathcal{X}(T)$ есть пространство вещественнозначных функций $x(t)$, $t \in T$, на параметрическом множестве T , и задано измеримое отображение $\xi(\omega)$ основного вероятностного пространства $\langle \Omega, \mathfrak{F}, \mathbf{P} \rangle$ в $\langle \mathcal{X}, \mathfrak{B} \rangle$, то случайный элемент $\xi(\omega) = \xi(\omega, t)$ будет *случайным процессом* (см. § 17.1), если $\{x : x(t) < u\} \in \mathfrak{B}$ при всех t, u . В этом случае речь в (21)–(24) будет идти о слабой сходимости распределений случайных процессов, которая уже изучалась нами в гл. 19.

В метрическом пространстве \mathcal{X} для любого $A \in \mathcal{X}$ определена граница $\partial A = [A] - (A)$, где $[A]$ — замыкание множества A , (A) — его внутренность ($(A) = \mathcal{X} - [\bar{A}]$, \bar{A} — дополнение к A).

Определение 2 Множество A называется *множеством непрерывности* распределения \mathbf{P} , (\mathbf{P} -непрерывным множеством), если $\mathbf{P}(\partial A) = 0$. Класс всех \mathbf{P} -непрерывных множеств обозначим \mathcal{D}_P .

Имеет место следующий критерий слабой сходимости распределений.

Теорема 4. Следующие 4 условия эквивалентны

- (i) $\mathbf{P}_n \Rightarrow \mathbf{P}$,
- (ii) $\lim_{n \rightarrow \infty} \mathbf{P}_n(A) = \mathbf{P}(A)$ для всех $A \in \mathcal{D}_P$,
- (iii) $\limsup_{n \rightarrow \infty} \mathbf{P}_n(F) \leq \mathbf{P}(F)$ для всех замкнутых множеств $F \subset \mathcal{X}$,
- (iv) $\liminf_{n \rightarrow \infty} \mathbf{P}_n(G) \geq \mathbf{P}(G)$ для всех открытых множеств $G \subset \mathcal{X}$.

Отметим, что если $P_n \Rightarrow P$, то сходимость (21)–(23) имеет место для более широкого класса функционалов чем $C_b(X)$ ($C(X)$), а именно, для так называемых P -непрерывных функционалов (или непрерывных с P -вероятностью 1). Мы будем называть так функционалы f , для которых $f(x_n) \rightarrow f(x)$ при $\rho(x_n, x) \rightarrow 0$ не для всех $x \in X$, а лишь для $x \in A$, $P(A) = 1$. Класс P -непрерывных функционалов мы будем обозначать $C_P(X)$.

Классы \mathcal{D} и $C_P(X)$, а также классы всех замкнутых и открытых множеств, фигурирующих в теореме 4, являются очень широкими, что делает проверку условий теоремы 4 делом весьма трудным и неудобным. Эти классы можно значительно сузить, если рассматривать не произвольные последовательности P_n , а лишь *секвенциально компактные* (из любой подпоследовательности P'_n можно выбрать сходящуюся). Такой подход мы уже использовали в § 6.3).

Определение 3. Говорят, что класс \mathcal{D} множеств из \mathcal{B} определяет меру P , если какова бы ни была мера Q , равенства $P(A) = Q(A)$ для всех $A \in \mathcal{D}$ влекут за собой $Q = P$.

Класс \mathcal{D} определяет меру P , если \mathcal{D} есть алгебра и $\sigma(\mathcal{D}) = \mathcal{B}_X$ (условие $\sigma(\mathcal{D}) = \mathcal{B}_X$ недостаточно (см. [4]).

Аналогично вводится класс Φ функционалов f , определяющих распределение P элемента $\xi = \xi^P$: для любого Q совпадение распределений $f(\xi^P)$ и $f(\xi^Q)$ для $f \in \Phi_{C_P}(X)$ влечет за собой $P = Q$. σ -алгебру \mathcal{B} пространства X мы будем обозначать, где это потребуется, \mathcal{B}_X .

Теорема 5. Для сходимости $P_n \Rightarrow P$ необходимо и достаточно, чтобы

- 1) Последовательность P_n была секвенциально компактной;
- 2) Существовал класс событий $\mathcal{D} \subset \mathcal{B}_X$, определяющий меру P такой, что $P_n(A) \rightarrow P(A)$ для каждого $A \in \mathcal{D}$.

Альтернативой условию 2) может быть существование класса функционалов Φ , определяющего меру P такого, что $P(f(\xi_n) < t) \Rightarrow P(f(\xi) < t)$ для всех $f \in \Phi_{C_P}(X)$.

При выяснении условий компактности $\{P_n\}$ важную роль играет понятие *плотности* $\{P_n\}$.

Определение 4. Семейство распределений $\{P_n\}$ на X называется *плотным*, если для любого $\varepsilon > 0$ существует компакт $K = K_\varepsilon \subset X$ такой, что $P_n(K) > 1 - \varepsilon$ при всех n .

Справедлива следующая

Теорема 6 (Прохоров). Если семейство $\{P_n\}$ плотно, то оно секвенциально компактно. Если X — полное сепарабельное пространство, то верно и обратное утверждение.

Так как компакты для многих функциональных пространств (в частности, пространств $C(0, T)$, $D(0, T)$), явно описаны, то мы получаем возможность исследовать условия сходимости $P_n \Rightarrow P$ в этих пространствах. Хорошо известно, например, что в этих пространствах компакты имеют вид множеств $\{x : \omega_\Delta(x) \leq \varepsilon(\Delta)\}$, где $\omega_\Delta(x)$ — так называемый «модуль непрерывности» (соответственно в пространствах C или D), $\varepsilon(\Delta) \geq 0$ — любая функция такая, что $\varepsilon(\Delta) \downarrow 0$ при $\Delta \downarrow 0$.

Доказательства теорем 4–6 можно найти, например, в [1]. Мы не приводим их здесь, так как они выходят в какой-то мере за рамки этой книги, а с другой стороны, сами эти теоремы в полном их объеме в основном тексте книги не используются. Используются лишь специальные случаи этих теорем, изложенные в §§ 6.2, 6.3.

Принцип инвариантности в § 17.1 является теоремой о слабой сходимости распределений в пространстве $C(0, 1)$. Если для доказательства этой теоремы использовать теоремы 5, 6, то в качестве класса \mathcal{D} следует выбрать класс цилиндрических множеств. Сходимость P_n к P на множествах из такого \mathcal{D} есть сходимость конечномерных распределений процессов $s_n(t)$, порожденных суммами случайных величин (см. § 17.1). Так как приращения $s_n(t)$ по сути независимы, то доказательство этой части теоремы сводится к доказательству асимптотической нормальности этих приращений, что немедленно получается из центральной предельной теоремы. Условие компактности семейства распределений в $C(0, 1)$ требует доказательства, согласно теореме 6, сходимости по вероятности к нулю модуля непрерывности траектории $s_n(t)$ (подробнее см., например, в [1]). Это может быть доказано с помощью неравенства Колмогорова в следствии 10.3.

2. Сходимость по вариации. Итак, если мы рассматриваем слабую сходимость распределений в пространствах общей природы $\langle \mathcal{X}, \mathcal{B} \rangle$, то для этого в этом пространстве необходимо введение топологии, а это не всегда удобно и возможно. Существует другой вид сходимости распределений на $\langle \mathcal{X}, \mathcal{B} \rangle$, не требующий введения топологий. Это есть сходимость по вариации.

Определение 5. Пусть γ — обобщенная конечная мера на $\langle \mathcal{X}, \mathcal{B} \rangle$. Вариацией γ (или нормой $\|\gamma\|$ по вариации) называется число

$$\|\gamma\| = \text{Var } \gamma = \sup_{f: |f| \leq 1} \left| \int f(x) d\gamma(x) \right|, \quad (25)$$

где \sup берется по классу всех \mathcal{B} -измеримых функций $f(x)$ таких, что $|f(x)| \leq 1$.

Ясно, что \sup в (25) будет достигаться, грубо говоря, на таких f , что $f(x) = 1$ в точках x , для которых $d\gamma(x) > 0$ и $f(x) = -1$ в точках x , для которых $d\gamma(x) < 0$. Поэтому (25) можно записать в виде

$$\|\gamma\| = \int |d\gamma(x)|. \quad (26)$$

Точный смысл этому выражению можно придать с помощью теоремы Хана о разложении (см. следствие 1), из которой следует, что

$$\|\gamma\| = \gamma^+(\mathcal{X}) + \gamma^-(\mathcal{X}). \quad (27)$$

Правую часть этого равенства и надо принять за определение $\int |d\gamma(x)|$.

Лемма 6. Если $\gamma(\mathcal{X}) = 0$, то $\|\gamma\| = 2 \sup_{B \in \mathcal{B}} \gamma(B)$.

Доказательство. Из (25) следует, что для любого B (\bar{B} — дополнение к B , $\gamma(B) + \gamma(\bar{B}) = 0$)

$$\|\gamma\| \geq |\gamma(B)| + |\gamma(\bar{B})| = 2|\gamma(B)|.$$

Поэтому $\|\gamma\| \leq 2 \sup_{B \in \mathcal{B}} |\gamma(B)|$.

Чтобы получить обратное неравенство, воспользуемся следствием 1 теоремы Хана о разложении. Как мы уже отмечали, согласно этой теореме (определение множеств D^\pm см. в теореме Хана)

$$\|\gamma\| = \gamma^+(\mathcal{X}) + \gamma^-(\mathcal{X}) = \gamma^+(D^+) + \gamma^-(\bar{D}^+) = \gamma(D^+) - \gamma(\bar{D}^+) = 2\gamma(D^+) \leq 2 \sup_{B \in \mathcal{B}} \gamma(B). \quad \blacksquare$$

Определение 6. Пусть даны распределение P и последовательность распределений P_n , $n = 1, 2, \dots$, на $\langle X, \mathfrak{B} \rangle$. Мы будем говорить, что P_n *сходится к P по вариации*: $P_n \xrightarrow{\text{Var}} P$, если $\|P_n - P\| = \text{Var}(P_n - P) \rightarrow 0$ при $n \rightarrow \infty$.

Сходимость по вариации является весьма сильной формой сходимости. Если $\langle X, \mathfrak{B} \rangle$ является метрическим пространством и $P_n \xrightarrow{\text{Var}} P$, то $P_n \Rightarrow P$. Действительно, так как функционал $f \in C_b(X)$ ограничен: $|f(x)| < b$, то

$$\left| \int f(dP_n - dP) \right| \leq b \int |d(P_n - P)| = b \|P_n - P\| \rightarrow 0.$$

Таким образом, здесь $\int f dP_n \rightarrow \int f dP$ даже без условия непрерывности f .

Обратное утверждение о сходимости $P_n \xrightarrow{\text{Var}} P$, если $P_n \Rightarrow P$, не верно. Пусть, например, $X = [0, 1]$, P_n есть равномерное распределение на множестве точек $\{0, 1/n, \dots, n/n\}$, $P = U_{0,1}$. Ясно, что все P_n сосредоточены на счетном множестве \mathcal{N} всех рациональных чисел. Стало быть, $P_n(\mathcal{N}) = 1$, $P(\mathcal{N}) = 0$, $\|P_n - P\| = P_n(\mathcal{N}) + P(X \setminus \mathcal{N}) = 2$.

Пусть теперь распределение P имеет плотность p относительно некоторой меры μ (можно взять, в частности, $\mu = P$, так что $p(x) \equiv 1$). Обозначим p_n плотность (относительно μ) абсолютно непрерывной компоненты P_n^a (относительно μ) распределения P_n .

Теорема 7. Для сходимости $P_n \xrightarrow{\text{Var}} P$ необходимо и достаточно, чтобы p_n сходилась к p по мере μ , т. е. для любого $\varepsilon > 0$

$$\mu\{x : |p_n(x) - p(x)| > \varepsilon\} \rightarrow 0 \quad \text{при } n \rightarrow \infty.$$

Доказательство. Имеем

$$\int |d(P_n - P)| = \int |p_n - p| \mu(dx) + \text{Var } P_n^c,$$

где P_n^c — сингулярная компонента P_n относительно меры μ .

Пусть $\|P_n - P\| \rightarrow 0$. Тогда

$$\int |p_n - p| d\mu \rightarrow 0,$$

$$\mu\{x : |p_n(x) - p(x)| > \varepsilon\} \leq \varepsilon^{-1} \int |p_n - p| d\mu \rightarrow 0. \quad (28)$$

Пусть теперь $p_n \xrightarrow{\mu} p$. Обозначим

$$B_\varepsilon = \{x : p(x) \geq \varepsilon\}, \quad A_{n,\varepsilon} = \{x : |p_n(x) - p(x)| \leq \varepsilon^2\}.$$

Тогда

$$1 \geq \int_{B_\varepsilon} p d\mu \geq \varepsilon \mu(B_\varepsilon), \quad \mu(B_\varepsilon) \leq \frac{1}{\varepsilon}.$$

Рассмотрим

$$\int |p_n - p| d\mu = \int_{B_\varepsilon A_{n,\varepsilon}} + \int_{\overline{B_\varepsilon A_{n,\varepsilon}}}. \quad (29)$$

Здесь первый интеграл в правой части не превосходит ε . Так как $\lim_{\varepsilon \rightarrow 0} \int_{B_\varepsilon} p \, d\mu \rightarrow 1$, то для заданного $\delta > 0$ и достаточно малых ε будет выполняться $\int_{B_\varepsilon} p \, d\mu > 1 - \delta$, и при достаточно больших n

$$\int_{B_\varepsilon A_{n,\varepsilon}} p \, d\mu > 1 - 2\delta, \quad \int_{B_n A_{n,\varepsilon}} p_n \, d\mu > 1 - 3\delta. \quad (30)$$

Из этих двух неравенств следует, что второй интеграл в (29) не превосходит 5δ . Это доказывает (28). Кроме того, из (30) следует, что $\text{Var } P_n^a > 1 - 3\delta$, $\text{Var } P_n^c < 3\delta$. Теорема доказана. ■

Из теоремы вытекает, что если $P_n \xrightarrow{\text{Var}} P$, то абсолютно непрерывная компонента P_n^a распределения P_n относительно $\mu = P$ имеет плотность $p_n(x) \xrightarrow{p} 1$, $P_n^a(\mathcal{X}) \rightarrow 1$.

Приложение 4

Теоремы Хелли и Арцела—Асколи

Пусть \mathcal{F} — класс всех функций распределения и \mathcal{G} — класс функций G , которые обладают свойствами $F1, F3$ § 3.2 (монотонность и непрерывность слева) и свойствами $G(-\infty) \geq 0$, $G(\infty) \leq 1$. Будем писать $G_n \Rightarrow G$, $G \in \mathcal{G}$, если $G_n(x) \rightarrow G(x)$ во всех точках x непрерывности функции G .

Теорема Хелли. *Всякая последовательность $F_n \in \mathcal{F}$ содержит подпоследовательность $F_{n_k} \Rightarrow F \in \mathcal{G}$.*

Нам понадобится

Лемма 1. *Для сходимости $F_n \Rightarrow F \in \mathcal{G}$ достаточно, чтобы*

$$F_n(x) \rightarrow F(x), \quad x \in D,$$

при $n \rightarrow \infty$ на каком-нибудь всюду плотном множестве D вещественных чисел.

Доказательство. Пусть x — произвольная точка непрерывности $F(x)$. Для произвольных $x', x'' \in D$ таких, что $x' \leq x \leq x''$, выполняется

$$F_n(x') \leq F_n(x) \leq F_n(x'').$$

Следовательно,

$$\lim_{n \rightarrow \infty} F_n(x') \leq \liminf_{n \rightarrow \infty} F_n(x) \leq \limsup_{n \rightarrow \infty} F_n(x) \leq \lim_{n \rightarrow \infty} F_n(x'').$$

В силу условий леммы получаем отсюда

$$F(x') \leq \liminf_{n \rightarrow \infty} F_n(x) \leq \limsup_{n \rightarrow \infty} F_n(x) \leq F(x'').$$

Полагая $x' \uparrow x$ и $x'' \downarrow x$ по множеству D и учитывая, что x — точка непрерывности $F(x)$, получим:

$$\lim_{n \rightarrow \infty} F_n(x) = F(x).$$

Доказательство теоремы. Пусть $D = \{x_n\}$ — произвольное счетное всюду плотное множество вещественных чисел. Последовательность чисел $\{F_n(x_1)\}$ ограничена и поэтому содержит сходящуюся подпоследовательность $\{F_{1_n}(x_1)\}$. Обозначим предел этой подпоследовательности $F(x_1)$. Рассмотрим теперь последовательность чисел $\{F_{1_n}(x_2)\}$. Она также содержит сходящуюся подпоследовательность $\{F_{2_n}(x_2)\}$ с пределом $F(x_2)$. При этом

$$\lim_{n \rightarrow \infty} F_{2_n}(x_1) = F(x_1).$$

Продолжая этот процесс, мы для любого числа k получим k последовательностей

$$\{F_{kn}(x_i)\}, \quad i = 1, \dots, k,$$

таких, что $\lim_{n \rightarrow \infty} F_{kn}(x_i) = F(x_i)$.

Рассмотрим диагональную последовательность функций распределений $\{F_{nn}(x)\}$. Для любого $x_k \in D$ только $k - 1$ первых членов последовательности чисел $\{F_{nn}(x_k)\}$ могут не принадлежать последовательности $F_{kn}(x_k)$. Поэтому $\lim_{n \rightarrow \infty} F_{nn}(x_k) = F(x_k)$.

Ясно, что $F(x)$ есть заданная на D неубывающая ограниченная функция. Ее легко продолжить по непрерывности слева до неубывающей функции на всей прямой. Теперь видно, что последовательность $\{F_{nn}\}$ и функция F удовлетворяют условию леммы 1.

Условия теоремы Хелли могут быть ослаблены. Именно, в качестве F можно рассматривать более широкий класс \mathcal{H} монотонных, непрерывных слева (т.е. удовлетворяющих F_1, F_3) функций H , ограниченных в каждой точке x : $|H(x)| \leq N(x) < \infty$, где N — заданная функция, характеризующая класс \mathcal{H} . Неограниченный рост $|H(x)|$ (или $N(x)$) при $|x| \rightarrow \infty$ не исключается.

Обобщенная теорема Хелли. *Всякая последовательность $H_n \in \mathcal{H}$ содержит подпоследовательность H_{n_n} , сходящуюся к $H \in \mathcal{H}$ в каждой точке непрерывности H .*

Доказательство обобщенной теоремы Хелли полностью повторяет доказательство, приведенное выше.

С каждой функцией $H_n \in \mathcal{H}$ можно ассоциировать меру μ_n , положив $\mu_n([a, b)) = H_n(b) - H_n(a)$. Обобщенная теорема Хелли будет означать тогда существование для любой последовательности μ_n , порожденной функциями из \mathcal{H} , подпоследовательности мер μ_{n_n} , слабо сходящейся на каждом конечном интервале, концы которого не являются атомами предельной меры μ .

Приведем еще один аналог теоремы Хелли, относящийся к совокупности равностепенно непрерывных функций g_n . Напомним, что последовательность $\{g_n\}$ называется равностепенно непрерывной, если для каждого $\varepsilon > 0$ найдется $\delta > 0$ такое, что из $|x_1 - x_2| < \delta$ вытекает $|g_n(x_1) - g_n(x_2)| < \varepsilon$ при всех n .

Теорема Арцела—Асколи. *Пусть $\{g_n\}$ — последовательность равномерно ограниченных и равностепенно непрерывных функций вещественной переменной. Тогда найдется последовательность g_{n_k} , сходящаяся к непрерывному пределу g . Эта сходимость равномерна на каждом конечном интервале.*

Доказательство. Выберем снова счетное всюду плотное множество $\{x_n\}$ и подпоследовательность $\{g_{n_k}\}$, сходящуюся в точках x_1, x_2, \dots . Обозначим ее предел в точке x_j через $g(x_j)$. Имеем

$$|g_{n_k}(x) - g_{n_r}(x)| \leq |g_{n_k}(x) - g_{n_k}(x_j)| + |g_{n_k}(x) - g_{n_r}(x_j)| + |g_{n_k}(x_j) - g_{n_r}(x_j)|. \quad (1)$$

По предположению последнее слагаемое справа сходится к 0 при $n_k \rightarrow \infty$, $n_r \rightarrow \infty$. Вследствие равностепенной непрерывности для любой точки x найдется точка x_j такая, что при всех n

$$|g_n(x) - g_n(x_j)| < \varepsilon. \quad (2)$$

Если x принадлежит интервалу I , то в I можно найти конечное число точек x_j таких, что будет выполнено (2). Отсюда вытекает, что правая часть в (1) будет меньше 3ε для всех достаточно больших n_k , n_r равномерно по $x \in I$. Итак, существует предел $g(x) = \lim g_{n_k}(x)$, и в силу (2) справедливо $|(g(x) - g(x_j))| \leq \varepsilon$, откуда следует непрерывность g .

Приложение 5

Доказательство теоремы Берри—Эссена

Докажем следующее утверждение (см. § 8.5):

Теорема (Берри—Эссен). Пусть ξ_k независимы, одинаково распределены, $E\xi_k = 0$, $D\xi_k = 1$, $\mu = E|\xi_k|^3 < \infty$, $S_n = \sum_{k=1}^n \xi_k$, $\zeta_n = S_n/\sqrt{n}$. Тогда при всех n

$$\Delta_n = \sup_x |P(\zeta_n < x) - \Phi(x)| < \frac{c\mu}{\sqrt{n}},$$

где c — абсолютная постоянная.

Доказательство будет проводиться методом композиций. Основанием для оценки Δ_n будет, как и в § 8.5, оценка малости $Eg(\zeta_n) - Eg(\eta)$, $\eta \in \Phi_{0,1}$, для гладких g . Чтобы получить оценку для Δ_n , мы в § 8.5 выбирали g с изменением, сосредоточенным на малом интервале. Следующая лемма показывает, что такой выбор g не обязателен. Пусть G — какая-нибудь функция распределения, $\gamma \in G$ и не зависит от ζ_n , η . Положим $g(z) = G\left(\frac{x-z}{\varepsilon}\right)$, так что $Eg(\zeta_n) = EG\left(\frac{x-\zeta_n}{\varepsilon}\right) = P\left(\gamma < \frac{x-\zeta_n}{\varepsilon}\right) = P(\zeta_n + \varepsilon\gamma < x)$, $Eg(\eta) = P(\eta + \varepsilon\gamma < x)$.

Обозначим

$$\begin{aligned} \Delta_{n,\varepsilon} &= \sup_x \left| EG\left(\frac{x-\zeta_n}{\varepsilon}\right) - EG\left(\frac{x-\eta}{\varepsilon}\right) \right| = \\ &= \sup_x \left| P(\zeta_n + \varepsilon\gamma < x) - P(\eta + \varepsilon\gamma < x) \right| = \\ &= \sup_x \left| \int dG(y) [P(\zeta_n < x - \varepsilon y) - P(\eta < x - \varepsilon y)] \right|. \end{aligned}$$

Очевидно, что $\Delta_{n,\varepsilon} \leq \Delta_n$. Нашей целью будет получение обратного для Δ_n неравенства.

Лемма 1. Пусть $v > 0$ таково, что $G(v) - G(-v) \geq 3/4$. Тогда при любом $\varepsilon > 0$

$$\Delta_n \leq 2\Delta_{n,\varepsilon} + \frac{3v\varepsilon}{\sqrt{2\pi}}.$$

Доказательство. Предположим, что \sup_x в определении Δ_n достигается на положительном значении $\Delta_n(x) \equiv F_n(x) - \Phi(x)$ (отрицательное $\Delta_n(x)$ рассматривается аналогично), и пусть для заданного $\delta > 0$ значение x_δ таково, что

$$\Delta_n(x_\delta) = F_n(x_\delta) - \Phi(x_\delta) \geq \Delta_n - \delta,$$

где F_n — функция распределения ζ_n . При увеличении аргумента значение $\Delta_n(x_\delta)$ меняется мало в следующем смысле. Пусть $|y| < v$. Тогда $v - y > 0$,

$$\begin{aligned}\Delta_n(x_\delta + \varepsilon(v - y)) &= F_n(x_\delta + \varepsilon(v - y)) - \Phi(x_\delta + \varepsilon(v - y)) \geq \\ &\geq F_n(x_\delta) - \Phi(x_\delta) - [\Phi(x_\delta + \varepsilon(v - y)) - \Phi(x_\delta)].\end{aligned}$$

Здесь разность в квадратных скобках не превосходит $\varepsilon(v - y)\Phi'(0) \leq 2v\varepsilon/\sqrt{2\pi}$ и, значит,

$$\Delta_n(x_\delta + \varepsilon(v - y)) \geq \Delta_n - \delta - \frac{2v\varepsilon}{\sqrt{2\pi}}.$$

Поэтому

$$\begin{aligned}\Delta_{n,\varepsilon} &\geq \int dG(y)\Delta_n(x_\delta + \varepsilon v - \varepsilon y) = \\ &= \int_{|y|<v} + \int_{|y|\geq v} \geq \frac{3}{4}\left(\Delta_n - \delta - \frac{2v\varepsilon}{\sqrt{2\pi}}\right) - \frac{1}{4}\Delta_n = \frac{\Delta_n}{2} - \frac{3}{4}\left(\delta + \frac{2v\varepsilon}{\sqrt{2\pi}}\right).\end{aligned}$$

Так как δ произвольно, то отсюда следует утверждение леммы. ■

Следствие 1. Для $G = \Phi$ ($\gamma \in \Phi_{0,1}$) значение $v = 6/5$ удовлетворяет условиям леммы 1,

$$\Delta_n \leq 2(\Delta_{n,\varepsilon} + \varepsilon). \quad (1)$$

В качестве следующего этапа доказательства оценим $\Delta_{n,\varepsilon}$, где, собственно, и используется метод композиций. Обозначим

$$u(n) = \max_{k \leq n} \Delta_k \frac{\sqrt{k}}{\mu}, \quad \alpha^2 = \varepsilon^2 n.$$

Буквой c (c индексами или без) будем обозначать абсолютные постоянные, не обязательно одни и те же.

Лемма 2. При $\alpha \geq 1$

$$\Delta_{n,\varepsilon} \leq c\mu\left(\frac{1}{\sqrt{n}} + \frac{\mu u(n-1)}{\alpha\sqrt{n}}\right). \quad (2)$$

Доказательство. Пусть $H_n = \sum_{k=1}^n \eta_k$, $\eta_k \in \Phi_{0,1}$ и независимы. Основу метода композиций составляет тождество (ср. с теоремой 8.6 и тождеством (8.20))

$$\begin{aligned}P(\zeta_n + \varepsilon\gamma < x) - P(\eta + \varepsilon\gamma < x) &= P(S_n + \alpha\gamma < x\sqrt{n}) - P(H_n + \alpha\gamma < x\sqrt{n}) = \\ &= \sum_{m=1}^n [P(S_{m-1} + (H_n - H_m) + \xi_m + \alpha\gamma < x\sqrt{n}) - \\ &\quad - P(S_{m-1} + (H_n - H_m) + \eta_m + \alpha\gamma < x\sqrt{n})].\end{aligned}$$

Так как при $\gamma \in \Phi_{0,1}$ выполняется $H_n - H_m + \alpha\gamma \in \Phi_{0,n-m+\alpha^2}$, то последняя сумма равна $\sum_{m=1}^n D_m$, где

$$\begin{aligned}D_m &= E\left[\Phi\left(\frac{x\sqrt{n} - S_{m-1} - \xi_m}{d_m}\right) - \Phi\left(\frac{x\sqrt{n} - S_{m-1} - \eta_m}{d_m}\right)\right] = \\ &= E\left[\Phi\left(T_m - \frac{\xi_m}{d_m}\right) - \Phi\left(T_m - \frac{\eta_m}{d_m}\right)\right],\end{aligned}$$

$$d_m^2 = n - m + \alpha^2, \quad T_m = \frac{x\sqrt{n} - S_{m-1}}{d_m}.$$

Для оценки слагаемых D_m воспользуемся тем же подходом, что и в лемме 8.3. Так как первые два момента ξ_m и η_m совпадают, то, пользуясь разложением в ряд, получим

$$|D_m| \leq \frac{2\mu}{d_m^3} \sup_t \mathbb{E} \varphi''(T_m + t),$$

где $\varphi(x) = \Phi'(x)$, $\varphi'' = \Phi'''$. Так как функция φ'' ограничена, то

$$|D_m| \leq \frac{c\mu}{d_m^3}. \quad (3)$$

Нам понадобится также другая оценка D_m . Чтобы ее получить, рассмотрим $R_m = \sup_t |\mathbb{E} \varphi''(T_m + t)| \leq \sup_t |\mathbb{E} [\varphi''(T_m + t) - \varphi''(V_m + t)]| + \sup_t |\mathbb{E} \varphi''(V_m + t)|$, (4)

где V_m определяется так же, как T_m , но с заменой S_{m-1} на H_{m-1} . Интегрированием по частям получаем

$$\begin{aligned} |\mathbb{E} [\varphi''(T_m + t) - \varphi''(V_m + t)]| &= \left| \int \varphi''(u + t) d[\mathbb{P}(T_m < u) - \mathbb{P}(V_m < u)] \right| = \\ &= \left| \int \varphi'''(u + t) [\mathbb{P}(T_m < u) - \mathbb{P}(V_m < u)] du \right| \leq \Delta_{m-1} \int \varphi'''(u) du = c\Delta_{m-1}, \end{aligned}$$

поскольку $|\mathbb{P}(T_m < u) - \mathbb{P}(V_m < u)| \leq \Delta_{m-1}$ (T_m и V_m отличаются от $S_{m-1}/\sqrt{m-1}$ и $H_{m-1}/\sqrt{m-1}$, соответственно, одним и тем же линейным преобразованием).

Для оценки второго слагаемого в (4) заметим, что

$$\mathbb{E} \varphi''(V_m + t) = \int \varphi''(u + t) \frac{1}{r_m} \varphi\left(\frac{u - a_m}{r_m}\right) du,$$

где

$$a_m = x \frac{\sqrt{n}}{d_m}, \quad r_m = \sqrt{\frac{m-1}{n-m+\alpha^2}},$$

так что $1/r_m \varphi\left(\frac{u-a_m}{r_m}\right)$ есть плотность $V_m = (x\sqrt{n} - H_{m-1})/d_m$. Интегрируя два раза по частям, находим

$$|\mathbb{E} \varphi''(V_m + t)| = \frac{1}{r_m^3} \left| \int \varphi(u + t) \varphi\left(\frac{u - a_m}{r_m}\right) du \right| \leq \frac{c}{r_m^3}.$$

Итак

$$R_m \leq c \left(\Delta_{m-1} + \frac{1}{r_m^3} \right), \quad D_m \leq c\mu \left(\frac{\Delta_{m-1}}{d_m^3} + \frac{1}{(m-1)^{3/2}} \right).$$

Полученные оценки D_m от x не зависят. Поэтому, используя при $m > n/2$ только что полученную оценку, а при $m \leq n/2$ — оценку (3), в силу которой $|D_m| \leq c\mu/n^{3/2}$, получим

$$\Delta_{n,\varepsilon} \leq c\mu \left[\sum_{m \leq n/2} n^{-3/2} + \sum_{m > n/2} \frac{\Delta_{m-1}}{d_m^3} + \sum_{m > n/2} \frac{1}{(m-1)^{3/2}} \right]. \quad (5)$$

Здесь первая сумма превосходит $n/2 \cdot n^{-3/2} = 1/(2\sqrt{n})$, последняя не превосходит $\int_{n/2-1}^n \frac{ds}{s^{3/2}} \leq c/\sqrt{n}$. Остается оценить среднюю сумму $\left(u(n) = \max_{k \leq n} (\Delta_k \sqrt{k})/\mu\right)$:

$$\sum_{m > n/2}^n \frac{\Delta_{m-1}}{d_m^3} \leq \mu u(n-1) \sqrt{\frac{2}{n}} \sum_{m > n/2}^n \frac{1}{(n-m+\alpha^2)^{3/2}}.$$

Последняя сумма не превосходит

$$\sum_{k=0}^{\infty} \frac{1}{(k+\alpha^2)^{3/2}} \leq \frac{1}{\alpha^3} + \int_0^{\infty} \frac{dt}{(t+\alpha^2)^{3/2}} = \frac{1}{\alpha^3} + \frac{1}{2\alpha} \leq \frac{3}{2\alpha},$$

если $\alpha \geq 1$. Резюмируя (5) и проведенные оценки, получим утверждение леммы. ■

Обратимся теперь непосредственно к доказательству теоремы. В силу (1), (2)

$$v(n) \equiv \frac{\Delta_n \sqrt{n}}{\mu} \leq \frac{2}{\mu} \sqrt{n} \Delta_{n,\varepsilon} + \frac{2\alpha}{\mu} \leq 2c + \frac{2c\mu u(n-1)}{\alpha} + \frac{2\alpha}{\mu}.$$

Положим здесь $\alpha = \max(1, 4c\mu)$. Тогда $(\mu \geq 1)$

$$v(n) \leq c_1 + \frac{u(n+1)}{2}.$$

Отсюда вытекает, что $u(n) \leq 2c_1$ при всех n . Чтобы убедиться в этом, воспользуемся индукцией. Очевидно, что $u(1) = v(1) \leq 1 \leq 2c_1$. Пусть $u(n-1) \leq 2c_1$. Тогда $v(n) \leq 2c_1$, $u(n) = \max(v(n), u(n-1)) \leq 2c_1$. Теорема доказана. ■

Приложение 6

Теоремы восстановления

Основной целью этого раздела является доказательство теоремы 4 гл. 9 — основной теоремы восстановления для нерешетчатого случая. Будут рассмотрены также ее уточнения и обобщения.

Рассмотрим сначала положительные одинаково распределенные случайные величины ξ_j с функцией распределения F , $a = E\xi_j < \infty$. Здесь нам будет удобнее под функцией восстановления понимать ее непрерывную слева версию

$$H(t) = \sum_{k=0}^{\infty} F^{*k}(t), \quad t \geq 0,$$

где F^{*k} — k -ая свертка функции распределения F , являющаяся функцией распределения $S_k = \xi_1 + \dots + \xi_k$.

Теорема 1. Если g — непосредственно интегрируемая функция (см. гл. 9), то

$$\int_0^t g(t-u) dH(u) \rightarrow \frac{1}{a} \int_0^{\infty} g(u) du.$$

Для доказательства теоремы, которое в целом будет следовать пути, предложенному в [22], нам понадобятся несколько вспомогательных утверждений.

Лемма 1. Пусть g — измеримая ограниченная функция. Интеграл

$$G(t) = \int_0^t g(t-u) dH(u) = g * H(t) \quad (1)$$

является единственным решением уравнения

$$G(t) = g(t) + \int_0^t G(t-u) dF(u) = g(t) + G * F(t) \quad (2)$$

в классе функций, ограниченных на конечных интервалах.

$G = H$ является решением (2) при $g \equiv 1$. $G \equiv 1$ является решением (2) при $g = 1 - F$.

Уравнение (2) называется уравнением восстановления.

Как уже отмечалось в теореме 9.4, с функциями H и F очевидным образом можно ассоциировать меры H и F , и интегралы в (1), (2), записывать, соответственно, как интегралы по мере $\int_0^t g(t-u)H(du)$ и $\int_0^t G(t-u)F(du)$.

Доказательство леммы 1. Положим

$$H_n(t) = \sum_{k=0}^n F^{*k}(t).$$

Функции $G_n = g * H_n$ удовлетворяют уравнению $G_{n+1} = g + G_n * F$ и образуют монотонно возрастающую последовательность $G_n \uparrow$, которая в силу леммы 9.2 ограничена. Поэтому $G_n \uparrow G$ и, переходя к пределу в уравнении для G_n , получим, что G удовлетворяет (2). Для доказательства единственности заметим, что разность $V = G^{(1)} - G^{(2)}$ для двух решений $G^{(1)}$ и $G^{(2)}$ должна удовлетворять однородному уравнению $V = V * F$ и, стало быть, соотношениям $V = V * (F^{*k})$ или, что то же,

$$V(t) = \int_0^t V(t-u) dF^{*k}(u).$$

Но $F^{*k}(u) \rightarrow 0$ при $k \rightarrow \infty$ на $[0, t]$. Так как по условию $|V(u)| < c$ на $[0, t]$, то $V(t) \rightarrow 0$ при $k \rightarrow \infty$. Так как V от k не зависит, то $V(t) \equiv 0$. Последнее утверждение леммы проверяется непосредственно. ■

Отметим, что если бы мы рассматривали функции g ограниченной вариации, то утверждение леммы 1 немедленно следовало бы из уравнения для преобразования

Лапласа—Стилтьеса $\tilde{G}(\lambda) = \int_0^\infty e^{-\lambda t} dG(t)$, вытекающего из (2):

$$\tilde{G}(\lambda) = \tilde{g}(\lambda) + \tilde{G}(\lambda)\varphi(\lambda), \quad (3)$$

где $\tilde{g}(\lambda) = \int_0^\infty e^{-\lambda t} dg(t)$, $\varphi(\lambda) = \int_0^\infty e^{-\lambda t} dF(t)$. Действительно, из (3) следует

$$\tilde{G}(\lambda) = \frac{\tilde{g}(\lambda)}{1 - \varphi(\lambda)},$$

что эквивалентно (1).

Мы будем говорить, что точка t является точкой роста функции распределения F , если $F(t+\varepsilon) - F(t) > 0$ при любом $\varepsilon > 0$.

Лемма 2. Пусть распределение F нерешетчато и Z есть множество, образованное точками роста H , т. е. точками роста F, F^{*2}, F^{*3}, \dots . Тогда Z «асимптотически плотно на бесконечности», то есть для любого заданного $\varepsilon > 0$ и всех достаточно больших x пересечение $(x, x+\varepsilon) \cap Z$ не пусто.

Доказательство. Заметим предварительно, что если t_1 — точка роста распределения F_1 случайной величины ξ , а t_2 — точка роста распределения F_2 случайной величины ζ , то $t = t_1 + t_2$ будет точкой роста распределения $F_1 * F_2$ величины $\xi + \zeta$. Действительно, для независимых ξ и ζ

$$P(t \leq \xi + \zeta < t + \varepsilon) \geq P\left(t_1 \leq \xi < t_1 + \frac{\varepsilon}{2}\right) P\left(t_2 \leq \zeta < t_2 + \frac{\varepsilon}{2}\right).$$

Пусть далее $x < y$ — две точки множества Z и $\Delta = y - x$. Имеет место альтернатива:

- 1) Либо при каждом $\varepsilon > 0$ можно найти x и y такие, что $\Delta < \varepsilon$.
- 2) Либо существует $\delta > 0$ такое, что $\Delta \geq \delta$ при всех x и y из Z .

Обозначим $I_n = [xn, yn]$. Если $n\Delta > x$, то этот интервал содержит $[nx, (n+1)x]$ в качестве подинтервала и, следовательно, любая точка $v > v_0 = x^2/\Delta$ принадлежит по крайней мере одному из интервалов I_1, I_2, \dots .

В силу сделанного выше замечания $(n+1)$ точек $n\Delta + k\Delta = (n-k)\Delta + ky$, $k = 0, \dots, n$, принадлежат Z и разбивают I_n на n подинтервалов длины Δ . Это значит, что всякая точка $v > v_0$ удалена от точек из Z не более чем на $\Delta/2$.

В случае 1) отсюда вытекает утверждение леммы.

В случае 2) можно считать, что x и y выбраны так, чтобы $\Delta < 2\delta$. Тогда точками вида $n\Delta + k\Delta$ исчерпываются все точки из Z , лежащие внутри I_n . Так как точка $(n+1)\Delta$ находится среди этих точек, то x кратно Δ , и все точки Z внутри I_n кратны Δ . Пусть теперь z — произвольная точка роста F . При достаточно большом n интервал I_n содержит точку вида $z + k\Delta$, а так как она принадлежит Z , то z также кратно Δ . Таким образом, распределение F является решетчатым и случай 2) в наших условиях не может иметь места. ■

Лемма 3. Пусть q — ограниченная равномерно непрерывная функция, определенная на $(-\infty, \infty)$ и такая, что $q(x) \leq q(0)$ при всех x ,

$$q(x) = \int_0^{\infty} q(x-y) dF(y). \quad (4)$$

Тогда $q(x) \equiv q(0)$.

Доказательство. Уравнение (4) означает, что $q = q * F = \dots q * F^{*k}$ при любом $k \geq 1$. Правая часть (4) не превосходит $q(0)$ и, стало быть, при $x = 0$ равенство (4) возможно, если только $q(-y) = q(0)$ для всех $y \in Z_k$, где Z_k — множество точек роста F^{*k} , и, стало быть, для всех $y \in Z$. В силу леммы 2 и равномерной непрерывности q это означает, что $q(-y) \rightarrow q(0)$ при $y \rightarrow \infty$. Далее, для произвольно большого заданного N можно выбрать k таким образом, что $q(x)$ будет сколь угодно близким к $\int_N^{\infty} q(x-y) dF^{*k}(y)$, поскольку $F^{*k}(N) \rightarrow 0$ при $k \rightarrow \infty$. Это означает, в свою очередь, что $q(x)$ будет близко к $q(0)$. Так как $q(x)$ ни от N , ни от k не зависит, то $q(x) = q(0)$. ■

Лемма 4. Пусть g — непрерывная функция, обращающаяся в 0 вне отрезка $[0, b]$. Тогда соответствующее g решение G уравнения восстановления (2) равномерно непрерывно, и для любого u

$$G(x+u) - G(x) \rightarrow 0 \quad (5)$$

при $x \rightarrow \infty$.

Доказательство. Имеем в силу леммы 9.3

$$\begin{aligned} |G(x+\delta) - G(x)| &= \left| \int_{x-b}^{x+\delta} (g(x+\delta-y) - g(x-y)) dH(y) \right| \leq \\ &\leq \max_{0 \leq x \leq b+\delta} |g(x+\delta) - g(x)| (c_1 + c_2(b+\delta)). \end{aligned} \quad (6)$$

Это означает, что равномерная непрерывность g влечет за собой равномерную непрерывность G .

Допустим теперь, что g имеет непрерывную производную g' . Тогда G' существует и удовлетворяет уравнению восстановления

$$G'(x) = g'(x) + \int_0^x G'(x-y) dF(y).$$

Это означает, что производная G' ограничена и равномерно непрерывна. Пусть

$$\limsup_{x \rightarrow \infty} G'(x) = s.$$

Выберем последовательность $t_n \rightarrow \infty$ такую, что $G'(t_n) \rightarrow s$. Семейство функций q_n , определенных равенствами

$$q_n(x) = G'(t_n + x),$$

равностепенно непрерывно и

$$q_n(x) = g'(t_n + x) + \int_0^{x+t_n} q_n(x-y) dF(y) = g'(t_n + x) + \int_0^\infty q_n(x-y) dF(y). \quad (7)$$

В силу теоремы Арцела—Асколи (см. Приложение 4) существует подпоследовательность t_{n_r} такая, что q_{n_r} сходится к пределу q . Из (7) следует, что этот предел удовлетворяет условиям леммы 3 и поэтому $q(x) = q(0) = s$ при всех x . Таким образом, $G'(t_{n_r} + x) \rightarrow s$ при всех x и, стало быть,

$$G(t_{n_r} + x) - G(t_{n_r}) \rightarrow sx.$$

Поскольку это верно при любом x , а функция G ограничена, то $s = 0$.

Мы доказали лемму для непрерывно дифференцируемой g . Но произвольная непрерывная функция g , равная 0 вне $[0, b]$, может быть приближена непрерывно дифференцируемой функцией g_1 , обращающейся в 0 вне $[0, b]$. Пусть G_1 — соответствующее g_1 решение уравнения восстановления. Тогда из $|g - g_1| < \varepsilon$ следует $|G - G_1| < c\varepsilon$, $c = c_1 + c_2b$ (см. лемму 9.3) и, стало быть,

$$|G(x+u) - G(x)| < (2c+1)\varepsilon$$

при всех достаточно больших x . Это доказывает (5) для произвольной непрерывной функции g . ■

Доказательство теоремы 1. Рассмотрим произвольную последовательность $t_n \rightarrow \infty$ и меры μ_n , порожденные функциями $H_{(n)}(u) = H(t_n+u) - H(t_n)$ ($\mu_n([u, v]) = H_{(n)}(v) - H_{(n)}(u)$). Эти функции удовлетворяют условиям обобщенной теоремы Хелли (см. Приложение 4). Поэтому существует подпоследовательность t_{nn} , соответствующая ей подпоследовательность μ_{nn} и предельная мера μ такие, что μ_{nn} слабо сходится к μ на любом конечном интервале.

Пусть теперь g непрерывная функция, равная 0 вне $[0, b]$. Тогда

$$\begin{aligned} G(t_{nn} + x) &= \int_{-b}^0 g(-u) dH(t_{nn} + x + u) = \\ &= \int_{-b}^0 g(-u) d(H(t_{nn} + x + u) - H(t_{nn})) \rightarrow \int_0^b g(u) \mu(x + du). \end{aligned}$$

По лемме 4 тот же предел будет существовать для $G(t_{nn} + y)$. Это значит, что мера $\mu(x + du)$ от x не зависит и, следовательно, $\mu([u, v))$ пропорционально длине интервала (u, v) :

$$\mu((u, v)) = c(v - u), \quad \mu(du) = c du.$$

Итак, мы доказали, что

$$G(t_{nn} + x) \rightarrow c \int_0^{\infty} g(u) du \quad (8)$$

для любой непрерывной функции, равной 0 вне $[0, b]$. Но для любой интегрируемой по Риману на $[0, b]$ функции g и заданного $\varepsilon > 0$ найдутся непрерывные функции g_1 и g_2 : $g_1 < g < g_2$, равные 0 вне $[0, b + 1]$ и такие, что $\int_0^b (g_2 - g_1) du < \varepsilon$. Это означает, что сходимость (8) справедлива также для любой интегрируемой по Риману функции, равной 0 вне $[0, b]$.

Рассмотрим теперь произвольную непосредственно интегрируемую функцию g . В силу свойства 2) этих функций (см. определение 9.4) можно так выбрать $b > 0$, что для функции

$$g_{(b)}(u) = \begin{cases} g(u) & \text{при } u \leq b, \\ 0 & \text{при } u > b, \end{cases}$$

правая и левая части в (8) будут сколь угодно мало отличаться от таких же выражений, соответствующих исходной функции g (для правой части это очевидно, для левой следует из того, что

$$\left| \int_0^t g(t-s) dH(s) - \int_0^t g_{(b)}(t-s) dH(s) \right| = \left| \int_0^{t-b} g(t-s) dH(s) \right| \leq \sum_{k>b-1} (c_1 + c_2)g_k \rightarrow 0$$

при $b \rightarrow \infty$ (см. лемму 9.3)). Поэтому (8) доказано для любой непосредственно интегрируемой функции g . Полагая $g = 1 - F$, мы получим в силу леммы 1

$$1 = c \int_0^{\infty} (1 - F(u)) du = ac, \quad c = \frac{1}{a}.$$

Таким образом, предел в (8) один и тот же для любой исходной последовательности t_n . Отсюда следует, что при $t \rightarrow \infty$

$$G(t) \rightarrow \frac{1}{a} \int_0^{\infty} g(u) du. \quad \blacksquare$$

Теорема 9.4 является простым следствием теоремы 1 и рассмотрений, использованных в доказательстве теоремы 9.2А, в которой основная теорема восстановления в решетчатом случае обобщалась на случай разнозначных ξ_j , $j \geq 2$ и произвольного ξ_1 . Мы предоставляем читателю повторить эти рассуждения применительно к нерешетчатому случаю.

Приведем еще ряд следствий теоремы 1. В § 9.4 мы получили уточнение теоремы восстановления в случае, когда $m_2 = E\xi_j^2 < \infty$. Подходы, развитые в доказательстве теоремы 1, позволяют получить альтернативное доказательство следующего утверждения, совпадающего с теоремой 9.4А.

Теорема 2. Пусть выполнены условия теоремы 1 и $m_2 < \infty$. Тогда

$$0 \leq H(t) - \frac{t}{a} \rightarrow \frac{m_2}{2a^2}.$$

Доказательство. Функция $G(t) = H(t) - t/a$ является решением уравнения восстановления (2), соответствующего функции

$$g(t) = \frac{1}{a} \int_t^\infty (1 - F(u)) du.$$

Так как g непосредственно интегрируема, то

$$G(t) \rightarrow \frac{1}{a} \int_0^\infty \int_v^\infty (1 - F(u)) du dv = \frac{m_2}{2a^2}.$$

Теорема 3 (локальная теорема восстановления). Предположим, что F имеет плотность $f = F'$, и эта плотность непосредственно интегрируема. Тогда H имеет плотность $h = H'$,

$$h(t) \rightarrow \frac{1}{a} \text{ при } t \rightarrow \infty.$$

Доказательство. Обозначим $f_n(x)$ плотность $S_n = \xi_1 + \dots + \xi_n$. Имеем

$$h(t) = H'(t) = \sum_{n=1}^\infty f_n(t) = f(t) + \int h(t-u)f(u) du = f(t) + h * F(t).$$

Это значит, что $h(t)$ удовлетворяет уравнению восстановления с функцией $g = f$. Поэтому в силу теоремы 1

$$h(t) \rightarrow \frac{1}{a} \int_0^\infty f(u) du = \frac{1}{a}.$$

Рассмотрим теперь некоторые обобщения теоремы 1. Функцию g , определенную на $(-\infty, \infty)$, мы будем называть непосредственно интегрируемой, если непосредственно интегрируемы функции $g(t)$ и $g(-t)$, $t \geq 0$.

Теорема 4. Если выполнены условия теоремы 1 и g непосредственно интегрируема, то

$$G(t) = \int_0^\infty g(t-u)H(du) \rightarrow \frac{1}{a} \int_{-\infty}^\infty g(u) du.$$

Доказательство получается незначительными и вполне очевидными модификациями рассуждений в доказательстве теоремы 1, главное из которых состоит в том, чтобы вместо функций g , равных нулю вне $[0, b]$, надо теперь рассматривать функции, обращающиеся в 0 вне $[-b, b]$.

Другое обобщение касается второй версии функции восстановления

$$U(t) = \sum_{k=0}^\infty F^{*k}(t), \quad -\infty < t < \infty,$$

в случае разнозначных ξ_j .

Теорема 5. Если g непосредственно интегрируема, $E\xi_j = a > 0$, то

$$G(t) = \int_{-\infty}^{\infty} g(t-u)U(du) \rightarrow \frac{1}{a} \int_{-\infty}^{\infty} g(u) du \quad \text{при } t \rightarrow \infty,$$

$U(t+u) - U(t) \rightarrow 0$ при $t \rightarrow -\infty$ и любом фиксированном u .

Доказательство также получается путем некоторой модификации доказательства теоремы 1 (см. также [18]).

Приложение 7

Таблицы

Таблица 1. Нормальное распределение

Таблица значений функции $\Phi_0(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-t^2/2} dt$. Функция нормального распределения $\Phi(x)$ связана с $\Phi_0(x)$ соотношениями $\Phi(x) = 1 - \Phi_0(x) = \Phi_0(-x)$.

x	0	1	2	3	4
0,0	0,5000	0,4960	0,4920	0,4880	0,4840
0,1	4602	4562	4522	4483	4443
0,2	4207	4168	4129	4090	4052
0,3	3821	3783	3745	3707	3669
0,4	3446	3409	3372	3336	3300
0,5	3085	3050	3015	2981	2946
0,6	2743	2709	2676	2643	2611
0,7	2420	2389	2358	2327	2297
0,8	2119	2090	2061	2033	2005
0,9	1841	1814	1788	1762	1736
1,0	1587	1562	1539	1515	1492
1,1	1357	1335	1314	1292	1271
1,2	1151	1131	1112	1093	1075
1,3	0968	0951	0934	0918	0901
1,4	0808	0793	0778	0764	0749
1,5	0668	0655	0643	0630	0618
1,6	0548	0537	0526	0516	0505
1,7	0446	0436	0427	0418	0409
1,8	0359	0351	0344	0336	0329
1,9	0288	0281	0274	0268	0262
2,0	0228	0222	0217	0212	0207
2,1	0179	0174	0170	0166	0162
2,2	0139	0136	0132	0129	0125
2,3	0107	0104	0102	0099	0096
2,4	0082	0080	0078	0075	0073
2,5	0062	0060	0059	0057	0055
2,6	0047	0045	0044	0043	0041
2,7	0035	0034	0033	0032	0031
2,8	0026	0025	0024	0023	0023
2,9	0019	0018	0018	0017	0016
x = $\Phi_0(x)$ =	3,0 0,0013	3,1 0,0010	3,2 0,0007	3,3 0,0005	3,4 0,0003

Окончание таблицы 1

x	5	6	7	8	9
0,0	0,4801	0,4761	0,4721	0,4681	0,4641
0,1	4404	4364	4325	4286	4247
0,2	4013	3974	3936	3897	3859
0,3	3632	3594	3557	3520	3483
0,4	3264	3228	3192	3156	3121
0,5	2912	2877	2843	2810	2776
0,6	2578	2546	2514	2483	2451
0,7	2266	2236	2206	2177	2148
0,8	1977	1949	1922	1894	1867
0,9	1711	1685	1660	1635	1611
1,0	1469	1446	1423	1401	1379
1,1	1251	1230	1210	1190	1170
1,2	1056	1038	1020	1003	0985
1,3	0885	0869	0853	0838	0823
1,4	0735	0721	0708	0694	0681
1,5	0606	0594	0582	0571	0559
1,6	0495	0485	0475	0465	0455
1,7	0401	0392	0384	0375	0367
1,8	0322	0314	0307	0301	0294
1,9	0256	0250	0244	0239	0233
2,0	0202	0197	0192	0188	0183
2,1	0158	0154	0150	0146	0143
2,2	0122	0119	0116	0113	0110
2,3	0094	0091	0089	0087	0084
2,4	0071	0069	0068	0066	0064
2,5	0054	0052	0051	0049	0048
2,6	0040	0039	0038	0037	0036
2,7	0030	0029	0028	0027	0026
2,8	0022	0021	0021	0020	0019
2,9	0016	0015	0015	0014	0014
$x =$ $\Phi_0(x) =$	3,5 0,0002	3,6 0,0002	3,7 0,0001	3,8 0,0001	3,9 0,0000

Таблица 2. Распределение Пуассона

Таблица значений функции $\Pi(x) = \sum_{k=x}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda}$.

	Параметр распределения λ					
x	0,1	0,2	0,3	0,4	0,5	0,6
1	0,0952	0,1813	0,2592	0,3297	0,3935	0,4512
2	0047	0175	0369	0616	0902	1219
3	0002	0011	0036	0079	0144	0232
4	0000	0001	0003	0008	0018	0034
5		0000	0000	0001	0002	0004
6				0000	0000	0000

Продолжение таблицы 2

	Параметр распределения λ					
x	0,7	0,8	0,9	1,0	1,2	1,4
1	0,5034	0,5507	0,5934	0,6321	0,6988	0,7534
2	1558	1912	2275	2642	3374	4082
3	0341	0474	0629	0803	1205	1665
4	0058	0091	0135	0190	0338	0537
5	0008	0014	0023	0037	0077	0143
6	0001	0002	0003	0000	0015	0032
7					0003	0006
8					0000	0001
9						0000

Продолжение таблицы 2

	Параметр распределения λ					
x	1,6	1,8	2,0	2,2	2,4	2,6
1	0,7981	0,8347	0,8647	0,8892	0,9093	0,9257
2	4751	5372	5940	6454	6916	7326
3	2166	2694	3233	3773	4303	4816
4	0788	1087	1429	1806	2212	2640
5	0237	0364	0527	0725	0859	1226
6	0060	0104	0166	0249	0357	0490
7	0013	0026	0045	0075	0116	0172
8	0003	0006	0011	0020	0033	0053
9	0000	0001	0002	0005	0009	0015
10		0000	0000	0001	0002	0004
11				0000	0000	0001
12						0000

Продолжение таблицы 2

	Параметр распределения λ					
x	2,8	3,0	4,0	5,0	6,0	7,0
1	0,9392	0,9502	0,9817	0,9933	0,9975	0,9991
2	7689	8009	9084	9596	9826	9227
3	5305	5768	7619	8753	9380	9704
4	3081	3528	5665	7350	8488	9182
5	1523	1847	3712	5595	7149	8270
6	0651	0839	2149	3840	5543	2993
7	0244	0335	1107	2378	3937	5503
8	0081	0119	0511	1334	2560	4013
9	0024	0038	0214	0681	1528	2709
10	0007	0011	0081	0318	0839	1695
11	0001	0003	0028	0137	0426	0985
12	0000	0001	0009	0055	0201	0534
13			0002	0020	0088	0270
14			0001	0007	0036	0128
15			0000	0002	0014	0057
16				0001	0005	0024
17				0000	0002	0010
18					0001	0004
19					0000	0001
20						0000

Продолжение таблицы 2

	Параметр распределения λ					
x	8,0	9,0	10,0	11,0	12,0	13,0
1	0,9997	0,9999	1,0000	1,0000	1,0000	1,0000
2	9970	9988	0,9995	0,9998	0,9999	1,0000
3	9862	9938	9972	9988	9995	0,9998
4	9576	9788	9897	9950	9977	9990
5	9004	9450	9707	9849	9924	9963
6	8088	8843	9329	9625	9797	9893
7	6866	7932	8699	9214	9542	9741
8	5470	6761	7798	8568	9105	9460
9	4075	5443	6672	7680	8450	9002

Окончание таблицы 2

x	Параметр распределения λ					
	8,0	9,0	10,0	11,0	12,0	13,0
10	2834	4126	5421	6594	7576	8342
11	1841	2940	4170	5401	6528	7483
12	1119	1967	3032	4207	5384	6468
13	0638	1242	2084	3113	4240	5369
14	0342	0739	1355	2187	3185	4270
15	0173	0415	0835	1460	2280	3249
16	0082	0220	0487	0926	1556	2364
17	0037	0111	0270	0559	1013	1645
18	0016	0053	0143	0322	0630	1095
19	0007	0024	0072	0177	0374	0698
20	0003	0011	0035	0093	0213	0427
21	0001	0004	0016	0047	0116	0250
22	0000	0002	0007	0023	0061	0141
23		0001	0003	0010	0030	0076
24		0000	0001	0005	0015	0040
25			0000	0002	0007	0020
26				0001	0003	0010
27				0000	0001	0005
28					0001	0002
29					0000	0001

Примечание. При $\lambda > 13$ вероятность $\Pi(x)$ можно вычислять приближенно по формуле (ср. с замечаниями в § 4 гл. 5)

$$\Pi(x) \approx \Phi_0\left(\frac{x - 0,5 - \lambda}{\sqrt{\lambda}}\right).$$

Более подробные таблицы можно найти, например, в [2, 8].

Список литературы

1. *Биллингсли П.* Сходимость вероятностных мер. М.: Наука, 1977.
2. *Большев Л. Н., Смирнов Н. В.* Таблицы математической статистики. М.: Наука, 1965.
3. *Боровков А. А.* Вероятностные процессы в теории массового обслуживания. М.: Наука, 1972.
4. *Боровков А. А.* Сходимость мер и случайных процессов // Успехи математических наук, 1976. Т. 31, № 2, стр. 3–68.
5. *Боровков А. А.* Теория вероятностей. 2-е изд., перераб. и доп. М.: Наука, 1986.
6. *Боровков А. А.* Эргодичность и устойчивость случайных процессов. М.: УРСС, 1998.
7. *Гихман И. И., Скороход А. В.* Введение в теорию случайных процессов. М.: Наука, 1977.
8. *Гнеденко Б. В.* Курс теории вероятностей. М.: Наука, 1969.
9. *Гнеденко Б. В., Колмогоров А. Н.* Предельные распределения для сумм независимых случайных величин. М.: Гостехиздат, 1949.
10. *Гренадер У.* Вероятности на алгебраических структурах. М.: Мир, 1965.
11. *Ибрагимов И. А., Линник Ю. В.* Независимые и стационарно связанные величины. М.: Наука, 1960.
12. *Колмогоров А. Н.* Цепи Маркова со счетным числом состояний // Бюллетень МГУ. 1937. Т. 1, № 3.
13. *Колмогоров А. Н.* Теория вероятностей. Математика, ее содержание, методы и значение. Т. 2. М.: Изд-во АН СССР, 1956.
14. *Крамер Г., Лидбеттер М.* Стационарные случайные процессы. М.: Мир, 1969.
15. *Лампурти Дж.* Вероятность. М.: Наука, 1973.
16. *Лозе М.* Теория вероятностей. М.: Иностранная литература, 1962.
17. *Натансон И. П.* Теория функций вещественной переменной. М.: Наука, 1974.
18. *Нумеллин Э.* Общие неприводимые цепи Маркова и неотрицательные операторы. М.: Мир, 1989.
19. *Петров В. В.* Суммы независимых случайных величин. М.: Наука, 1972.
20. *Скороход А. В.* Случайные процессы с независимыми приращениями. М.: Наука, 1986.
21. *Файнштейн А.* Основы теории информации. М.: Иностранная литература, 1960.
22. *Феллер В.* Введение в теорию вероятностей и ее приложения. Т. 1, 2. М.: Мир, 1984.
23. *Халмош П.* Теория меры. М.: Иностранная литература, 1953.

24. *Хинчин А. Я.* Понятие энтропии в теории вероятностей // Успехи математических наук. 1953. Т. 8.
25. *Ширяев А. Н.* Вероятность. М.: Наука, 1980.
26. *Kifer Yu.* Ergodic Theory of Random Transformations. Boston, Birkhauser, 1986.
27. *Meyn S. P., Tweedie R. L.* Markov chains and stochastic stability. N.-Y., Berlin, Springer Verlag, 1993.
28. *Van Beek P.* An application of Fourier methods to the problem of sharpening the Berry // Esseen inequality. Z. Wahrscheinlichkeitstheorie und Verw. 1972. B. 23, № 3. S. 187–196.

Предметный указатель

Абсолютная непрерывность меры, 437
Абсолютно непрерывное распределение, 44, 48

Авторегрессия обобщенная, 344

Алгебра (множеств), 23

Безобидная игра, 70

Бернулли схема, 18, 96

Бернштейна полиномы, 97

Большие отклонения, 104, 112

Борелевская функция, 46

Борелевское множество, 24

— поле, 24

Бореля—Кантелли лемма (критерий), 213

Броуновский мост, 413

Вероятность, 14, 25

—, классическое определение, 15

— переходная, 267, 387

— условная, 30

Выборка, 15

Генеральная совокупность, 15

Дефект случайного блуждания, 204

Дисперсия, 76

Закон больших чисел, 81, 155, 157, 210

— — — усиленный, 96, 213, 219, 220, 223, 229

— нуля и единицы, 213–215

— повторного логарифма, 360, 361, 378, 379

Измеримое отображение, 37

— пространство, 24

Индикатор множества, 427

Интеграл, 57, 427

— от произвольной функции, 428

— по множеству, 427

Информации количество, 284

Кольцо (множеств), 23

Корреляции коэффициент, 78

Критерий Колмогорова, 223, 351

Локальная предельная теорема, 97, 174

Маркова цепь, 245

— — возвратная, 253

— — однородная, 267

— — разложимая (неразложимая), 248, 249

— — счетная (дискретная), 245

Мартиггал, 292–296, 300

Математическое ожидание, 64

— — условное, 68, 84, 85

Матрица дважды стохастическая, 262

— стохастическая, 247

Мера, 426

— сингулярная, 437

Момент, 79

— остановки, 73, 268, 296

— — несобственный, 299

- смешанный, 80
- центральный, 79

Независимые испытания, 32

- классы событий, 52
- случайные величины, 50, 55, 130
- события, 30

Неравенство Гёльдера, 80

- Иенсена, 80, 88
- Колмогорова, 217, 308
- Коши—Буняковского, 80, 88

— Минковского, 80**— Чебышева, 81, 87****Носитель меры, 437****Обслуживания системы, 234, 394****Плотная последовательность распределений, 127****Плотность, 443**

- меры, 435
- переходная, 388
- распределения, 45, 48, 62
- условная, 90

Последовательность верхняя (нижняя), 216**— событий обновляющая, 333****Преобразование Крамера, 181, 304****— Лапласа, 132****Принцип инвариантности, 371, 372, 377****Прогноз, 414, 415****Произведение мер прямое, 60****Производная Радона—Никодима, 435, 438****Производящая функция, 135, 151****Пространство вероятностное, 26**

- выборочное, 41, 48, 265, 348, 442
- измеримое, 24, 27
- с мерой, 426

- элементарных событий (исходов), 13, 23, 25

Процесс восстановления, 195

- — однородный, 200, 209
- гауссовский, 413
- диффузионный, 404
- марковский, 385, 386, 388
- непредсказуемый, 410
- непрерывный в среднем, 355
- пуассоновский, 209, 364, 367, 381, 393
- — обобщенный, 366
- с независимыми приращениями, 357
- сепарабельный, 354
- случайный, 347, 349, 442
- стационарный в узком смысле, 412
- — — широко —, 411
- стохастически непрерывный, 355

Процессы ветвящиеся, 151**Псевдомомент, 171****Равномерная интегрируемость, 118****Распределение Г (гамма-распределение), 147**

- χ^2 (хи-квадрат), 148
- абсолютно непрерывное, 44
- безгранично делимое, 188, 357
- биномиальное, 18, 42
- вероятностей, 14, 26
- вырожденное, 41
- геометрическое, 43
- гипергеометрическое, 17, 19
- дискретное, 44
- конечномерное, 348
- Коши, 43, 133
- нормальное, 42, 132
- — многомерное, 49, 92, 145
- показательное, 42, 133, 148
- полиномиальное, 49
- процесса, 348, 349, 367
- Пуассона, 33, 43

- равномерное, 38, 42, 133, 262
- — дискретное, 15
- симметричное, 237
- сингулярное, 45, 222
- случайной величины, 37, 41, 47
- стационарное (инвариантное), 204, 235, 257, 269
- условное, 89
- устойчивое, 188
- Эрланга, 148

Свертка, 62

Симметрическая разность, 418

Случайная величина, 37

- — марковская, 73, 268
- — многомерная, 47
- —, не зависящая от будущего, 72
- — несобственная, 38
- — нормированная, 78
- — остаточная, 215

Случайное блуждание, 195, 252, 254

Случайные величины, положительно (отрицательно) коррелированные, 79

Случайный вектор, 47

Событие, 13

- дополнительное (противоположное), 14, 25
- достоверное, 14, 25
- невозможное, 25
- обновляющее, 333
- остаточное, 214
- перестановочное, 215

Событий произведение, 13

- разность, 13
- сумма, 13

События независимые, 30, 31

— несовместные, 14, 25

Состояние возвратное (невозвратное), 249, 250

- нулевое (ненулевое), 249, 250
- поглощающее, 248

— существенное (несущественное), 247

Сочетаний число, 16

Спектральная мера процесса, 369

Среднее значение, 64

Стандартное отклонение, 76

Стохастическая матрица, 247

Стохастически эквивалентные процессы, 350

Суб(супер)мартингал, 292, 293

Схема Бернулли, 96

Сходимость по вероятности, 114

- — мере, 427
- — распределению, 123
- почти наверное (почти всюду, с вероятностью 1), 114, 213, 426
- слабая, 121, 144, 441

Теорема аппроксимации, 53

- Берри—Эссена, 174, 449
- Бохнера—Хинчина, 133
- Вейерштрасса, 97
- восстановления, 197, 199, 201, 458
- Дуба, 297, 308, 311, 354
- единственности, 136
- Каратеодори, 28, 421
- Колмогорова, 57, 214, 220, 314
- Лебега, 438
- Леви, 313
- Муавра—Лапласа, 102
- непрерывности, 118, 137, 145
- о монотонной сходимости, 87, 430, 437
- — продолжении меры, 28, 419
- — сходимости мажорируемой последовательности, 120, 430, 437
- о повторном интегрировании, 61
- Прохорова, 73
- Пуанкаре, 324
- Пуассона, 141
- Радона—Никодима, 437
- Фату—Лебега, 431
- Фубини, 432

- Хана, 439
- Хелли, 126, 446, 448
- центральная предельная, 156, 164, 165, 177
- эргодическая, 257, 412
- Тождество Вальда, 74, 280, 296, 302, 303
- Поллачека—Спитцера, 238
- факторизационное, 225, 230, 235

Уравнения Колмогорова, 391, 406

Условие Крамера, 179

— Линдеберга, 157, 164, 404

— Ляпунова, 166, 167

Условная вероятность, 30

— функция распределения, 68

Условное математическое ожидание, 68, 82, 85

Факторизации компоненты, 229

Факторизационные тождества, 225, 230

Формула Байеса, 35

— обращения, 135, 144

— полной вероятности, 32, 68, 88

— Стирлинга, 98, 99

Функция борелевская, 46

— верхняя (нижняя), 213, 361

— восстановления, 196

— ковариационная, 410

— неотрицательно определенная, 410, 411

— переходная, 267, 387, 388

— производящая, 135, 151

— простая, 427

— распределения, 37, 39, 47

— — условная, 68

— уклонений, 181

— характеристическая, 130, 143

Центральная предельная теорема, 156, 165, 177

Цепь Маркова, 245

— — счетная (дискретная), 245

Цилиндрическое множество, 348, 349, 424

Экссесс случайного блуждания, 197

Элементарные события (исходы), 13

Энтропия, 285

— условная, 287

— цепи Маркова, 288

σ -алгебра, 24

σ -алгебра борелевская, 24

σ -алгебра, порожденная классом множеств, 24

σ -алгебра, порожденная случайной величиной, 54, 56