Введение в теорию вероятностей: конспект лекций

В.И. Богачев

мат. фак. ВШЭ, 3 курс, осень 2017

Оглавление

Предисловие

Вводный курс теории вероятностей важен для последующего изучения элементов математической статистики и некоторых других дисциплин (теории защиты информации, некоторых курсов физики и т. д.). Несомненно, этот курс полезен для расширения кругозора всем математикам независимо от их основной области. Наконец, освоение этого введения необходимо для практической работы в таких направлениях, как прикладная статистика, финансы, лингвистика, не говоря уже о статистических методах большинства областей естествознания. При этом оказывается разумной и несколько более высокая по сравнению с реальными приложениями абстракция такого курса, ибо иначе он вместе с курсом статистики рискует стать похожим на кулинарную книгу. Разумеется, эта повышенная абстракция тоже должна иметь разумные пределы, ибо здесь несомненно возникает опасность превращения курса теории вероятностей в утонченного дублера теории меры Лебега. В данном семестровом курсе нужный баланс достигался переносом на семинары рассмотрений конкретных примеров и задач.

Первым опубликованным трактатом по теории вероятностей считается сочинение голландца Х. Гюйгенса (1629–1695) «О расчетах в азартных играх» (1657), появившееся под влиянием дискуссии французов Б. Паскаля (1623–1662) и П. Ферма (1601–1665) о расчете шансов при игре в кости, известной с глубокой древности. Первым систематическим изложением теории стала работа швейцарца Я. Бернулли (1655–1705) «Искусство предположений», опубликованная в полном виде в 1713 году, но еще до этого получившая широкое распространение. В том же XVIII веке Байес (1702–1761), английский священник и математик, открыл свою знаменитую формулу. Еще одно классическое произведение — книга английского математика французского происхождения А. де Муавра (1667–1754) «Учение о случаях» (1718).

В работах де Муавра появились первые варианты центральной предельной теоремы и (в неявном виде) нормальное (гауссовское) распределение. Позже французский математик П.-С. Лаплас (1749–1827) значительно развил эти исследования. В XIX веке выдающийся вклад в развитие теории вероятностей внесли также французский математик С.-Д. Пуассон (1781–1840), немецкий математик К.Ф. Гаусс, русские математики П.Л. Чебышёв (1821–1894) и А.А. Марков (1856–1922), а английский математик К. Пирсон (1857–1936) заложил основы математической статистики. Хотя все упомянутые ученые были известны именно как выдающиеся математики (помимо них интересовались вероятностными и статистическими проблемами и другие знаменитые математики, например Л. Эйлер (1707–1783), О. Коши (1789–1857)), теория вероятностей все же не считалась вполне математической наукой. В частности, знаменитый список проблем Гильберта, поставленных в 1900 году, включал аксиоматизацию теории вероятностей. Общепризнанное теперь решение этой проблемы на основе теории меры было предложено крупнейшим математиком XX века А.Н. Колмогоровым (1903–1987) в его компактной монографии „Основные понятия теории вероятностей", которую полезно хотя бы пролистать всякому изучающему этот предмет.

За полтора века существования теории вероятностей как учебного предмета в университетских программах было издано огромное число учебных пособий по ней, различающихся объемом, методическими принципами и предполагаемой аудиторией. Чрезвычайно велико число таких пособий и на русском языке. В этом легко убедиться поиском в интернете. Поэтому отмечу лишь следующие книги:

Ламперти Дж. Вероятность. Наука, М., 1973 — хорошо написанное краткое пособие для математиков, первая половина близка данному конспекту,

Кораллов Л.Б., Синай Я.Г. Теория вероятностей и случайные процессы. МЦНМО, М., 2013 — первая половина книги содержит компактный интенсивный курс теории вероятностей (более объемный, чем семестровый),

Прохоров Ю.В., Пономаренко Л.С. Лекции по теории вероятностей и математической статистике. Изд-во МГУ, М., 2012 — первая половина может служить конспектом семестрового курса теории вероятностей,

Ширяев А.Н. Вероятность. Т. 1, 2. 3-е изд. МЦМНО, М., 2004 — весьма обстоятельное и известное современное пособие, содержащее обширный материал по многим вопросам, в том числе не входящим в обязательные программы, Феллер В. Введение в теорию вероятностей и ее приложения. Т. 1, 2. Мир, М., 1984 — перевод классического трактата одного из создателей современной теории вероятностей, написанного более полувека назад, но не устаревшего и содержащего решения множества конкретных задач, включая счетно-комбинаторные,

Гнеденко Б.В. Курс теории вероятностей. 4-е изд. Наука, М., 1965 — классический учебник, написанный в период интенсивного создания современной теории более полувека назад активным участником и хорошо передающий дух теории вероятностей,

Тутубалин В.Н. Теория вероятностей. Академия, М., 2008 — интересное и полезное дополнительное чтение к базовому курсу, в том числе дающее впечатление о прикладной стороне дела и соотношении ее с теоретической.

Несколько известных задачников:

Ширяев А.Н. Задачи по теории вероятностей. МЦМНО, М., 2006,

Зубков А.М., Севастьянов Б.А., Чистяков В.П. Сборник задач по теории вероятностей. 2-е изд. Наука, М., 1989,

Кельберт М.Я., Сухов Ю.М. Вероятность и статистика в примерах и задачах. Т.1. МЦНМО, М., 2007,

Мешалкин Л.Д. Сборник задач по теории вероятностей. Изд-во МГУ, М., 1963,

Прохоров А.В., Ушаков В.Г., Ушаков Н.Г. Задачи по теории вероятностей. Наука, М., 1986.

Все эти книги следуют колмогоровской аксиоматике, в частности используют меру и интеграл Лебега. Однако, вопреки распространенному заблуждению, необходимая для введения в недискретной теории вероятностей идеология и техника лебеговского интегрирования может быть изложена довольно кратко и оказывается ничуть не сложнее известной из первого курса математического анализа. Применительно к основным приложениям по своей сути метод Лебега представляет собой удачный способ единообразного обращения с рядами и интегралами, причем способ, гораздо более гибкий, чем основанный на обычном интеграле Римана. Полезно также полистать старые классические курсы теории вероятностей, имевшие широкое распространение до появления колмогоровской аксиоматики:

Чебышев П.Л. Теория вероятностей. Лекции академика П.Л. Чебышева, читанные в 1879-80 гг. По записи А.М. Ляпунова; изданы акад. А.Н. Крыловым; АН СССР. Изд-во АН СССР, М.-Л., 1936,

Бернштейн С.Н. Теория вероятностей. Гос. издат., М.-Л., 1927.

Зачем читают лекции по теории вероятностей и зачем пишут конспекты? Почему бы не предложить студентам самим почитать какой-либо учебник и сдать экзамен? Хотя последнее в принципе возможно и для отдельных студентов даже может быть более полезным, многолетняя практика убедительно показывает, что для большинства учащихся самым рациональным является обсуждение на лекциях и семинарах всяких разных мелочей, мимо которых легко пройти при самостоятельном чтении. Поэтому полезные лекции должны существенно отличаться и от учебника, и от даже лекторского конспекта самих этих лекций. По этой же причине чужой конспект или конспект лектора не могут заменить ни своего собственного, даже коряво написанного, ни, главное, активного участия в обсуждении предмета на лекциях и семинарах.

Глава 1

Вероятностные пространства и случайные величины

Курс теории вероятностей предлагает математические основы изучения явлений и количественных характеристик, которые в соответствии с общепринятыми представлениями характеризуются как случайные. Понятие случайного не формализуется, но обычно таковым считается то, что не имеет (по крайней мере на данный исторический момент) детерминированного поведения или может иметь множественные исходы, которые наблюдатель не может предсказать. С самого начала важно иметь в виду, что отсутствие формализации понятия случайного никоим образом не препятствует разработке эффективных вероятностных методов анализа таких явлений. Более того, даже в тех ситуациях, про которые заведомо известно, что они вполне детерминированы и никакой случайности в себе не таят, бывает полезно применение вероятностных подходов. Например, так бывает при анализе явлений и событий, подверженных влиянию очень многих факторов, каждый из которых играет ничтожно малую роль. Именно такие явления представляют собой предмет математической статистики (теоретической и прикладной), курс которой примыкает к вводному курсу теории вероятностей (а во многих учебных планах и просто в него включается). Сама же теория вероятностей имеет дело с весьма идеализированными моделями реальности, которые, тем не менее, находят довольно хорошее подтверждение на практике.

Такая практика насчитывает уже многие столетия, а сам ее анализ математическими методами (т. е. собственно теория вероятностей или теория шансов) ведется уже более 300 лет. Эта практика наложила заметный отпечаток на базовые концепции теории вероятностей. Уже само понятие вероятности как некоторого числа между нулем и единицей, иногда трактуемое как «шанс» и измеряемое в процентах (чего обычно не делается в теоретических построениях), отражает опыт измерения частот. Например, классический пример — бросание монеты, падающей орлом или решкой вверх, когда после N бросаний появляются два числа из $[0,1]$: N_1/N и N_2/N, где N_1 и N_2 — количества выпавших орлов и решек соответственно (в этой модели считается, что иных исходов нет, скажем, монета не может стать на ребро, в отличие от известного студенческого теста «идти ли на лекцию», решаемого бросанием монеты). Конечно, из здравого смысла понятно, что при очень большом числе бросаний N можно ожидать, что N_1/N и N_2/N должны мало отличаться в случае монеты без смещений. Но вот уже и теоретический вопрос (который несколько столетий назад был практическим для азартных игроков в эту некогда популярную игру): сколь велико должно быть расхождение между N_1/N и N_2/N, чтобы заподозрить какое-либо смещение центра тяжести монеты? То, что искусственное смещение очень даже возможно, показывают найденные при раскопках древних городов игральные кости и монеты с несомненно намеренно сделанными впайками, нарушающими естественное центрирование. Здесь уместно сказать, что азартные игры (игральные кости, карты, монеты для бросания и т.п.) оказали очень заметное влияние на возникновение и развитие теории вероятностей. При этом созданные для их анализа методы и приемы оказались полезны далеко за пределами их сомнительной области происхождения. Позже более важными источниками вероятностных задач и моделей стали вопросы анализа и учета больших данных (демография, медицина, страховое дело, экономическая статистика и т.д.). В настоящее время приложения теории вероятностей и статистики имеются во многих областях помимо «прямого наследника» истоков — анализа финансовых рынков («финансовой математики»): экономике, медицине, биологии, физике, страховом деле, лингвистике и т. д.

§ 1.1. Вероятностные пространства

Одной из первых возникших вероятностных моделей была модель N-кратного бросания симметричной двусторонней монеты с измерением количеств выпавших орлом или решкой сторон. В приближенном к современному описанию речь идет о наборах длины N из нулей и единиц. Симметрия монеты означает, что все 2^N комбинаций равновероятны, т. е. имеют шанс осуществления $1/2^N$. Такие наборы считают «элементарными исходами» и им приписывают вероятности 2^{-N}, а их различные сочетания считаются «событиями». Например, событием является «превышение числом орлов числа решек». Скажем, при четырех бросаниях ($N = 4$) элементарные события — четверки типа

$(0,1,0,1)$ с вероятностями $1/16$, а упомянутое событие «превышения числа орлов» объединяет элементарные события с тремя или четырьмя нулями, т. е. состоит из пяти точек (единица может быть на одном из четырех мест или вообще отсутствовать) и потому имеет шанс осуществиться $5/16$.

Более общим образом, конечное вероятностное пространство Ω состоит из n точек $\omega_1, \dots, \omega_n$, имеющих вероятности $p_n \in [0,1]$ (необязательно равные) с условием $p_1 + \dots + p_n = 1$. Эти точки считаются элементарными событиями, а все прочие события — какие-либо их объединения, причем вероятность события A задается формулой

$$P(A) = \sum_{i\colon\, \omega_i \in A} p_i.$$

Даже в простейшей ситуации бросания симметричной монеты с числом бросаний N возникает очень большое число 2^N точек пространства Ω при не таких уж больших N (порядка десятков). Поэтому по указанной формуле бывает трудно или даже невозможно найти вероятность какого-либо анализируемого неэлементарного события. Из решения типовых задач будет видно, что обычно используются всякие специальные соображения, позволяющие иным способом находить вероятности сложных событий. Правда, надо еще отметить, что нередко проблемой бывает и правильный выбор вероятностного пространства, адекватный задаче (многие вероятностные задачи изначально формулируются без какой-либо апелляции к конкретным вероятностным пространствам).

Один из весьма употребительных приемов расчета состоит в значительном уменьшении вероятностного пространства путем использования так называемых условных вероятностей и формулы полной вероятности. Формальная конструкция такова. Пусть дано множество $B \subset \Omega$ с $P(B) > 0$. Тогда его можно превратить в отдельное вероятностное пространство, рассматривая лишь события вида $A \cap B$ (т. е. части B) и приписывая им вероятности $P(A \cap B)/P(B)$, где деление на $P(B)$ необходимо для сохранения вероятности 1 для всего нового пространства. Эти вероятности

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

называют условными вероятностями событий A при условии осуществления события B. Основной их смысл состоит в том, что A может лишь частично входить в B.

Если Ω разбито на дизъюнктные части B_i с $P(B_i) > 0$, то такие условные вероятности возникают для всех B_i и приводят к так называемой *формуле полной вероятности*

$$P(A) = \sum_{i=1}^{n} P(A|B_i)P(B_i).$$

Эта формула совершенно тривиальна, так как

$$P(A) = P(A \cap B_1) + \cdots + P(A \cap B_i)$$

из-за дизъюнктности разбиения, так что остается лишь разделить и умножить каждое слагаемое на $P(B_i)$. Таким образом, фактически речь идет о вычислении большой суммы разбиением на части. Как ни странно, эта банальная на первый взгляд процедура вместе с некоторым искусством правильного выбора разбиения позволяет вычислять вероятности сложных событий в довольно неочевидных ситуациях (чем занимаются на семинарах).

Аналогично проверяется *формула Байеса*

$$P(B_k|A) = \frac{P(A|B_k)P(B_k)}{\sum_{i=1}^{n} P(A|B_i)P(B_i)}$$

для события A с $P(A) > 0$. Тривиальность обоснования таких формул не должна смущать: в теории вероятностей применение нескольких тривиальных формул и шагов часто приводит к совершенно неочевидным выводам.

Однако все эти понятия приходится модифицировать при рассмотрении континуальных вероятностных пространств, в которых точки имеют нулевые меры. Конечно, здесь тоже можно было бы считать точки элементарными событиями, а какие-то их объединения рассматривать как неэлементарные события, но при желании наделить такие события числовыми характеристиками (вероятностями) возникает проблема: в большинстве задач важно, чтобы область определения $\mathfrak{B}$ вероятностной меры допускала счетные операции объединения, пересечения и дополнения, т. е. была σ-алгеброй (что включает еще естественное требование, что $\Omega, \varnothing \in \mathfrak{B}$), а сама вероятностная мера $P\colon \mathfrak{B} \to [0,1]$ была счетно-аддитивной, т. е. удовлетворяла равенству

$$P(\cup_{n=1}^{\infty} B_n) = \sum_{n=1}^{\infty} P(B_n)$$

для всяких попарно непересекающихся множеств $B_n \in \mathfrak{B}$. Название «вероятностная» заключает в себе еще и требование $P(\Omega) = 1$.

Нельзя сказать, что перечисленные требования откуда-то вытекают, но долгая практика показывает, что обычно они весьма естественны и полезны. Однако комбинация этих двух требований (область определения выдерживает счетные операции и есть счетная аддитивность) практически исключает континуальные пространства, на которых вероятностные меры заданы на всех множествах и равны нулю на одноточечных множествах. Разумеется, без последнего ограничения проблемы нет: мера Дирака δ_a равна единице в фиксированной точке $a \in \Omega$ и нулю на ее дополнении, так что мера всякого множества A есть 1 или 0 в зависимости от того, входит ли в него точка a. Можно взять счетную сумму таких мер в разных точках с весами. Однако иного способа нет без специальных дополнительных теоретико-множественных предположений, тем более нет никакого конструктивного способа.

В итоге довольствуются следующей конструкцией. Обычно бывает можно построить меру со свойством счетной аддитивности на алгебре $\boldsymbol{\mathcal{A}}$ вместо σ-алгебры (в этом случае нужное равенство требуется для таких счетных объединений дизъюнктных множеств из $\boldsymbol{\mathcal{A}}$, что и их объединение лежит в $\boldsymbol{\mathcal{A}}$). Алгебра отличается от σ-алгебры тем, что допускает конечные операции вместо счетных. Вот типичный пример: обычная длина на $[0,1)$, заданная на конечных объединениях промежутков вида $[\alpha,\beta)$ в $[0,1)$. Конечно, ее счетная аддитивность требует проверки. После этого оказывается возможным продолжить P с сохранением свойства счетной аддитивности на минимальную σ-алгебру $\sigma(\boldsymbol{\mathcal{A}})$, содержащую $\boldsymbol{\mathcal{A}}$. Эта минимальная σ-алгебра формально допускает очень простое описание: надо пересечь все вообще σ-алгебры множеств отрезка, содержащие $\boldsymbol{\mathcal{A}}$ (такие есть, например, класс всех множеств). Однако конструктивного описания множеств из $\sigma(\boldsymbol{\mathcal{A}})$ нет, поэтому обоснование упомянутого результата о продолжении — довольно трудная задача (впервые решенная Лебегом именно для классической длины на промежутках). Кроме того, продолжение еще и единственно. Задается оно очень естественной явной формулой

$$P^*(B) = \inf\Bigl\{\sum_{n=1}^{\infty} P(A_n)\colon A_n \in \boldsymbol{\mathcal{A}},\ B \subset \bigcup_{n=1}^{\infty} A_n\Bigr\},$$

где inf берется по всем указанным покрытиям. В высшей степени неочевидно лишь, что эта формула дает решение. На самом деле, тот факт, что она дает решение, доказывается весьма окольным путем: вводится класс множеств $\boldsymbol{\mathcal{L}}$ с тем свойством, что

$$P^*(B) + P^*(\Omega\backslash B) = 1,$$

затем проверяется, что этот класс есть σ-алгебра, содержащая $\mathcal{A}$, причем функция P^* на $\mathcal{L}$ счетно-аддитивна. На классе всех множеств функция P^* может не быть счетно-аддитивной в предположении аксиомы выбора.

Наименьшая σ-алгебра подмножеств прямой, содержащая все открытые множества, называется борелевской σ-алгеброй и обозначается символом $\mathcal{B}(\mathbb{R})$. Аналогично вводится борелевская σ-алгебра $\mathcal{B}(\mathbb{R}^d)$ подмножеств $\mathbb{R}^d$.

Например, замкнутые множества являются борелевскими как дополнения открытых. Борелевскими будут и их счетные объединения. Эти построения можно усложнять, используя счетные операции, но явного описания борелевских множеств это не даст. Вряд ли можно вполне понять, что такое борелевские множества, скорее, к ним можно просто привыкнуть, как привыкают к вещественным числам, и тогда такая привычка будет казаться нам пониманием.

Тройка $(\Omega, \mathcal{B}, P)$, состоящая из пространства Ω с σ-алгеброй и вероятностной мерой P, и называется *вероятностным пространством*.

Важнейшим недискретным вероятностным пространством является отрезок $[0, 1]$ с обычной мерой Лебега, заданной на борелевской σ-алгебре или на σ-алгебре измеримых по Лебегу множеств (с помощью упомянутой выше внешней меры). В большинстве приложений можно считать, что вероятностное пространство есть отрезок, прямая или их счетная степень, а мера задана на борелевской σ-алгебре.

§ 1.2. Случайные величины

Напомним, что индикатор (или индикаторная функция) I_A множества A задается формулой

$$I_A(\omega) = 1, \text{ если } \omega \in A, \quad I_A(\omega) = 0, \text{ если } \omega \notin A.$$

1.2.1. Определение. *Функция $f\colon \Omega \to \mathbb{R}$ на пространстве Ω, наделенном σ-алгеброй $\mathcal{B}$, называется измеримой относительно $\mathcal{B}$ (или $\mathcal{B}$-измеримой), если множество $\{\omega\colon f(\omega) < c\}$ входит в $\mathcal{B}$ для каждого $c \in \mathbb{R}$.*

Функция на прямой, измеримая относительно борелевской σ-алгебры, называется борелевской.

Если $(\Omega, \mathcal{B}, P)$ — вероятностное пространство, то $\mathcal{B}$-измеримая функция называется случайной величиной (сокращенно с.в.). При этом случайной величиной называют также всякую такую функцию ξ, которая определена лишь на некотором множестве $\Omega_0 \in \mathcal{B}$ с $P(\Omega_0) = 1$ и на нем измерима в предыдущем смысле.

1.2.2. Лемма. *Приведенное определение измеримости равносильно более сильному свойству*:

$$f^{-1}(B) \in \mathcal{B} \quad \text{для всех борелевских множеств } B \subset \mathbb{R}.$$

Доказательство. Класс множеств $\mathcal{B}_1 = \{B \in \mathcal{B}(\mathbb{R}) : f^{-1}(B) \in \mathcal{B}\}$ является σ-алгеброй, что проверяется непосредственно. Поскольку для измеримой функции f он содержит все лучи, то $\mathcal{B}_1 = \mathcal{B}(\mathbb{R})$. □

Свойство, выполняющееся для всех точек вероятностного пространства, кроме точек множества меры нуль, называется справедливым *почти наверное* (п.н.) или *почти всюду* (п.в.).

1.2.3. Теорема. (i) *Если функция f измерима относительно $\mathcal{B}$, а функция φ на прямой измерима относительно борелевской σ-алгебры $\mathcal{B}(\mathbb{R})$, то композиция $\varphi(f)$ измерима относительно $\mathcal{B}$.*

(ii) *Если функции f и g измеримы относительно $\mathcal{B}$, то таковы и $\alpha f + \beta g$ для всех чисел α, β, а также fg. Если $g \neq 0$, то измерима и f/g.*

(iii) *Если последовательность $\mathcal{B}$-измеримых функций f_n сходится к функции f в каждой точке, то f измерима. Поэтому предел почти всюду сходящейся последовательности случайных величин является случайной величиной.*

Доказательство. Первое утверждение очевидно из леммы. Второе проверяется аналогично третьему, которое мы и докажем. Множество $\{\omega \colon f(\omega) < c\}$ есть объединение по натуральным k множеств $\Omega_k = \bigcup_{N=1}^{\infty} \bigcap_{n \geqslant N} \{\omega \colon f_n(\omega) < c - 1/k\}$, которые входят в $\mathcal{B}$. Второе утверждение сводится к первому на множестве с дополнением меры нуль (объединение множеств, где нет сходимости и не определены какие-то из рассматриваемых с.в.). □

Среди измеримых функций выделяются *простые* функции, т. е. принимающие лишь конечное множество значений. Такая функция может быть записана в виде линейной комбинации индикаторов

$$f = \sum_{i=1}^{n} c_i I_{B_i}, \quad \text{где} \quad B_i \in \mathcal{B}.$$

Числа c_i и множества B_i определены при этом, конечно, не однозначно, но всегда можно перейти к представлению с дизъюнктными множествами, которое очевидным образом однозначно.

1.2.4. Определение. *Функция распределения с.в. ξ задается равенством*
$$F_\xi(t) = P\big(\omega\colon\ \xi(\omega) < t\big), \quad t \in \mathbb{R}.$$

Функция распределения возрастает, непрерывна слева, имеет предел 0 на $-\infty$ и 1 на $+\infty$. Непрерывность функции распределения равносильна отсутствию значений с положительными вероятностями прообразов.

Например, если $\xi = 0$, то $F_\xi(t) = 0$ при $t \leqslant 0$ и $F_\xi(t) = 1$ при $t > 0$. Если ξ принимает конечное число разных значений $c_1, \ldots, c_n$, то F_ξ постоянна на $(-\infty, c_1]$, $(c_1, c_2]$ и т. д. до $(c_{n-1}, c_n]$, а также на луче $(c_n, +\infty)$.

В случае непрерывной функции распределения для каждого значения $\alpha \in (0,1)$ непусто множество $F_\xi^{-1}(\alpha)$. Точки этого множества, т. е. решения уравнения
$$F_\xi(t_\alpha) = \alpha,$$
называют *квантилями*. Квантиль с $\alpha = 1/2$ называют *медианой* распределения, причем чаще медиану определяют как наименьшее число в $F_\xi^{-1}(1/2)$. Для разрывных функций распределения обычно медиану задают как $\inf\{s\colon\ F_\xi(s) \geqslant 1/2\}$.

Отметим, что медиану бывает полезно вводить и для конечных наборов чисел $c_1, \ldots, c_n$, необязательно связанных с вероятностными распределениями. Для этого такой набор упорядочивают по возрастанию и получают числа $b_1 \leqslant \cdots \leqslant b_n$ (это называется вариационным рядом), затем в случае нечетного n берут срединное из полученных чисел, а в случае четного берут попавшее на место $n/2$ (или следующее). Во многих случаях медианное значение является гораздо более показательным, чем среднее. Например, при вычислении «средней» заработной платы в организации очень высокие зарплаты отдельных сотрудников могут заметно поднять среднюю арифметическую по организации, но на медианное значение они почти не влияют (именно поэтому в такого рода вычислениях обычно используются совершенно бессмысленные средние арифметические показатели). Условный пример: в некотором университете около 200 сотрудников получали в месяц примерно по 25000 рублей; после удвоения фонда зарплаты пятерым стали платить по 1000000 рублей в месяц, что примерно удвоило среднюю зарплату, но никак не повлияло на медиану.

Квантили распределений активно используются в статистике.

Помимо функции распределения полезно ввести и само *распределение случайной величины* ξ как борелевскую вероятностную меру, задаваемую формулой

$$P_\xi(B) := P\big(\xi^{-1}(B)\big), \quad B \in \mathcal{B}(\mathbb{R}).$$

Условие измеримости позволяет определить правую часть, при этом из равенства $\xi^{-1}(\cup_n B_n) = \cup_n \xi^{-1}(B_n)$ легко получить счетную аддитивность P_ξ.

Аналогично задаются в $\mathbb{R}^d$ распределения случайных векторов.

Отметим, что для всякой возрастающей непрерывной слева функции F, имеющей пределы 0 и 1 в $-\infty$ и $+\infty$, есть случайная величина с такой функцией распределения. Проще всего взять функцию $\xi(t) = t$ на прямой с борелевской σ-алгеброй и мерой P, которая получена продолжением с алгебры конечных объединений промежутков вида $[a, b)$, где a и b могут быть $-\infty$ и $+\infty$, причем $P([a, b)) = F(b) - F(a)$. Для применения теоремы существования нужна счетная аддитивность P на этой алгебре, вытекающая из непрерывности F слева.

Помимо сходимости случайных величин почти наверное, т. е. соотношения $\xi_n(\omega) \to \xi(\omega)$ для почти всех ω, часто и используется более слабая сходимость по вероятности.

1.2.5. Определение. *Случайные величины ξ_n сходятся по вероятности к случайной величине ξ (все они заданы на одном пространстве), если для каждого $\varepsilon > 0$ мы имеем*

$$\lim_{n\to\infty} P(\omega\colon |\xi_n(\omega) - \xi(\omega)| \geqslant \varepsilon) = 0.$$

Отметим, что при сходимости по вероятности возможна ситуация, когда $P(\omega\colon |\xi_n(\omega) - \xi(\omega)| > 0) = 1$, скажем, можно взять $\xi_n = 1/n$ и $\xi = 0$. Ниже приведен простой пример сходящейся по вероятности последовательности, которая не сходится ни в одной точке. Кроме того, мы проверим, что сходимость почти наверное влечет сходимость по вероятности.

§ 1.3. Математическое ожидание и дисперсия

Теперь введем математическое ожидание (интеграл Лебега) случайной величины ξ на $(\Omega, \mathcal{B}, P)$. Для простой случайной величины ξ вида $c_1 I_{A_1} + \cdots + c_n I_{A_n}$ положим

$$\mathrm{M}\xi := \int_\Omega \xi(\omega)\, P(d\omega) := \int_\Omega \xi\, dP = \sum_{i=1}^n c_i P(A_i).$$

Из аддитивности меры следует корректность определения (независимость от представления).

Если $\xi \geqslant 0$, то положим

$$\mathrm{M}\xi := \int_\Omega \xi(\omega)\, P(d\omega) := \int_\Omega \xi\, dP = \sup_{\eta\in S_\xi} \mathrm{M}\eta,$$

где S_ξ — множество всех простых случайных величин η, для которых $\eta \leqslant \xi$ п.н., причем интегрируемыми (имеющими математическое ожидание будем считать только те с.в., для которых этот sup конечен).

Наконец, в общем случае запишем ξ в виде

$$\xi = \xi^+ - \xi^-, \quad \text{где}\ \ \xi^+ = \max(\xi,0),\ \xi^- = -\min(\xi,0).$$

Будем считать ξ интегрируемой, если обе с.в. ξ^+ и ξ^- интегрируемы, при этом положим

$$\mathrm{M}\xi := \mathrm{M}\xi^+ - \mathrm{M}\xi^-.$$

Таким образом, определение таково, что вместе с ξ интегрируемой обязательно оказывается и $|\xi|$ (в отличие от несобственного интеграла Римана).

Множество $\mathfrak{L}^1(P)$ интегрируемых с.в. обладает тем свойством, что содержит линейные комбинации своих элементов, причем интеграл Лебега линеен на нем:

$$\mathrm{M}(\alpha\xi + \beta\eta) = \alpha\mathrm{M}\xi + \beta\mathrm{M}\eta \quad \forall\, \alpha,\beta \in \mathbb{R},$$

однако это множество не является линейным пространством (для тех, кто помнит определение линейного пространства). Проблема с нулем: в линейном пространстве $x - x = 0$ причем нулевой элемент только один, а здесь мы разрешили случайным величинам быть не всюду определенными, поэтому функция $\xi(\omega) - \xi(\omega)$ тоже может быть не всюду определена, так что она не есть штатный нуль (тождественно нулевая функция). Обычный выход в таких ситуациях: переход к пространству $L^1(P)$ классов эквивалентности, где эквивалентными объявляются функции, равные почти всюду. Тогда $L^1(P)$ оказывается не только линейным пространством, но еще и нормированным, если в качестве нормы взять $\mathrm{M}|\xi|$. Более того, оно полно с этой нормой (т. е. является банаховым пространством).

Из определения очевидно, что переопределение случайной величины на множестве нулевой вероятности не влияет на ее интегрируемость и на значение интеграла. Кроме того, если $\xi \leqslant \eta$ п.н. и обе с.в. интегрируемы, то $\mathrm{M}\xi \leqslant \mathrm{M}\eta$. Наконец, если с.в. ξ и η таковы, что $|\xi| \leqslant |\eta|$ п.н., то интегрируемость η влечет интегрируемость ξ.

1.3.1. Определение. *Пусть* ξ *— такая с.в., что* $\xi^2 \in \mathcal{L}^1(P)$. *Тогда величина*
$$\mathrm{D}\xi := \mathrm{M}(\xi - \mathrm{M}\xi)^2 = \mathrm{M}\xi^2 - (\mathrm{M}\xi)^2$$
называется дисперсией ξ.

Отметим, что интегрируемость ξ^2 дает существование математического ожидания (это видно из приводимого ниже неравенства Коши–Буняковского).

В случае нулевого математического ожидания дисперсия совпадает со вторым моментом (k-й момент есть $\mathrm{M}|\xi|^k$).

Для интегрируемых случайных величин есть два очень важных неравенства.

1.3.2. Теорема. (Неравенство Чебышёва) *Пусть* $\xi \in \mathcal{L}^1(P)$. *Тогда для всякого* $R > 0$ *верно неравенство*
$$P\bigl(\omega\colon\ |\xi(\omega)| \geqslant R\bigr) \leqslant \frac{\mathrm{M}|\xi|}{R}.$$

Доказательство. Пусть Ω_R — множество в левой части этого неравенства. Тогда $RI_{\Omega_R} \leqslant |\xi|$. □

1.3.3. Следствие. *Для всякой случайной величины* ξ *с конечной дисперсией имеем*
$$P(|\xi - \mathrm{M}\xi| \geqslant R) \leqslant \frac{\mathrm{D}\xi}{R^2}.$$

Из этой оценки следует, что $\mathrm{D}\xi = 0$ в точности тогда, когда ξ п.н. совпадает с постоянной.

1.3.4. Пример. («Правило трех сигма») Для всякой случайной величины ξ с дисперсией $\sigma^2 > 0$ имеем
$$P(|\xi - \mathrm{M}\xi| \geqslant 3\sigma) \leqslant \frac{\mathrm{D}\xi}{9\sigma^2} = \frac{\sigma^2}{9\sigma^2} = \frac{1}{9}.$$
Таким образом, с вероятностью $8/9 \approx 0.9$ случайная величина с конечной дисперсией не отклоняется от своего среднего больше, чем на утроенный корень из дисперсии. Здесь интересно то, что указанная вероятность не зависит от дисперсии (последняя измеряет лишь размах отклонения). Для нулевой дисперсии это тоже верно, ибо в этом случае случайная величина п.н. равна ее среднему.

Через $\mathcal{L}^p(P)$ обозначим множество случайных величин ξ, для которых $|\xi|^p \in \mathcal{L}^1(P)$.

1.3.5. Теорема. (Неравенство Гёльдера) *Пусть* $\xi \in \mathcal{L}^p(P)$, *где* $p \in (1, +\infty)$. *Тогда для всякого* $q \in [1, p]$ *имеем* $\xi \in \mathcal{L}^q(P)$, *причем верно неравенство*

$$(\mathrm{M}|\xi|^q)^{1/q} \leqslant (\mathrm{M}|\xi|^p)^{1/p}.$$

При $p = 2$, $q = 1$ *получаем неравенство Коши–Буняковского*

$$\mathrm{M}|\xi| \leqslant (\mathrm{M}|\xi|^2)^{1/2}.$$

Как водится, интеграл приходится переставлять с пределом и суммированием. На этот счет есть три основные теоремы, в каждой из которых одинаковое начало: даны интегрируемые с.в. ξ_n и

$$\xi(\omega) = \lim_{n\to\infty} \xi_n(\omega) \quad \text{п.н.}$$

1.3.6. Теорема. (Теорема Лебега о мажорируемой сходимости) *Пусть существует* $\Phi \in \mathcal{L}^1(P)$ *с*

$$|\xi_n(\omega)| \leqslant \Phi(\omega) \quad \textit{п.н.}$$

при всех n. *Тогда* $\xi \in \mathcal{L}^1(P)$ *и*

$$\mathrm{M}\xi = \lim_{n\to\infty} \mathrm{M}\xi_n. \tag{1.3.1}$$

Кроме того, $\lim\limits_{n\to\infty} \mathrm{M}|\xi - \xi_n| = 0$.

1.3.7. Теорема. (Теорема Беппо Леви о монотонной сходимости) *Пусть*

$$\xi_n(\omega) \leqslant \xi_{n+1}(\omega) \quad \textit{п.н.}$$

при всех n, *причем допускается и бесконечный предел. Тогда справедливо заключение предыдущей теоремы, в частности величина* $\xi(\omega)$ *почти всюду конечна.*

1.3.8. Теорема. (Теорема Фату) *Пусть*

$$\xi_n \geqslant 0 \quad \textit{и} \quad \sup_n \mathrm{M}\xi_n < \infty.$$

Тогда $\xi \in \mathcal{L}^1(P)$ *и*

$$\mathrm{M}\xi \leqslant \liminf_{n\to\infty} \mathrm{M}\xi_n.$$

Отметим, что в теоремах Лебега и Фату можно заменить сходимость почти наверное сходимостью по вероятности (в теореме Беппо Леви это делать бессмысленно, так как возрастающая последовательность и так стремится к конечному или бесконечному пределу).

Если Ω счетно (или конечно) и состоит из точек ω_n, то обычно можно считать, что $\mathcal{B}$ есть класс всех множеств в Ω, тогда все функции измеримы, а мера P задается такой же формулой, как и в конечном случае:

$$P(A) = \sum_{n\colon \omega_n \in A} P(\omega_n).$$

При этом интегрируемость случайной величины ξ равносильна сходимости ряда из $|\xi(\omega_n)|P(\omega_n)$, а математическое ожидание задается формулой

$$\mathrm{M}\xi = \sum_{n=1}^{\infty} \xi(\omega_n) P(\omega_n).$$

Все это можно проверить по определению, но проще воспользоваться теоремами о сходимости: считая, что $\xi \geqslant 0$ (к этому по определению сводится общий случай), вводим случайные величины ξ_n, совпадающие с ξ на $\{\omega_1, \ldots, \omega_n\}$ и равные нулю на дополнении, получаем монотонную последовательность с математическими ожиданиями $\sum_{i=1}^{n} \xi(\omega_i) P(\omega_i)$. Если ξ интегрируема, то эти числа равномерно ограничены, причем верное и обратное.

Из теоремы Лебега или теоремы Беппо Леви следует, что всякая неотрицательная функция $\varrho \in \mathcal{L}^1(P)$ задает конечную меру по формуле

$$Q(B) := \varrho \cdot P(B) := \int_B \varrho(\omega)\, P(d\omega), \quad B \in \mathcal{B}.$$

Такая мера называется мерой с плотностью ϱ относительно P. Теорема Радона – Никодима описывает все такие меры Q: мера $Q \geqslant 0$ на $\mathcal{B}$ задается неотрицательной плотностью относительно P в точности тогда, когда $Q(B) = 0$ для всякого $B \in \mathcal{B}$ с $P(B) = 0$.

Приведем простой пример использования этих теорем.

1.3.9. Лемма. (Лемма Бореля – Кантелли для произвольных событий) *Пусть $\{B_n\}$ — последовательность событий и*

$$\sum_{n=1}^{\infty} P(B_n) < \infty.$$

Тогда почти наверное происходит лишь конечное число событий B_n (почти всякая точка ω входит лишь в конечное число из них), т. е. множество $B = \bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} B_n$ имеет вероятность 0.

Доказательство. Сходимость ряда из $\mathrm{M} I_{B_n} = P(B_n)$ дает сходимость ряда из $I_{B_n}(\omega)$ п.н., но ряд из нулей и единиц сходится лишь тогда, когда в нем конечное число единиц (что соответствует принадлежности ω лишь к конечному числу множеств B_n). $\square$

1.3.10. Определение. *Случайная величина ξ имеет плотность распределения p_ξ, если p_ξ — такая интегрируема по Лебегу функция на прямой, что*

$$P_\xi(B) = \int_{\mathbb{R}} p_\xi(t)\,dt \quad \forall\, B \in \mathcal{B}(\mathbb{R}).$$

Аналогично задается плотность распределения случайного вектора в пространстве $\mathbb{R}^d$.

Не всякая случайная величина имеет плотность распределения. Например, если $\xi = 0$, то P_ξ — мера Дирака в нуле, которая не задается плотностью. Напомним, что по теореме Радона – Никодима борелевская мера ν задается плотностью относительно меры Лебега в точности тогда, когда она равна на нулю на всех борелевских множествах лебеговской меры нуль.

Как по функции распределения узнать, задается ли распределение плотностью? Простое достаточное (но не необходимое) условие состоит в непрерывной дифференцируемости F_ξ. В самом деле, пусть $\varrho(t) = F'_\xi(t)$ — непрерывная функция. Тогда

$$P_\xi\bigl([a,b)\bigr) = F_\xi(b) - F_\xi(a) = \int_a^b \varrho(t)\,dt = \int_{[a,b)} \varrho(t)\,dt$$

для всех $[a,b)$. Как известно, это тождество равносильно тому, что мера P_ξ задается плотностью ϱ (равенство мер на полуинтервалах влечет равенство на всех борелевских множествах).

Необходимое и достаточное условие шире и состоит в том, что функция F_ξ должна быть абсолютно непрерывна на каждом отрезке I, т. е. для всякого $\varepsilon > 0$ должно найтись такое $\delta > 0$, что для всякого набора дизъюнктных интервалов (a_i, b_i) суммарной длины не более δ верно неравенство $\sum_{i=1}^n |F_\xi(b_i) - F_\xi(a)| < \varepsilon$. Конечно, это условие вряд ли можно считать легко проверяемым.

В случае его выполнения верно равенство $p_\xi(t) = F'_\xi(t)$ почти всюду. Отметим еще, что конечная производная $F'_\xi(t)$ существует почти всюду для всякой функции распределения, но не всегда по ней можно восстановить $F_\xi(t)$, скажем, так обстоит дело для $\xi = 0$ (ниже будет пример, когда $F'_\xi(t) = 0$ почти всюду даже для непрерывной функции F_ξ).

Далее систематически используется следующий общий факт.

1.3.11. Теорема. (Формула замены переменных) *Пусть ξ — случайная величина на пространстве $(\Omega, \mathcal{B}, P)$. Тогда для каждой ограниченной борелевской функции f верно равенство*

$$\mathrm{M}f(\xi) = \int_\Omega f\bigl(\xi(\omega)\bigr)\, P(d\omega) = \int_{\mathbb{R}} f(t)\, P_\xi(dt). \qquad (1.3.2)$$

Это же равенство верно и для неограниченной борелевской функции f при условии, что хотя бы один из двух интегралов существует. Аналогичные утверждения верны и для распределений случайных векторов.

Доказательство. Если f — индикатор множества $B \in \mathcal{B}(\mathbb{R})$, то это верно по определению P_ξ, так как $I_B \circ \xi = I_{\xi^{-1}(B)}$. Значит, (1.3.2) верно для простых функций, а тогда остается в силе для всех ограниченных борелевских, ибо это равенство сохраняется при равномерных пределах по f, а всякая ограниченная борелевская функция есть равномерный предел простых функций. Действительно, если $|f| \leqslant M$, то для заданного $\varepsilon > 0$ можно разделить $[-M, M]$ на конечное число промежутков $J_1, \ldots, J_n$ длины не более ε и взять функцию $g = \sum_{k=1}^n y_k I_{f^{-1}(J_k)}$, где y_k — левый конец J_k, для которой $|f(x) - g(x)| \leqslant \varepsilon$ при всех x, ибо это верно на каждом из множеств $f^{-1}(J_k)$, на которые разбивается все пространство.

Если $f \geqslant 0$ не является ограниченной, но один из двух интегралов существует, то для функций $f_N = \min(f, N)$ (они борелевские ограниченные) равенство верно, причем обе части остаются равномерно ограниченными по N (тем интегралом, который существует). По теореме Беппо Леви при $N \to \infty$ получаем равенство и для f, ибо $f_N(\xi)$ возрастают к $f(\xi)$. Общий случай получается разложением $f = f^+ - f^-$. Все это верно и для случайных векторов. □

При наличии плотности p_ξ интеграл от $f(t)$ по мере P_ξ вычисляется как интеграл от $f(t)p_\xi(t)$ по мере Лебега (что обосновывается по стандартной схеме: имеем это для индикаторов, значит, для простых f, а тогда для ограниченных борелевских, после чего можно перейти и к интегрируемым). Поэтому приходим к такому представлению.

1.3.12. Следствие. *Если P_ξ имеет плотность p_ξ, то*

$$\mathrm{M}f(\xi) = \int_{-\infty}^{+\infty} f(t)p_\xi(t)\, dt.$$

Почему в формуле замены переменных нет якобианов, как в анализе? Потому что интегрирование ведется по мере P_ξ, а не по мере

Лебега. Вот переход в последнем следствии к мере Лебега неявно содержит «якобиан» в плотности p_ξ. Если исходная мера P была задана на $\mathbb{R}^d$ плотностью ϱ относительно меры Лебега, а $\xi\colon \mathbb{R}^d \to \mathbb{R}^d$ было дифференцируемым с невырожденным якобианом, то мера P_ξ тоже будет задаваться плотностью, причем в p_ξ будет входить якобиан ξ, так что будет согласие с формулами замены переменных из анализа.

1.3.13. Пример. Если ξ имеет математическое ожидание, то

$$\mathrm{M}\xi = \int_{\mathbb{R}} t\, P_\xi(dt),$$

что при наличии плотности равно

$$\mathrm{M}\xi = \int_{\mathbb{R}} t\, P_\xi(dt) = \int_{-\infty}^{+\infty} t p_\xi(t)\, dt.$$

Аналогично для дисперсии:

$$\mathrm{D}\xi = \int_{\mathbb{R}} (t - \mathrm{M}\xi)^2\, P_\xi(dt),$$

$$\mathrm{D}\xi = \int_{\mathbb{R}} (t - \mathrm{M}\xi)^2\, P_\xi(dt) = \int_{-\infty}^{+\infty} (t - \mathrm{M}\xi)^2 p_\xi(t)\, dt.$$

1.3.14. Пример. Для случайного вектора $\xi = (\xi_1, \xi_2)$ с распределением P_ξ:

$$\mathrm{M}(\xi_1\xi_2) = \int_{\mathbb{R}^2} x_1 x_2\, P_\xi(dx_1\, dx_2).$$

Если P_ξ имеет плотность распределения p_ξ, то получим

$$\int_{\mathbb{R}^2} x_1 x_2\, p_\xi(x_1, x_2)\, dx_1\, dx_2.$$

Наконец, для случайного вектора ξ с распределением P_ξ в $\mathbb{R}^d$ имеем

$$\mathrm{M} f(\xi) = \int_{\mathbb{R}^d} f(x_1, \ldots, x_d)\, P_\xi(dx)$$

для всякой ограниченной борелевской функции f (или такой борелевской функции, что $f(\xi)$ имеем математическое ожидание). Если же ξ имеет плотность распределения ϱ_ξ, то

$$\mathrm{M} f(\xi) = \int_{\mathbb{R}^d} f(x_1, \ldots, x_d) \varrho_\xi(x_1, \ldots, x_d)\, dx_1 \cdots dx_d.$$

§ 1.4. Сходимость случайных величин

Интуитивно наиболее понятной является сходимость случайных величин почти наверное, когда $\xi_n(\omega) \to \xi(\omega)$ для всех точек ω, кроме точек множества нулевой вероятности.

Из этой сходимости вытекает сходимость по вероятности, ибо для всякого $\varepsilon > 0$ множества

$$A_n = \{\omega\colon \sup_{k \geqslant n} |\xi_k(\omega) - \xi(\omega)| \geqslant \varepsilon\}$$

убывают, а их пересечение имеет меру нуль по условию. Значит, мы имеем $P(A_n) \to 0$, что влечет сходимость по вероятности.

Однако обратное неверно. Например, можно взять отрезок $[0, 1]$ с обычной мерой Лебега и на нем последовательность функций f_n, полученную последовательной нумерацией следующих конечных наборов: $f_1 = 1$, $f_{2,1} = 1$ на $[0, 1/2]$ и $f_{2,1} = 0$ на $(1/2, 1]$, затем отрезок делится на четыре равных части, а функции $f_{3,1}, \ldots, f_{3,4}$ — индикаторы этих четырех отрезков, аналогично для всякого n. Тогда $\{f_n(x)\}$ содержит бесконечно много нулей и единиц для всякого $x \in (0, 1)$, т. е. нет сходимости ни в одной точке. При этом очевидна сходимость к нулю по мере, ибо мера множества, где $f_n \neq 0$, стремится к нулю.

Несмотря на этот пример, имеется положительное общее утверждение (теорема Рисса): если случайные величины ξ_n сходятся к ξ по мере, то найдется подпоследовательность, сходящаяся почти всюду.

Сходимость в среднем определяется как соотношение

$$\mathrm{M}|\xi_n - \xi| \to 0.$$

Из него следует сходимость по мере в силу неравенства Чебышёва:

$$P(|\xi - \xi_n| \geqslant \varepsilon) \leqslant \varepsilon^{-1}\mathrm{M}|\xi - \xi_n|,\ \varepsilon > 0.$$

С другой стороны, сходимость в среднем получается из сходимости по мере добавлением условия наличия общей интегрируемой мажоранты (теорема Лебега).

Сходимость в среднем порядка p определяется как соотношение $\mathrm{M}|\xi_n - \xi|^p \to 0$. При $p = 2$ это называют сходимостью в среднем квадратическом.

Неравенство Коши – Буняковского говорит, что из сходимости в среднем квадратическом следует сходимость в среднем.

§ 1.5. Некоторые классы случайных величин

Простейшая непостоянная случайная величина — принимающая два значения 1 и 0 с вероятностями p и $1-p$. Она называется *бернуллиевской*. С ее помощью строятся другие важные случайные величины (как дискретные, так и с непрерывными распределениями).

Пуассоновская случайная величина ξ с параметром $\lambda > 0$ принимает целые неотрицательные значения $k = 0, 1, 2, \ldots$ с вероятностями $\lambda^k e^{-\lambda}/k!$ (их сумма равна единице).

Равномерное распределение в отрезке $[a, b]$ — мера Лебега, деленная на $b-a$; случайная величина с таким распределением называется равномерно распределенной в $[a, b]$; такие величины — основа построения прочих случайных величин с плотностями распределения.

Экспоненциальное распределение на $[0, +\infty)$ с параметром $\lambda > 0$ задается плотностью $\lambda e^{-\lambda x} I_{[0,+\infty)}$.

Гауссовское (или *нормальное*) распределение с параметрами $a \in \mathbb{R}$ и $\sigma \geqslant 0$ задается так. Если $\sigma = 0$, то это мера Дирака в точке a. Если $\sigma > 0$, то это мера с плотностью

$$\frac{1}{\sigma\sqrt{2\pi}} \exp\Big(-\frac{(x-a)^2}{2\sigma^2}\Big).$$

Конечно, при этом необходимо проверить, что интеграл от указанной плотности равен 1. Сделать это можно так: квадрат интеграла от e^{-x^2} запишем как произведение интегралов от e^{-x^2} и e^{-y^2}, что даст интеграл по плоскости от $e^{-x^2-y^2}$. Вычисляя его в полярных координатах (не забыть бы, какой там якобиан!), получаем интеграл от $2\pi r e^{-r^2}$ по лучу $(0, +\infty)$, т. е. π.

Мера с параметрами $a = 0$ и $\sigma = 1$ называется *стандартной гауссовской*. При $a = 0$ гауссовская мера называется симметричной или центрированной. Случайная величина с гауссовским распределением называется гауссовской. Такая величина ξ имеет вид $a + \sigma \cdot \eta$, где η — стандартная гауссовская случайная величина.

Распределение Коши задается плотностью

$$\frac{1}{\pi}\frac{1}{1+x^2}.$$

Логнормальное распределение есть распределение случайной величины e^{ξ}, где ξ имеет нормальное распределение.

Распределение Парето задается плотностью p следующего вида: $p(x) = 0$ при x, меньших некоторого x_0,

$$p(x) = \frac{\alpha}{x_0}\Big(\frac{x_0}{x}\Big)^{\alpha}, \quad x > x_0,$$

где $\alpha > 1$ — некоторый параметр.

Ясно, что распределение Коши и распределение Парето при $\alpha \leqslant 2$ не имеют среднего. В качестве упражнения предлагается вычисление математических ожиданий и дисперсий указанных распределений в случаях их существования.

Всякое распределение на прямой может быть получено возрастающим преобразованием равномерного распределения в $(0, 1)$.

1.5.1. Теорема. *Для всякой вероятностной борелевской меры μ на прямой найдется такая возрастающая функция $\xi\colon\ (0,1) \to \mathbb{R}$, что μ совпадает с распределением ξ относительно меры Лебега на $(0, 1)$.*

Доказательство — задача 1.7.14.

§ 1.6. Независимые случайные величины

При рассмотрении случайных величин и событий важнейшую роль играет понятие независимости. Именно это понятие и различные его модификации отличают теорию вероятностей от раздела теории меры. Вводимое ниже техническое понятие независимости не следует смешивать с независимостью в философском или житейско-бытовом смысле, хотя некоторые параллели и аналогии здесь иногда уместны.

1.6.1. Определение. *Случайные величины $\xi_1, \ldots, \xi_n$ на общем вероятностном пространстве $(\Omega, \mathcal{B}, P)$ называют независимыми, если*

$$P\big(\omega\colon\ \xi_1(\omega) < c_1, \ldots, \xi_n(\omega) < c_n\big) = \\ = P\big(\omega\colon\ \xi_1(\omega) < c_1\big) \cdots P\big(\omega\colon\ \xi_n(\omega) < c_n\big) \quad \forall\, c_1, \ldots, c_n \in \mathbb{R}.$$

Случайные величины из набора $\{\xi_t\}_{t\in T}$ называют независимыми, если всякий конечный поднабор в нем состоит из независимых величин.

Приведенное определение независимости равносильно следующему: распределение случайного вектора $(\xi_1, \ldots, \xi_n)$ равно произведению $P_{\xi_1} \otimes \cdots \otimes P_{\xi_n}$ распределений этих случайных величин.

Напомним, что произведение двух мер μ и ν на пространствах $(X, \mathcal{A})$ и $(Y, \mathcal{B})$ соответственно есть мера на пространстве $X \times Y$ с σ-алгеброй $\mathcal{A} \otimes \mathcal{B}$, порожденной множествами вида $A \times B$, где $A \in \mathcal{A}$, $B \in \mathcal{B}$, задаваемая формулой $\mu\otimes\nu(A\times B) = \mu(A)\nu(B)$ на множествах последнего вида. Разумеется, при этом обосновывается существование такой меры на $\mathcal{A} \otimes \mathcal{B}$ (это делается путем проверки ее счетной аддитивности на алгебре конечных объединений произведений $A \times B$). Проверяется также, что для всех $E \in \mathcal{A} \otimes \mathcal{B}$ имеет место формула

$$\mu \otimes \nu(E) = \int_Y \mu(E_y)\,\nu(dy), \quad E_y = \{x \in X \colon (x,y) \in E\}.$$

Верна также симметричная формула

$$\mu \otimes \nu(E) = \int_X \nu(E^x)\,\mu(dx), \quad E^x = \{y \in Y \colon (x,y) \in E\}.$$

Эти формулы — частные случаи теоремы Фубини, согласно которой интеграл от функции f на $X \times Y$, интегрируемой по мере $\mu \otimes \nu$ вычисляется посредством

$$\int_{X\times Y} f\,d(\mu \otimes \nu) = \int_Y \int_X f(x,y)\,\mu(dx)\,\nu(dy) = \\ = \int_X \int_Y f(x,y)\,\nu(dy)\,\mu(dx).$$

Частью теоремы Фубини является утверждение о том, что внутренние интегралы по одному переменному автоматически существует при почти всех фиксированных значениях второго переменного.

По индукции определяется произведение $\mu_1 \otimes \cdots \otimes \mu_n$ мер μ_i, заданных на σ-алгебрах $\mathcal{B}_i$. Оно определено на σ-алгебре $\mathcal{B}_1 \otimes \cdots \otimes \mathcal{B}_n$, порожденной произведениями вида $B_1 \times \cdots \times B_n$, где $B_i \in \mathcal{B}_i$.

Несколько сложнее обстоит дело со счетным произведением вероятностных пространств $(\Omega_n, \mathcal{B}_n, \mu_n)$. Аналогично конечному произведению можно взять пространство $\Omega = \prod_{n=1}^\infty \Omega_n$ и рассмотреть σ-алгебру $\otimes_{n=1}^\infty \mathcal{B}_n$, порожденную множествами $\prod_{n=1}^\infty B_n$, где $B_n \in \mathcal{B}_n$. Для этого на алгебре всех множеств вида $B \times \Omega_{n+1} \times \Omega_{n+2}$, где B входит в $\mathcal{B}_1 \otimes \cdots \mathcal{B}_n$, задают меру $\otimes_{n=1}^\infty \mu_n$ формулой

$$\otimes_{n=1}^\infty \mu_n(B \times \Omega_{n+1} \times \Omega_{n+2}) = \mu_1 \otimes \cdots \otimes \mu_n(B).$$

Однако не является тривиальной проверка ее счетной аддитивности. После проведения этой проверки получается вероятностная мера на $\otimes_{n=1}^\infty \mathcal{B}_n$. После этого случай несчетного произведения вероятностных пространств $(\Omega_\alpha, \mathcal{B}_\alpha, \mu_\alpha)$ получается очень легко: он сводится к счетному случаю, так как обычное объединение всевозможных σ-алгебр, получаемых из счетных произведений сомножителей, оказывается σ-алгеброй. При этом множество из счетного произведения рассматривается как множество из произведения всех Ω_α дописыванием всего пространства сомножителем при отсутствующем индексе.

Равносильность исходного определения независимости утверждению о совпадении совместного распределения с произведением распределений компонент следует из того факта, что равенство борелевских мер на $\mathbb{R}^n$ вытекает из равенства их значений на всех множествах вида $\{x\colon x_1 < c_1, \ldots, x_n < c_n\}$.

Независимость случайных величин $\xi_1, \ldots, \xi_n$ равносильна также тождеству

$$P\bigl(\omega\colon\ \xi_1(\omega) \in B_1, \ldots, \xi_n(\omega) \in B_n\bigr) = \\ = P\bigl(\omega\colon\ \xi_1(\omega) \in B_1\bigr) \cdots P\bigl(\omega\colon\ \xi_n(\omega) \in B_n\bigr) \quad \forall\, B_1, \ldots, B_n \in \mathcal{B}(\mathbb{R}).$$

Это вытекает из сказанного выше.

Данное техническое определение приводит к тому, что для всякой случайной величины ξ и всякой постоянной η мы получаем независимость ξ и η, что кажется не отвечающим интуитивному восприятию независимости. К этому следует относиться как к некоторому особому случаю. Более того, если ξ — случайная величина, а f — такая борелевская функция, что случайная величина $\eta = f(\xi)$ имеет распределение, отличное от дираковского, т. е. η не есть постоянная почти наверное, то ξ и η не могут быть независимыми (иначе говоря, наличие нетривиальной зависимости исключает независимость). Чтобы в этом убедиться, возьмем такое борелевское множество $B_2 \subset \mathbb{R}$, что $0 < P\bigl(\eta^{-1}(B_2)\bigr) < 1$; его существование как раз и означает, что $f(\xi)$ имеет отличное от дираковского распределение. Тогда число $P\bigl(\eta^{-1}(B_2)\bigr)$ отлично от своего квадрата, поэтому

$$P\bigl(\omega\colon\ \xi(\omega) \in f^{-1}(B_2), \eta(\omega) \in B_2\bigr) \neq \\ \neq P\bigl(\omega\colon\ \xi(\omega) \in f^{-1}(B_2)\bigr) P\bigl(\omega\colon\ \eta(\omega) \in B_2\bigr),$$

ибо слева стоит вероятность множества $\eta^{-1}(B_2)$, а справа — ее квадрат.

Подчеркнем, что распределения случайных величин ξ и η по отдельности не позволяют судить о независимости этих величин (кроме случая, когда одно из распределений дираковское).

Для всякого конечного набора случайных величин $\eta_1, \ldots, \eta_n$ легко построить набор независимых случайных величин с теми же распределениями: достаточно взять на $\mathbb{R}^n$ произведение $P_{\eta_1} \otimes \cdots \otimes P_{\eta_n}$ данных распределений, а в качестве искомых случайных величин взять просто координатные функции. Разумеется, при этом совместные распределения могут быть различны. Используя бесконечные произведения вероятностных пространств, точно так же получаем независимые случайные величины с заданными распределениями в произвольном количестве. В частности, всегда имеет смысл выражение «последовательность независимых случайных величин с заданными распределениями».

Одно из важнейших проявлений независимости случайных величин состоит в следующем.

1.6.2. Теорема. *Если независимые случайные величины ξ и η имеют математические ожидания, то таково и их произведение, причем*

$$\mathrm{M}(\xi\eta) = \mathrm{M}\xi\mathrm{M}\eta.$$

Доказательство. Отметим, что в общем случае произведение интегрируемых функций не всегда интегрируемо (например, квадрат интегрируемой функции может быть неинтегрируемым). Однако по теореме Фубини произведение интегрируемых функций f и g на пространствах с мерами (X,μ) и (Y,ν) интегрируемо относительно меры $\mu\otimes\nu$ на $X\times Y$. Поэтому функция xy на $\mathbb{R}^2$ интегрируема относительно произведения мер P_ξ и P_η. Условие независимости говорит, что эта функция интегрируема относительно распределения вектора (ξ,η). По формуле замены переменных получаем интегрируемость $\xi\eta$ относительно меры P. При этом $\mathrm{M}(\xi\eta)$ есть интеграл от xy по распределению (ξ,η), равному мере $P_\xi\otimes P_\eta$. По теореме Фубини это дает произведение интегралов от координатных функций по мерам P_ξ и P_η. Еще раз применяя формулу замены переменных, получаем произведение $\mathrm{M}\xi$ и $\mathrm{M}\eta$. □

По индукции из доказанного получаем, что для независимых случайных величин $\xi_1,\ldots,\xi_n$, имеющих математические ожидания, произведение $\xi_1\cdots\xi_n$ также имеет математическое ожидание, причем

$$\mathrm{M}(\xi_1\cdots\xi_n) = \mathrm{M}\xi_1\cdots\mathrm{M}\xi_n.$$

При этом используется независимость ξ_1 и $\xi_2\cdots\xi_n$.

Разумеется, последнее равенство не равносильно независимости.

1.6.3. Следствие. *Если независимые случайные величины $\xi_1,\ldots,\xi_n$ имеют конечные дисперсии, то такова и их сумма, причем*

$$\mathrm{D}(\xi_1+\cdots+\xi_n) = \mathrm{D}\xi_1+\cdots+\mathrm{D}\xi_n,$$

т. е. дисперсии суммируются подобно математическим ожиданиям. Кроме того, эти величины попарно некоррелированы, т. е.

$$\mathrm{M}(\xi_i-m_i)(\xi_j-m_j) = 0 \quad \text{при} \quad i\neq j.$$

Доказательство. Ясно, что достаточно рассмотреть случай двух независимых случайных величин ξ и η. Тогда

$$\mathrm{M}(\xi+\eta)^2 = \mathrm{M}(\xi^2) + 2\mathrm{M}(\xi\eta) + \mathrm{M}(\eta)^2 = \mathrm{M}(\xi^2) + 2\mathrm{M}\xi\mathrm{M}\eta + \mathrm{M}(\eta)^2,$$

что после вычитания $\big(\mathrm{M}(\xi+\eta)\big)^2$ дает нужное равенство. $\square$

Конечно, такого эффекта нет при наличии зависимости. Например, имеем $\mathrm{D}(\xi+\xi) = 4\mathrm{D}\xi$.

Частным случаем независимости случайных величин является независимость событий $A_1, \ldots, A_n \in \mathcal{B}$, понимаемая как независимость их индикаторов I_{A_i}. Равносильное условие:

$$P(A_{i_1} \cap \cdots \cap A_{i_k}) = P(A_{i_1}) \cdots P(A_{i_k}) \tag{1.6.3}$$

для каждого поднабора $\{i_1, \ldots, i_k\} \subset \{1, \ldots, n\}$ попарно различных индексов. Отметим, что это условие не равносильно такому равенству для всего набора. Ясно, что данное условие вытекает из независимости индикаторов. Обратно, независимость индикаторов из него следует. В самом деле, множество $\{\omega\colon I_{A_1}(\omega) \in B_1, \ldots, I_{A_n}(\omega) \in B_n\}$ непусто лишь для таких множеств B_i, которые содержат 0 или 1. Не содержащие таких точек множества приводят к нулевым вероятностям, содержащие обе можно исключить, а содержащие лишь одну из двух точек заменить этой точкой. При этом случай множества $B_i = \{0\}$ соответствует переходу в (1.6.3) от множества A_i к его дополнению. Однако такой переход возможен, поскольку для всяких A и B мы имеем

$$P\big(B \cap (\Omega \backslash A)\big) = P(B) - P(A \cap B),$$

так что фактически (1.6.3) влечет аналогичные равенства и при замене какой-либо части множеств на их дополнения. Разумеется, здесь также важно иметь (1.6.3) для всевозможных поднаборов.

Независимость событий не следует путать с их дизъюнктностью, которая в нетривиальных случаях как раз несовместима с независимостью (при $A \cap B = \varnothing$ имеем $P(A \cap B) = 0$, что не равно $PA)P(B)$ для событий с ненулевыми вероятностями).

В некоторых результатах бывает достаточно попарной независимости рассматриваемых случайных величин или событий, которая в общем случае слабее независимости. Например, на отрезке $[0,1]$ с мерой Лебега следующие три множества не являются независимыми, хотя попарно независимы: отрезок делим на четыре равных промежутка $I_1, \ldots, I_4$ длины $1/4$ и берем множества $I_1 \cup I_2$, $I_1 \cup I_3$, $I_1 \cup I_4$ меры $1/2$. Их попарные пересечения имеют меру $1/4$, как и пересечение всех трех, но произведение мер всех трех равно $1/8$, а не $1/4$.

Для независимых случайных величин заметно упрощается вычисление распределений функций от них.

1.6.4. Пример. Пусть случайные величины ξ и $\eta > 0$ независимы и имеют функции распределения F_ξ и F_η. Тогда их частное ξ/η имеет функцию распределения

$$F_{\xi/\eta}(t) = \int F_\xi(ty)\, P_\eta(dy).$$

Если есть плотности распределения ϱ_ξ и ϱ_η, то ξ/η также имеет плотность распределения

$$\varrho_{\xi/\eta}(t) = \int y\varrho_\xi(ty)\, \varrho_\eta(y)\, dy.$$

В самом деле, $F_{\xi/\eta}(t) = P(\xi < t\eta)$, что совпадает с мерой множества под прямой $x < ty$ относительно распределения (ξ, η) на плоскости, т. е. относительно произведения мер P_ξ и P_η.

Для независимых событий справедливо такое продолжение леммы Бореля – Кантелли.

1.6.5. Лемма. (Лемма Бореля – Кантелли для независимых событий) *Пусть $\{B_n\}$ — последовательность независимых событий и $\sum_{n=1}^\infty P(B_n) = \infty$. Тогда п.н. события B_n осуществляются бесконечно часто, т. е. множество $B = \bigcap_{k=1}^\infty \bigcup_{n=k}^\infty B_n$ имеет вероятность* 1.

Доказательство. Достаточно проверить, что при каждом k мы имеем $P\bigl(\bigcup_{n=k}^\infty B_n\bigr) = 1$. Для всякого $N \geqslant k$ имеем

$$1-P\Bigl(\bigcup_{n=k}^{\infty} B_n\Bigr) \leqslant 1-P\Bigl(\bigcup_{n=k}^{N} B_n\Bigr) = P\Bigl(\bigcap_{n=k}^{N} (\Omega\backslash B_n)\Bigr) = \prod_{n=k}^{N}\bigl(1-P(B_n)\bigr),$$

где последнее равенство следует из независимости событий $\Omega\backslash B_n$, вытекающей из независимости B_n (это пояснялось выше). Так как ряд из $P(B_n)$ расходится, то произведения $\prod_{n=k}^N \bigl(1 - P(B_n)\bigr)$ стремятся к нулю. $\square$

Разумеется, здесь важна независимость B_n. Для произвольных событий B_n при условии $P(B_n) \geqslant c > 0$ можно показать, что найдется такая подпоследовательность $\{n_k\}$, что пересечение $\bigcap_{k=1}^{\infty} B_{n_k}$ непусто. Для меры Лебега на $[0,1]$ можно даже найти подпоследовательность с континуальным пересечением, хотя не всегда можно получить положительную меру.

В заключение рассмотрим пример, показывающий, что вопрос о независимости тех или иных событий отнюдь не всегда легко определяется на основании интуитивных соображений. Пусть вероятностное пространство состоит из ста натуральных чисел $1,\ldots,100$ с равными вероятностями, событие A состоит в том, что выбранное наугад число четно, а событие B состоит в том, что оно делится на 5. Пересечение этих двух событий состоит из чисел, кратных 10, и имеет вероятность $10/100$, что совпадает с $P(A)P(B)$, ибо $P(A)=50/100$, $P(B)=20/100$. Таким образом, A и B независимы, что кажется интуитивно понятным. Однако теперь мы чуть изменим вероятностное пространство (возьмем числа до 99 или до 101). Количества точек в A, B и $A\cap B$ не изменились, но вероятности стали другими, так как теперь вероятность одной точки равна $1/99$ (или $1/101$ во втором случае). Равенство $P(A\cap B)=P(A)P(B)$ теперь нарушено: $10/99$ не равно произведению $50/99$ и $20/99$ (так же и для 101), поэтому теперь независимости нет. Можно исследовать вопрос, при каких N события A и B независимы в пространстве $1,\ldots,N$.

§ 1.7. Задачи

1.7.1. Сравнить с $1/2$ вероятность того, что в компании из 23 человек есть двое с совпадающим днем рождения.

1.7.2. На одном этаже некоего факультета есть три кофе-машины, из которых первая не работает никогда, вторая работает всегда, а третья работает или не работает с вероятностью $1/2$. Машины работают от рублевой монеты, и иностранный визитер (не знавший расстановки машин) с тремя такими монетами решил вычислить всегда работающую машину. Он опустил монету в одну машину, но успеха не достиг, затем он опустил монету в другую машину, которая сработала. Наконец, он опустил туда же последнюю монету, чтоб удостовериться в успехе, и машина сработала еще раз. Сравнить с $2/3$ вероятность того, что он попал на всегда работающую машину.

1.7.3. Найти вероятность того, что при двукратном бросании обычной игральной кости сумма выпавших чисел будет четна.

1.7.4. Найти вероятность того, что при трехкратном бросании игральной кости сумма выпавших чисел будет меньше 6.

1.7.5. Найти условную вероятность того, что при двукратном бросании игральной кости сумма выпавших чисел будет четна при условии, что четно первое выпавшее число.

1.7.6. Найти вероятность того, что в группе из 10 студентов не менее двух родились в один месяц.

1.7.7. Студент тратит треть стипендии на билеты государственной лотерии, а две трети на акции ЕЕЕ «Наколково». Какова вероятность полной потери вложенных средств, если в лотерии ничего не выигрывают 60 процентов билетов, а в ЕЕЕ «Наколково» не обесцениваются до нуля 10 процентов акций?

1.7.8. Случайная величина равномерно распределена в $[0, 2\pi]$. Найти ее дисперсию.

1.7.9. Случайная величина принимает значения $1, 2, 3$ с вероятностью $1/6$ и значения $4, 5$ с вероятностью $1/4$. Найти ее среднее.

1.7.10. Случайная величина с нулевым средним принимает значения $1, 2$ с вероятностью $1/3$ и еще одно значение p с вероятностью $1/3$. Чему равно p?

1.7.11. Пусть случайные величина ξ и η имеют конечные дисперсии $\mathrm{D}\xi$ и $\mathrm{D}\eta$, причем $\xi \leqslant \eta$. Верно ли, что $\mathrm{D}\xi \leqslant \mathrm{D}\eta$?

1.7.12. Случайная величина ξ принимает значения -1 и 1 с вероятностями p и $1-p$. При каком p ее дисперсия максимальна?

1.7.13. Найдите математические ожидания и дисперсии случайных величин с модельными распределения, указанными в § 1.5.

1.7.14. Доказать теорему 1.5.1, сначала рассмотрев случай, когда функция $\mu(-\infty, t)$ строго возрастает и непрерывна, используя обратную функцию.

1.7.15. Построить пример последовательности ограниченных случайных величин ξ_n, сходящихся к случайной величине ξ в среднем, но не имеющих общей интегрируемой мажоранты, т. е. таких, что случайная величина $\sup_n |\xi_n(\omega)|$ не является интегрируемой.

1.7.16. Доказать, что если последовательность случайных величин ξ_n сходится в среднем, то она имеет подпоследовательность $\{\xi_{n_k}\}$, обладающую интегрируемой мажорантой, т. е. такую, что случайная величина $\sup_k |\xi_{n_k}(\omega)|$ интегрируема.

1.7.17. Пусть $\xi_1, \ldots, \xi_n$ — независимые случайные величины. Доказать, что для всяких борелевских функций $f_1, \ldots, f_n$ на прямой случайные величины $f_1(\xi_1), \ldots, f_n(\xi_n)$ независимы.

1.7.18. Пусть $\xi_1, \ldots, \xi_n$ — независимые случайные величины. Доказать, что для всякой борелевской функции f на $\mathbb{R}^k$, где $k < n$, случайные величины $f(\xi_1, \ldots, \xi_k), \xi_{k+1}, \ldots, \xi_n$ независимы.

1.7.19. Пусть ξ и η — независимые случайные величины, причем распределение ξ симметрично относительно нуля, а η принимает значения -1 и 1 с вероятностью $1/2$. Доказать, что произведение $\xi\eta$ имеет такое же распределение, как ξ.

1.7.20. Привести пример случайных величин ξ и η с нулевыми математическими ожиданиями, не являющихся независимыми, но имеющих нулевое математическое ожидание произведения.

1.7.21. Привести пример стандартных гауссовских случайных величин ξ и η, не являющихся независимыми, но имеющих нулевое математическое ожидание произведения.

1.7.22. Привести пример стандартных гауссовских случайных величин ξ и η, для которых $\xi + \eta$ не является гауссовской.

1.7.23. Пусть ξ и η — независимые случайные величины, равномерно распределенные в $[0, 1]$. Найти $\mathrm{M}|\xi - \eta|$.

1.7.24. Пусть ξ и η — независимые стандартные гауссовские случайные величины. Доказать, что $\xi^2 + \eta^2$ и ξ/η независимы.

Глава 2

Предельные теоремы

В этой главе обсуждаются две важнейшие предельные теоремы теории вероятностей — закон больших чисел и центральная предельная теорема. Кроме того, вводится важное техническое средство — характеристические функционалы (преобразования Фурье) мер, играющее важную роль в установлении особого рода сходимости распределений случайных величин: слабой сходимости, кратко обсуждаемой в этой главе и имеющей непосредственное отношение к центральной предельной теореме.

§ 2.1. Закон больших чисел

Первой из двух великих предельных теорем теории вероятностей является так называемый закон больших чисел.

2.1.1. Теорема. (Закон больших чисел в форме Колмогорова) *Пусть $\{\xi_n\}$ — последовательность независимых случайных величин с одним и тем же распределением и средним m. Тогда*

$$\lim_{n\to\infty} \frac{\xi_1 + \cdots + \xi_n}{n} \to m \quad \textit{п.н.}$$

Несмотря на элементарность формулировки, доказательство при сделанных предположениях оказывается довольно трудным и не сразу было найдено (неслучайно, что для этого потребовался А.Н. Колмогоров). Однако при упрощающих предположениях доказательство становится несложным. Мы приведем один из таких результатов, но сначала установим следующий факт близкого характера (дающий меньше при предположениях, частично более слабых, частично более сильных).

2.1.2. Теорема. *Пусть $\{\xi_n\}$ — последовательность случайных величин с математическими ожиданиями m_i и равномерно ограниченными дисперсиями. Предположим, что эти случайные величины попарно некоррелированы, т. е.* $\mathrm{M}(\xi_i - m_i)(\xi_j - m_j) = 0$ *при* $i \neq j$ *(что* *имеет место при попарной независимости). Тогда*

$$\lim_{n\to\infty} \frac{(\xi_1 - m_1) + \cdots + (\xi_n - m_n)}{n} \to 0 \quad \textit{в среднем квадратическом,}$$

что дает также сходимость в среднем и по вероятности.

Доказательство. Ясно, что без ущерба для общности можно считать, что $m_i = 0$. Тогда надо проверить сходимость к нулю в L^2. Так как $\mathrm{D}[(\xi_1 + \cdots + \xi_n)/n] = n^{-2}(\mathrm{D}\xi_1 + \cdots + D\xi_n)$, то заключаем, что достаточно даже более слабого условия, чем ограниченность дисперсий: стремления к нулю величин $n^{-2}(\mathrm{D}\xi_1 + \cdots + \mathrm{D}\xi_n)$. □

Из этой теоремы следует, что имеется подпоследовательность $\{n_k\}$, для которой

$$\lim_{k\to\infty} \frac{(\xi_{n_1} - m_{n_1}) + \cdots + (\xi_{n_k} - m_{n_k})}{n_k} \to 0 \quad \text{п.н.}$$

Однако если соглашаться на подпоследовательности, то можно использовать следующий весьма общий результат Комлоша, в котором вообще почти нет условий (в частности, нет никаких условий независимости). Правда, доказательство его весьма трудно.

2.1.3. Теорема. (Теорема Комлоша) *Пусть $\{\xi_n\}$ — последовательность случайных величин с равномерно ограниченными математическими ожиданиями от $|\xi_n|$. Тогда имеется подпоследовательность $\{n_k\}$, для которой средние арифметические*

$$\lim_{k\to\infty} \frac{\xi_{n_1} + \cdots + \xi_{n_k}}{n_k}$$

сходятся почти наверное.

Условие равномерной ограниченности $\mathrm{M}|\xi_n|$ можно ослабить: достаточно иметь такую интегрируемую случайную величину $\eta > 0$, что равномерно ограничены $\mathrm{M}(|\xi_n|\eta)$. В самом деле, утверждение о сходимости п.н. не изменится при замене меры на эквивалентную.

Приведем теперь доказательство закона больших чисел при дополнительном сильно упрощающем дело предположении наличия четвертого момента, т. е. $\mathrm{M}|\xi_n|^4 < \infty$. Можно считать, что среднее равно нулю. Поэтому надо доказать сходимость средних арифметических п.н. к нулю. Множество B точек, где нет сходимости к нулю, представляет собой объединение счетного набора множеств B_k, $k \in \mathbb{N}$, таких точек ω, что для бесконечно многих номеров n выполнено неравенство $|\xi_1(\omega) + \cdots + \xi_n(\omega)|/n \geqslant k^{-1}$. Поэтому достаточно установить, что каждое B_k имеет нулевую вероятность. Здесь нам поможет лемма Бореля – Кантелли для произвольных событий, применяемая к множествам $A_n = \{\omega\colon |\xi_1(\omega) + \dots + \xi_n(\omega)|/n \geqslant k^{-1}\}$ при фиксированном k. Для оценки $P(A_n)$ заметим, что ввиду равенства нулю математических ожиданий и независимости

$$\mathrm{M}|\xi_1 + \dots + \xi_n|^4 = n\mathrm{M}\xi_1^4 + 6C_n^2(\mathrm{M}\xi_1^2)^2 \leqslant Cn^2$$

при некотором C, не зависящем от n. Теперь применим неравенство Чебышёва:

$$P(A_n) \leqslant k^4 n^{-4}\mathrm{M}|\xi_1 + \dots + \xi_n|^4 \leqslant Ck^4 n^{-2}.$$

Итак, ряд из $P(A_n)$ сходится, что завершает обоснование.

2.1.4. Пример. Вернемся к примеру 1.3.4 и применим его к среднему из n независимых случайных величин ξ_i с одинаковым распределением с математическим ожиданием m и дисперсией $\sigma^2 > 0$. Мы знаем, что

$$\mathrm{M}\frac{\xi_1 + \dots + \xi_n}{n} = m, \quad \mathrm{D}\frac{\xi_1 + \dots + \xi_n}{n} = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}.$$

Значит,

$$P\Bigl(\Bigl|\frac{\xi_1 + \dots + \xi_n}{n} - m\Bigr| \geqslant \frac{3\sigma}{\sqrt{n}}\Bigr) \leqslant \frac{1}{9}.$$

Следовательно, с вероятностью примерно 0.9 или выше $(\xi_1+\dots+\xi_n)/n$ отличается от m не больше, чем на $3\sigma/\sqrt{n}$. Скажем, если σ порядка единиц, то для n порядка десятков тысяч это отклонение составляет сотые. Эти довольно грубые расчеты могут служить основой примитивных статистических оценок: скажем, если проверяется некоторая модель с теоретической дисперсией 1 и теоретическим средним 10, то эмпирические данные с отклонением выборочного среднего от 10 заметно больше, чем на 0.03 при выборке размера 10000, становятся основанием для сомнений в правильности модели или чистоте эксперимента. В реальных статистических методах используются более тонкие средства, но причина популярности n в диапазоне $1500 - 2000$ в различных социологических опросах довольна близка к появлению 10000 выше.

В связи со сделанным ранее замечанием о предпочтительности медиан перед средними при оценке типичных значений следует подчеркнуть, что в сформулированном законе больших чисел речь шла о равнораспределенных независимых величинах.

§ 2.2. Характеристические функционалы

Характеристический функционал (преобразование Фурье) борелевской вероятностной меры μ на прямой задается формулой

$$\widetilde{\mu}(y) = \int \exp(iyx)\,\mu(dx), \quad y \in \mathbb{R}.$$

Комплексная функция $\widetilde{\mu}$ непрерывна (по теореме Лебега) и по модулю не превосходит 1, ибо $|\exp(iyx)| = 1$.

Для меры на $\mathbb{R}^d$ аналогично положим

$$\widetilde{\mu}(y) = \int \exp\bigl(i(y,x)\bigr)\,\mu(dx), \quad y \in \mathbb{R}^d.$$

Если мера μ задана плотностью ϱ, то $\widetilde{\mu}$ лишь множителем и умножением аргумента на -1 отличается от преобразования Фурье функции ϱ, задаваемого формулой

$$\widehat{f}(y) = (2\pi)^{-d/2} \int_{\mathbb{R}^d} \exp\bigl(-i(y,x)\bigr)\varrho(x)\,dx.$$

Если при этом функция $\widetilde{\mu}$ интегрируема (что не всегда верно), то плотность ϱ восстанавливается по ней с помощью так называемой формулы обращения

$$\varrho(x) = (2\pi)^{-d} \int_{\mathbb{R}^d} \exp\bigl(-i(y,x)\bigr)\widetilde{\mu}(y)\,dy. \tag{2.2.4}$$

Отметим, что эта формула дает сразу непрерывную версию плотности.

Характеристическим функционалом случайной величины ξ называется характеристический функционал ее распределения, т.е. функция

$$\varphi_\xi(y) = \mathrm{M}e^{iy\xi}.$$

Для меры Дирака в нуле имеем $\widetilde{\delta}(y) = 1$, для стандартной гауссовской меры на прямой имеем $\widetilde{\gamma}(y) = \exp(-y^2/2)$. Для остальных известных нам модельных распределений также явно находятся характеристические функционалы, однако в целом возможность явного вычисления скорее исключение из правил. Тем не менее характеристические функционалы играют важнейшую роль в теории и приложениях. Первым важным обстоятельством служит равенство мер с равными преобразованиями Фурье. Это отнюдь не очевидно из определения и может быть установлено несколькими способами. Один способ основан на равенстве Парсеваля

$$\int \widetilde{\mu}\, d\nu = \int \widetilde{\nu}\, d\mu, \tag{2.2.5}$$

вытекающем из теоремы Фубини для интеграла от $\exp\bigl(i(y,x)\bigr)$ по произведению мер μ и ν (в одном и другом порядке). Это равенство показывает, что меры μ_1 и μ_2 с равными преобразованиями Фурье приписывают одинаковые интегралы преобразованиям Фурье всевозможных мер. Тогда они приписывают одинаковые интегралы функциям класса C_0^∞, что видно из так называемой формулы обращения преобразования Фурье для функций этого класса: преобразование Фурье от функции $\widehat{f}$ в точке x равно $f(-x)$ (разумеется, эту формулу надо либо доказывать, либо считать известной из анализа).

Другой способ — обоснование следующей формулы для точек непрерывности функции распределения:

$$F_\mu(\beta) - F_\mu(\alpha) = \lim_{\sigma\to 0} \frac{1}{2\pi}\int_{-\infty}^{+\infty} \widetilde{\mu}(y) e^{-\sigma^2 y^2/2} \frac{e^{-iy\beta} - e^{-iy\alpha}}{-iy}\, dy.$$

2.2.1. Пример. Если $\xi_1, \ldots, \xi_n$ — независимые гауссовские случайные величины, то

$$c_1\xi_1 + \cdots + c_n\xi_n$$

есть гауссовская случайная величина для любых чисел $c_1, \ldots, c_n$. Ее математическое ожидание и дисперсия равны соответственно

$$c_1 \mathrm{M}\xi_1 + \cdots + c_n \mathrm{M}\xi_n, \quad c_1^2 \mathrm{D}\xi_1 + \cdots + c_n^2 \mathrm{D}\xi_n.$$

С помощью преобразования Фурье можно оценить степень концентрации меры в заданном интервале.

2.2.2. Лемма. *Пусть μ — борелевская вероятностная мера на прямой. Тогда*

$$\mu(\mathbb{R}\backslash[-2/r, 2/r]) \leqslant \frac{1}{r}\int_{-r}^{r} \bigl(1 - \widetilde{\mu}(y)\bigr)\, dy, \quad r > 0.$$

Доказательство. Правую часть с помощью теоремы Фубини мы можем переписать как

$$\frac{1}{r}\int_{-r}^{r}\int_{\mathbb{R}} (1 - e^{iyx})\, \mu(dx)\, dy = \int_{\mathbb{R}} 2\Bigl(1 - \frac{\sin(rx)}{rx}\Bigr)\, \mu(dx).$$

Кстати, из этого видно, что правая часть доказываемой формулы неотрицательна, что заранее не было очевидно. Так как $\sin s/s \leqslant 1$, то последнее выражение не меньше интеграла от подынтегральной функции лишь по дополнению отрезка $[-2/r, 2/r]$. На этом дополнении мы имеем $|rx| \geqslant 2$, т. е. $|\sin(rx)/rx| \leqslant 1/2$, откуда видно, что как раз получается не меньше μ-меры дополнения $[-2/r, 2/r]$. $\square$

Эта оценка нам понадобится ниже.

Для вывода центральной предельной теоремы ниже отметим, что если случайная величина ξ имеет конечный второй момент, то ее преобразование Фурье дважды дифференцируемо и справедливо равенство

$$\varphi_\xi^{(2)}(y) = -\mathrm{M}(\xi^2 e^{iy\xi}).$$

Более общим образом, если ξ имеет момент порядка k, то

$$\varphi_\xi^{(k)}(y) = i^k \mathrm{M}(\xi^k e^{iy\xi}).$$

В частности, $\varphi_\xi'(0) = i\mathrm{M}\xi$, $\varphi_\xi^{(2)}(0) = -\mathrm{M}\xi^2$.

Дифференцирование преобразования Фурье возможно по теореме о дифференцируемости интеграла по параметру (вытекающей из теоремы Лебега), согласно которой

$$\frac{d}{d\alpha}\int f(\alpha, x)\,\mu(dx) = \int \frac{\partial}{\partial\alpha} f(\alpha, x)\,\mu(dx),$$

если $|f(\alpha, x)| + |\partial f(\alpha, x)/\partial\alpha| \leqslant \Phi(x)$ для некоторой μ-интегрируемой функции Φ.

§ 2.3. Слабая сходимость распределений

В § 1.4 речь шла о некоторых видах сходимости случайных величин, близких к сходимости в индивидуальных точках (с точностью до подпоследовательности). Однако есть еще один вид сходимости иного рода, оказавшийся весьма важным в приложениях. Этот вид сходимости возник очень давно как поточечная сходимость функций распределения. Часто более удобна иная техническая формулировка.

2.3.1. Определение. *Последовательность вероятностных борелевских мер μ_n на $\mathbb{R}^d$ слабо сходится к борелевской вероятностной мере μ_n, если для каждой ограниченной непрерывной функции f на $\mathbb{R}^d$ верно равенство*

$$\lim_{n\to\infty} \int_{\mathbb{R}^d} f\,d\mu_n = \int_{\mathbb{R}^d} f\,d\mu. \tag{2.3.6}$$

Обозначение: $\mu_n \Rightarrow \mu$. *Случайные векторы ξ_n сходятся по распределению к случайному вектору ξ, если их распределения P_{ξ_n} слабо сходятся к P_ξ.*

В терминах математических ожиданий условие записывается как сходимость $\mathrm{M}f(\xi_n) \to \mathrm{M}f(\xi)$.

Ниже приведены равносильные описания слабой сходимости. Еще одно такое описание в терминах преобразования Фурье будет приведено в следующем параграфе.

Прежде рассмотренные виды сходимости влекут сходимость по распределению, так как сходимость с.в. $\xi_n \to \xi$ по вероятности дает сходимость по вероятности $f(\xi_n) \to f(\xi)$ для ограниченных непрерывных f, что по теореме Лебега влечет сходимость математических ожиданий.

Обратное очевидным образом неверно: скажем, можно взять не сходящуюся по вероятности последовательность случайных величин с одинаковыми распределениями (простейший вариант: $(-1)^n\xi$, где ξ принимает значения -1 и 1 с вероятностями $1/2$).

Однако А.В. Скороход обнаружил следующий замечательный факт (для общих метрических пространств, но здесь он приведен для прямой).

2.3.2. Теорема. *Если случайные величины ξ_n таковы, что их распределения P_{ξ_n} слабо сходятся к распределению случайной величины ξ, то найдутся такие случайные величины η_n, п.н. сходящиеся к случайной величине η, что $P_{\eta_n} = P_{\xi_n}$ для всех n и $P_\eta = P_\xi$.*

Таким образом, «подменой» случайных величин можно из слабой сходимости распределений получить сходимость п.н. самих случайных величин.

Очевидным следствием слабой сходимости является так называемая *равномерная плотность* мер μ_n:

$$\lim_{R\to\infty} \sup_n \mu_n(x \in \mathbb{R}^d\colon\ |x| \geqslant R) = 0.$$

Достаточно рассмотреть функции f, для которых $0 \leqslant f \leqslant 1$, причем $f = 1$ на достаточно большом шаре. Как мы увидим ниже, это необходимое условие довольно близко (с точностью до подпоследовательности) к достаточному.

2.3.3. Теорема. *Для слабой сходимости вероятностных борелевских мер μ_n на прямой к вероятностной мере μ необходимо и достаточно иметь равенство*

$$\lim_{n\to\infty}\int_{\mathbb{R}^d} f\,d\mu_n = \int_{\mathbb{R}^d} f\,d\mu, \quad f\in C_0^\infty(\mathbb{R}), \tag{2.3.7}$$

причем достаточно иметь его для функций f вида $f(x)=x^k\theta(x/m)$, где $k=0,1,\ldots$, $m\in\mathbb{N}$, $\theta\in C_0^\infty(\mathbb{R})$ — какая-либо фиксированная функция с $0\leqslant f\leqslant 1$, равная 1 на $[-1,1]$ и 0 вне $[-2,2]$. Аналогичное верно для мер на $\mathbb{R}^d$.

Доказательство. В проверке нуждается лишь утверждение о достаточности (2.3.7) с функциями указанного вида. Взяв $k=0$ и используя сделанное выше замечание, получаем равномерную плотность мер μ_n. Пусть $f\in C_b(\mathbb{R})$ и $\varepsilon>0$. Можно считать, что $|f|\leqslant 1$. Предположим сначала, что $f=0$ вне отрезка $[-R,R]$, причем число $R\in\mathbb{N}$ выбрано так, что

$$\mu_n(\mathbb{R}\backslash[-R,R])+\mu(\mathbb{R}\backslash[-R,R])<\varepsilon \quad \forall\, n.$$

Найдется многочлен ψ, для которого

$$|\psi(x)-f(x)|\leqslant\varepsilon \quad \text{при } |x|\leqslant 2R.$$

Заметим, что

$$|\theta(x/R)\psi(x)-\theta(x/R)f(x)|\leqslant\varepsilon \quad \forall\, x\in\mathbb{R},$$

так как $0\leqslant\theta(x/R)\leqslant 1$ и $\theta(x/R)=0$ при $|x|\geqslant 2R$. При этом $\theta(x/R)f(x)=f(x)$ для всех x.

По нашему условию для функции $g(x)=\psi(x)\theta(x/R)$ существует такое n_1, что

$$\left|\int g\,d\mu-\int g\,d\mu_n\right|<\varepsilon \quad \forall\, n\geqslant n_1.$$

Значит,

$$\left|\int f\,d\mu-\int f\,d\mu_n\right|\leqslant 3\varepsilon,$$

ибо интегралы от функций $f(x)=\theta(x/R)f(x)$ и $g(x)$ по всякой вероятностной мере отличаются не более чем на ε.

Наконец, случай общей функции $f\in C_b(\mathbb{R})$ сводится к доказанному в силу равномерной плотности мер μ_n путем замены $f(x)$ на $f(x)\theta(x/R)$ при достаточно большом R. $\square$

Отметим, что в этой теореме важно, что речь идет о вероятностных мерах. Например, интегралы от всякой функции с компактным носителем по мерам Дирака δ_n в точках n становятся нулевыми при больших n, но эти меры не сходятся слабо.

Важное практическое значение последнего утверждения теоремы состоит в том, что тестировать сходимость можно по счетному множеству функций, хотя пространство $C_b(\mathbb{R})$ с его $\sup$-нормой несепарабельно. Переход к счетному множеству функций дает возможность метризации слабой сходимости. В следующей теореме фигурирует одна из возможных метрик — метрика Канторовича (который ввел ее в случае мер на компакте, а для некомпактных множеств такую метрику называют метрикой Форте – Мурье или метрикой Канторовича – Рубинштейна), задаваемая на множестве вероятностных мер на прямой формулой

$$d_K(\mu,\nu) := \sup\biggl\{\int f\,d\mu - \int f\,d\nu\colon\ f\in \mathrm{Lip}_1,\ \sup_x |f(x)|\leqslant 1\biggr\},$$

где Lip_1 — класс 1-липшицевых функций, т. е. функций f, для которых $|f(x)-f(y)|\leqslant |x-y|$.

Легко видеть, что это действительно метрика.

2.3.4. Теорема. *Слабая сходимость вероятностных борелевских мер μ_n на прямой к борелевской вероятностной мере μ равносильна соотношению $d_K(\mu_n,\mu)\to 0$.*

Доказательство. Достаточность ясна из предыдущей теоремы. Необходимость же формально есть более сильное условие, а именно равномерность сходимости интегралов по функциям из указанного класса. Проверим, что такая равномерность автоматически имеет место. Пусть $\varepsilon>0$. Мы уже знаем, что найдется отрезок $[-R,R]$ с мерой более $1-\varepsilon$ для всех мер μ_n и μ. По теореме Асколи – Арцела множество 1-липшицевых функций на $[-R,R]$ с супремумом модуля не более 1 компактно по sup-норме. Значит, в нем есть конечная ε-сеть $f_1,\ldots,f_m$, т.е. для всякой функции f из этого множества найдется $i\leqslant m$, для которого $|f(x)-f_i(x)|\leqslant\varepsilon$ при всех $x\in[-R,R]$. Функции f_i продолжим постоянными значениями вне $[-R,R]$, что даст функции класса Lip_1 на прямой с модулем не более 1. Продолжения обозначим прежними символами. Для полученного конечного набора функций найдется такой номер n_1, что при $n\geqslant n_1$ интегралы от $f_1,\ldots,f_m$ по мерам μ_n отличаются от интегралов по μ меньше, чем на ε. Теперь для всякой функции $f\in\mathrm{Lip}_1$ со значениями в $[-1,1]$ мы можем взять функцию

f_i с $\sup_{|x|\leqslant R}|f(x)-f_i(x)|\leqslant\varepsilon$. Тогда при $n\geqslant n_1$ имеем

$$\begin{aligned}&\left|\int f\,d\mu_n-\int f\,d\mu\right|=\\&=\left|\int(f-f_i)\,d\mu_n+\int f_i\,d\mu_n-\int f_i\,d\mu-\int(f-f_i)\,d\mu\right|\leqslant\\&\leqslant 4\sup_{x\in[-R,R]}|f(x)-f_i(x)|+2\varepsilon+\varepsilon\leqslant 7\varepsilon,\end{aligned}$$

ибо интеграл от $f-f_i$ по дополнению $[-R,R]$ не больше удвоенной меры этого дополнения, а такие меры не больше ε по всем мерам μ_n и μ. □

Свяжем теперь слабую сходимость с классической сходимостью функций распределения. Для этого воспользуемся следующей формулой интегрирования по частям:

$$\int_{\mathbb{R}} f(t)\,P(dt)=-\int_{-\infty}^{+\infty}f'(t)\,F_P(t)\,dt,\quad f\in C_0^\infty(\mathbb{R}),$$

где $F_P(t)=P\bigl((-\infty,t)\bigr)$. Эта формула сводит интегралы от финитных гладких функций по общим вероятностным распределениям к римановским интегралам по отрезкам (формальная манипуляция состоит в замене $P(dt)$ выражением $dF_P(t)$ и последующим формальным преобразованием Ньютона – Лейбница).

Для обоснования запишем $f(t)$ в виде

$$f(t)=\int_{-\infty}^{t}f'(s)\,ds=\int_{-\infty}^{+\infty}I_{(-\infty,t)}(s)f'(s)\,ds,$$

а затем подставим это выражение в интеграл по мере P и переставим пределы интегрирования по теореме Фубини (что возможно при наших предположениях). Интеграл от функции $I_{(-\infty,t)}(s)=I_{[s,+\infty)}(t)$ по мере P даст $P\bigl([s,+\infty)\bigr)=1-F_P(s)$. При интегрировании по s этого выражения, умноженного на $f'(s)$, останется только интеграл от $-F_P(s)f'(s)$, ибо интеграл от f' равен нулю (здесь важно, что f обращается в нуль вне некоторого отрезка).

2.3.5. Теорема. *Слабая сходимость вероятностных борелевских мер μ_n на прямой к борелевской вероятностной мере μ равносильна соотношению $F_{\mu_n}(t)\to F_\mu(t)$ во всех точках непрерывности функции F_μ.*

Доказательство. Пусть $\mu_n \Rightarrow \mu$ и t — точка непрерывности F_μ (напомним, что возрастающая функция F_μ имеет лишь конечное или счетное множество точек разрыва). Пусть $\varepsilon > 0$. В силу непрерывности F_μ в t есть такое $\delta > 0$, что $F_\mu(t-\delta) > F_\mu(t) - \varepsilon$. Возьмем функцию f, равную 1 на $(-\infty, t-\delta]$, равную 0 на $[t, +\infty)$ и доопределенную аффинно на $[t-\delta, t]$. Интеграл от f по всякой вероятностной мере не больше значения этой меры на $(-\infty, t)$. При всех достаточно больших n имеем

$$F_{\mu_n}(t) \geqslant \int f\, d\mu_n \geqslant \int f\, d\mu - \varepsilon \geqslant F_\mu(t-\delta-\varepsilon > F_\mu(t) - 2\varepsilon.$$

Рассматривая аналогично кусочно линейную функцию, равную 1 на $(-\infty, t]$ и 0 на $[t+\delta, +\infty)$, получаем, что $F_{\mu_n}(t) \leqslant F_\mu(t) + 2\varepsilon$ при всех достаточно больших n.

Обратно, пусть дана сходимость функций распределения в точках непрерывности F_μ. Тогда для всякой функции $f \in C_0^\infty(\mathbb{R})$ имеет место сходимость $f'(t)F_{\mu_n}(t) \to f'(t)F_\mu(t)$ за исключением точек не более чем счетного множества. По теореме Лебега о мажорированной сходимости и доказанной выше формуле интегрирования по частям получаем сходимость интегралов от f по мерам μ_n к интегралу по мере μ. Теорема 2.3.3 дает слабую сходимость. $\square$

Можно проверить, что в этой теореме подходит любое всюду плотное множество точек, где есть сходимость функций распределения.

Завершим этот параграф следующей теоремой Хелли (или Хелли – Брея).

2.3.6. Теорема. *Пусть последовательность вероятностных мер на прямой равномерно плотна. Тогда из нее можно выбрать слабо сходящуюся подпоследовательность.*

Доказательство. Из равномерно ограниченной последовательности функций F_{μ_n} можно выбрать подпоследовательность, сходящуюся в каждой рациональной точке (диагональный прием Кантора). Получена возрастающая функция на множестве рациональных чисел. При этом ее пределы на $-\infty$ и $+\infty$ равны 0 и 1 соответственно, что вытекает из равномерной плотности данных мер. Легко видеть, что существует возрастающая функция F на прямой, продолжающая построенную. Заметим, что $F_{\mu_n}(t) \to F(t)$ в точках непрерывности F. В самом деле, пусть $\varepsilon > 0$. Найдутся рациональные точки $t_1 < t$ и $t_2 > t$, для которых $F(t_2) - F(t_1) < \varepsilon$. Найдется также такой номер n_1,

что
$$|F_{\mu_n}(t_1) - F(t_1)| < \varepsilon, \quad |F_{\mu_n}(t_2) - F(t_2)| < \varepsilon \quad \forall\, n \geqslant n_1.$$
Тогда $F_{\mu_n}(t) \geqslant F_{\mu_n}(t_1) \geqslant F(t_1) - \varepsilon \geqslant F(t) - 2\varepsilon$ при $n \geqslant n_1$. Аналогично $F_{\mu_n}(t) \leqslant F(t) + 2\varepsilon$. Таким образом, вне счетного множества есть сходимость $F_{\mu_n}(t) \to F(t)$.

Теперь переопределим F в точках разрыва (их не более счетного числа) и сделаем непрерывной слева. Получится функция распределения F_μ некоторой борелевской вероятностной меры μ. Это переопределение не испортит сходимость $F_{\mu_n}(t) \to F_\mu(t)$ вне некоторого счетного множества. Как отмечено выше, это влечет сходимость в точках непрерывности F_μ, что обеспечивает слабую сходимость мер μ_n к μ. □

Нетрудно перенести доказанное утверждение на случай $\mathbb{R}^d$. Более того, можно показать, что аналогичное утверждение верно для всякого полного сепарабельного метрического пространства. Фундаментальная теорема Ю.В. Прохорова утверждает, что если последовательность борелевских вероятностных мер μ_n на полном сепарабельном метрическом пространстве X слабо сходится, то она равномерно плотна в следующем смысле: для всякого $\varepsilon > 0$ найдется такой компакт K_ε, что
$$\sup_n \mu_n(X \backslash K_\varepsilon) \leqslant \varepsilon.$$
Обратно, если последовательность борелевских вероятностных мер μ_n на X равномерно плотна, то из нее можно выделить слабо сходящуюся подпоследовательность.

Естественно возникает вопрос об эффективной проверке условия равномерной плотности последовательности мер.

2.3.7. Предложение. *Последовательность вероятностных борелевских мер μ_n на $\mathbb{R}^d$ равномерно плотна в точности тогда, когда найдется такая возрастающая к бесконечности непрерывная функция $V\colon\ [0, +\infty) \to [0, +\infty)$, что*
$$\sup_n \int_{\mathbb{R}^d} V(|x|)\, \mu_n(dx) < \infty.$$

Доказательство. Из неравенства Чебышёва легко усмотреть достаточность этого условия, взяв $r_k \to +\infty$ так, что $V(|x|) \geqslant k$ вне шара радиуса r_k. Для установления необходимости надо взять возрастающие к бесконечности числа R_j так, что
$$\sup_n \mu_n(\mathbb{R}^d \backslash U_{R_j}) < 2^{-j}, \quad U_{R_j} = \{x\colon |x| < R_j\},$$

Затем можно взять непрерывную возрастающую функцию V, для которой $V(R_j) = j$. □

Например, для мер на прямой достаточно иметь оценку

$$\int |x|^{\alpha}\, \mu_n(dx) \leqslant C < \infty$$

при некотором $\alpha > 0$. В частности, равномерно плотна последовательность вероятностных мер с равномерно ограниченными средними и равномерно ограниченными дисперсиями.

Теперь докажем важную теорему П. Леви о слабой сходимости в терминах характеристических функционалов.

2.3.8. Теорема. *Слабая сходимость вероятностных борелевских мер μ_n на прямой к борелевской вероятностной мере μ равносильна поточечной сходимости характеристических функционалов*:

$$\widetilde{\mu_n}(y) \to \widetilde{\mu}(y).$$

Это же верно и для мер на $\mathbb{R}^d$.

Доказательство. Необходимость сходимости характеристических функционалов очевидна. Проверим ее достаточность двумя разными способами. Первый способ состоит в применении равенства Парсеваля (2.2.5) вместе с использованием менее элементарной формулы обращения (2.2.4) и того факта, что для всякой функции $f \in C_0^{\infty}(\mathbb{R}^d)$ ее преобразование Фурье (в смысле преобразования Фурье в L^1) входит в L^1. Таким образом, $f(-x)$ есть характеристический функционал комплексной меры ν с плотностью $(2\pi)^{-d}\widetilde{\sigma}$, где σ — мера с плотностью f. Тогда формула (2.2.5), очевидным образом верная и для комплексных мер, дает равенства

$$\int_{\mathbb{R}^d} f(-x)\, \mu_n(dx) = \int_{\mathbb{R}^d} \widetilde{\mu}_n(y)\, \nu(dy),$$

а также аналогичное равенство для μ. Поэтому в силу теоремы Лебега получаем сходимость интегралов от функций класса $C_0^{\infty}(\mathbb{R}^d)$. Кстати, из этого обоснования видно, что достаточно иметь сходимость $\widetilde{\mu}_n(y) \to \widetilde{\mu}(y)$ почти всюду.

Можно обойтись и без формулы обращения. Покажем сначала, что последовательность мер μ_n равномерно плотна. Пусть $\varepsilon > 0$. При фиксированном $r > 0$ интегралы от $1 - \widetilde{\mu_n}(y)$ по $[-r, r]$ с мерой Лебега сходятся к интегралу от $1 - \widetilde{\mu}(y)$ по $[-r, r]$. Ввиду непрерывности $\widetilde{\mu}$ и равенства $\widetilde{\mu}(0) = 1$ можно найти столь малое $r > 0$, что

$$\frac{1}{r}\int_{-r}^{r}\bigl(1-\widetilde{\mu}(y)\bigr)\,dy < \varepsilon.$$

Из сходимости указанных интегралов и леммы ясно, что тогда

$$\mu_n(\mathbb{R}\backslash[-2/r, 2/r]) < \varepsilon$$

при всех достаточно больших n. Итак, $\{\mu_n\}$ равномерно плотна.

Теперь вместе со сходимостью преобразований Фурье это дает и слабую сходимость. Для этого можно либо применить теорему Хелли – Брея и заметить, что по свойству единственности все слабо сходящиеся подпоследовательности имеют общий предел, либо использовать теорему 2.3.3. В этом случае надо лишь заметить, что функция $f \in C_0^\infty(\mathbb{R})$ на достаточно большом отрезке $[-\pi m, \pi m]$, содержащем ее носитель, равномерно приближается $2\pi m$-периодическими функциями вида $T(x) = \sum_{j=-k}^{k} c_j e^{ijx/m}$, причем для такой функции $\sup_x |T(x)| = \sup_{|x|\leqslant 2\pi m} |T(x)|$. От таких функций интегралы по мерам μ_n сходятся к интегралу по мере μ, что позволяет сделать вывод и о сходимости интегралов от f, ибо интеграл от функции T по мере μ_n есть $\sum_{j=-k}^{k} c_j \widetilde{\mu_N}(j/m)$. $\square$

Имеет место и многомерный вариант теоремы Леви, доказываемый аналогично.

§ 2.4. Центральная предельная теорема

В этом параграфе приведен один из важнейших результатов теории вероятностей — центральная предельная теорема, в которой речь идет о предельных распределениях нормированных сумм независимых случайных величин. Предварительно мы обсудим, что происходит с функциями распределения, плотностями и характеристическими функционалами при сложении независимых случайных величин. Ранее мы уже видели, что математические ожидания и дисперсии просто складываются (впрочем, для математических ожиданий при этом не нужна и независимость).

С преобразованием Фурье все очевидно: если ξ и η независимы, то

$$\mathrm{M}e^{iy(\xi+\eta)} = \mathrm{M}(e^{iy\xi}e^{iy\eta}) = \mathrm{M}e^{iy\xi}\mathrm{M}e^{iy\eta},$$

т.е. получаем произведение преобразований Фурье:

$$\varphi_{\xi+\eta} = \varphi_\xi\varphi_\eta.$$

С распределениями положение несколько сложнее, но и здесь ответ выражается простой формулой

$$P_{\xi+\eta} = P_\xi * P_\eta, \tag{2.4.8}$$

если ввести операцию свертки мер μ и ν по формуле

$$\mu * \nu(B) := \int \mu(B-x)\,\nu(dx).$$

Это можно сделать как на прямой, так и в $\mathbb{R}^d$. Технически проще всего задать свертку $\mu*\nu$ как образ произведения $\mu\otimes\nu$ при отображении $(x,y)\mapsto x+y$. Тогда значение $\mu*\nu(B)$ есть интеграл от I_B от образа произведения, что по формуле замены переменных есть интеграл от $I_B(x+y) = I_{B-y}(x)$ по произведению $\mu\otimes\nu$. Вычисление этого интеграла по теореме Фубини и дает указанное выше выражение. Заодно получаем и равенство

$$\mu*\nu = \nu*\mu.$$

Итак, свертка коммутативна. Множество ограниченных мер с операцией свертки превращается в коммутативную алгебру, а мера Дирака в нуле оказывается единицей этой алгебры. Интегралы функций по мере $\mu*\nu$ вычисляются по формуле

$$\int f(u)\,(\mu*\nu)(du) = \int\int f(x+y)\,\mu(dx)\,\nu(dy).$$

Для преобразований Фурье получаем формулу

$$\widetilde{\mu*\nu} = \widetilde{\mu}\widetilde{\nu},$$

из которой в силу однозначной определенности мер преобразованиями Фурье следует формула (2.4.8).

Если распределения P_ξ и P_η обладают плотностями ϱ_ξ и ϱ_η, то приходим к формуле

$$\varrho_{\xi+\eta}(x) = \varrho_\xi * \varrho_\eta(x) := \int \varrho_\xi(x-y)\,\varrho_\eta(y)\,dy, \tag{2.4.9}$$

где используется свертка интегрируемых функций. Отметим, что свертка интегрируемых функций f и g определена по указанной формуле и интегрируема. Это видно из теоремы Фубини, дающей интегрируемость функции двух переменных $f(u)g(v)$. Сделав замену $u=x-y$, $v=y$, получаем интегрируемую функцию $f(x-y)g(y)$ двух переменных, но тогда та же теорема Фубини говорит, что при почти каждом x функция одного переменного $y\mapsto f(x-y)g(y)$ интегрируема, причем ее интеграл по y даст функцию, интегрируемую по оставшемуся переменному x. Чтобы убедиться, что мера $P_\xi * P_\eta$ задается плотностью $\varrho_\xi * \varrho_\eta$, достаточно заметить, что преобразование Фурье меры с плотностью $\varrho_\xi * \varrho_\eta$ есть произведение преобразований Фурье мер с плотностями ϱ_ξ и ϱ_η, т. е. как раз P_ξ и P_η (это тоже ясно из теоремы Фубини).

Теперь понятно, что при сложении n независимых случайных величин $\xi_1, \ldots, \xi_n$ преобразование Фурье суммы оказывается равным $\varphi_{\xi_1} \cdots \varphi_{\xi_n}$. Нас будет интересовать еще более специальный случай, когда ξ_i имеют равные распределения. Тогда

$$\varphi_{\xi_1+\cdots+\xi_n} = \varphi_{\xi_1}^n.$$

Перейдем к основному результату — центральной предельной теореме.

2.4.1. Теорема. *Пусть* $\{\xi_n\}$ — *последовательность независимых случайных величин с одним и тем же распределением с нулевым средним и дисперсией* 1. *Тогда распределения нормированных сумм*

$$S_n := \frac{\xi_1 + \cdots + \xi_n}{\sqrt{n}}$$

слабо сходятся к стандартному гауссовскому распределению.

Доказательство. Из приведенных выше простых вычислений ясно, что

$$\varphi_{S_n}(y) = \varphi_{\xi_1+\cdots+\xi_n}(y/\sqrt{n}) = \varphi_1(y/\sqrt{n})^n.$$

Как отмечено выше, наличие дисперсии влечет существование второй производной функции φ_{ξ_1} и равенство

$$\varphi_{\xi_1}''(0) = -\mathrm{M}\xi_1^2 = -1.$$

При этом равенство нулю математического ожидание показывает, что $\varphi_{\xi_1}'(0) = 0$. Следовательно,

$$\varphi_{\xi_1}(s) = 1 - s^2/2 + o(s^2), \quad s \to 0,$$

откуда мы получаем, что

$$\varphi_{\xi_1}(y/\sqrt{n})^n = \bigl(1 - y^2/(2n) + o(n^{-1})\bigr)^n \to e^{-y^2/2},$$

что по теореме Леви дает слабую сходимость мер P_{S_n} к стандартной гауссовской. Стоит отметить, что случайные величины S_n имеют нулевые средние и дисперсии 1, поэтому неравенство Чебышёва сразу дает равномерную плотность их распределений, что вместе с поточечной сходимостью характеристических функционалов влечет слабую сходимость и без использования теоремы Леви. □

Важным отличием этой теоремы от закона больших чисел является не только то, что теперь в знаменателе стоит $\sqrt{n}$ вместо n, но и иной характер сходимости: здесь не утверждается сходимость почти наверное или по вероятности. Такой сходимости нет даже в том случае, когда все ξ_n уже сами являются стандартными гауссовскими случайными величинами, так что и нормированные суммы S_n таковы. При этом разность

$$S_{n+m} - S_n =$$
$$= \big((n+m)^{-1/2} - n^{-1/2}\big)(\xi_1 + \cdots + \xi_n) + (n+m)^{-1/2}(\xi_{n+1} + \cdots + \xi_{n+m})$$

есть (как видно из примера выше) гауссовская случайная величина с нулевым средним и дисперсией

$$n\big((n+m)^{-1/2} - n^{-1/2}\big)^2 + m(n+m)^{-1}.$$

При $m = n$ эта дисперсия больше $1/2$, поэтому $P(|S_{2n} - S_n| > 1)$ не стремится к нулю при $n \to +\infty$, т. е. последовательность случайных величин S_n не фундаментальна по вероятности.

Для случайных величин с ненулевым математическим ожиданием и ненулевой дисперсией $\sigma^2 > 0$ получаем такое утверждение.

2.4.2. Следствие. *Пусть $\{\xi_n\}$ — последовательность независимых случайных величин с одним и тем же распределением со средним m и дисперсией $\sigma^2 > 0$. Тогда распределения случайных величин*

$$\frac{\xi_1 + \cdots + \xi_n - nm}{\sigma\sqrt{n}}$$

слабо сходятся к стандартному гауссовскому распределению.

Ясно также, что в случае нулевого математического ожидания распределения указанных выше нормированных сумм S_n слабо сходятся к гауссовскому распределению с нулевым средним и дисперсией σ^2 (в этой форме, конечно, и для нулевой дисперсии утверждение верно).

Замечательная особенность сходимости в центральной предельной теореме заключается в независимости предела от конкретного вида распределения ξ_1 (при этом параметры предела определяются по среднему и дисперсии ξ_1). Это обстоятельство (наряду с законом больших чисел) лежит в основе множества результатов и методов статистики. В этой связи уместно привести несколько общих фактов, связанных с гауссовскими распределениями.

Быстрое убывание гауссовской плотности на бесконечности приводит к еще более быстрому убыванию меры множества больших значений:

$$\frac{1}{\sqrt{2\pi}}\int_R^{+\infty} e^{-x^2/2}\,dx \sim \frac{1}{R\sqrt{2\pi}}e^{-R^2/2}, \quad R\to+\infty.$$

Далее, вычисления показывают, что

$$\frac{1}{\sqrt{2\pi}}\int_{-3}^{3} e^{-x^2/2}\,dx \approx 0.997,$$

что приводит к так называемому *правилу трех сигма* (ср. примеры 1.3.4 и 2.1.4): для гауссовской величина ξ с дисперсией σ^2 выполнено приближенное равенство

$$P(|\xi - \mathrm{M}\xi| \geqslant 3\sigma) \approx 0.003,$$

т. е. распределение сосредоточено на отрезке от $\mathrm{M}\xi - 3\sigma$ до $\mathrm{M}\xi + 3\sigma$ с точностью до 0.003. Таким образом, для гауссовских случайных величин заметно повышается вероятность попадания в указанный диапазон (по сравнению с общим примером 2.1.4). Это обстоятельство существенно проявляется в прикладных статистических методах. В частности, центральная предельная теорема подсказывает такой способ статистических проверок: для эмпирической выборки $x_1,\ldots,x_n$ с теоретическим распределением с нулевым средним и дисперсией σ^2 можно рассматривать числа $S=(x_1+\cdots+x_n)/(\sigma\sqrt{n})$ и сравнивать их со значением R, найденным из уравнения $\Phi(R)=1-\alpha$, где Φ — стандартная нормальная функция распределения, с заданным малым уровнем α (например, $\alpha=0.1$ или $\alpha=0.05$). Приближенное значение для R можно найти из таблиц или с помощью вычислительных средств; грубые приближения можно сделать и вручную). Заметное превышение эмпирическим числом S «теоретического значения» R может трактоваться как причина сомнений в верности модели или чистоте эксперимента. Обратим внимание на то обстоятельство, что здесь происходит подмена реального распределения асимптотическим предельным нормальным распределением (именно для последнего имеются столь точные формулы). Конечно, возникает вопрос о правомерности такой подмены. Для оправдания этих действий нужны оценки сходимости в центральной предельной теореме, что оказывается заметно более сложным вопросом (которым на практике, однако, часто пренебрегают, предполагая, что теория вероятностей уже оправдала все такие манипуляции). Если ξ_1 имеет еще и третий момент, то теорема Берри–Эссеена дает скорость сходимости: в ситуации теоремы 2.4.1 справедливо неравенство

$$\sup_t |F_{S_n}(t) - \Phi(t)| \leqslant \frac{\mathrm{M}|\xi_1|^3}{\sqrt{n}},$$

где Φ — стандартная нормальная функция распределения. Как видим, и здесь возникает квадратный корень из n.

Еще одно очень важное свойство гауссовских распределений состоит в том, что если случайный вектор $(\xi_1, \ldots, \xi_n)$ имеет гауссовское распределение (что равносильно тому, что все линейные комбинации $c_1\xi_1 + \cdots + c_n\xi_n$ являются гауссовскими случайными величинами), то из некоррелированности компонент, т. е. равенства

$$\mathrm{M}[(\xi_i - \mathrm{M}\xi_i)(\xi_j - \mathrm{M}\xi_j)] = 0, \quad i \neq j,$$

следует их независимость. Для упрощения выкладок проверим это для двух случайных величин ξ_1 и ξ_2. Можно считать, что они имеют нулевые средние. Нам надо показать, что распределение вектора (ξ_1, ξ_2) равно произведению распределений компонент. Это равносильно тому, что произведение преобразований Фурье $\varphi_{\xi_1}(y_1)\varphi_{\xi_2}(y_2)$ равно значению преобразования совместного распределения в точке (y_1, y_2), что по формуле замены переменных есть $\mathrm{M}\exp(iy_1\xi_1 + iy_2\xi_2)$. Так как $y_1\xi_1 + iy_2\xi_2$ — гауссовская случайная величина с нулевым математическим ожиданием, причем ее дисперсия равна сумме дисперсий $y_1\xi_1$ и $y_2\xi_2$ из-за некоррелированности, то получаем равенство

$$\mathrm{M}\exp(iy_1\xi_1 + iy_2\xi_2) = \exp(-y_1^2\mathrm{D}_{\xi_1}/2 - y_2^2\mathrm{D}_{\xi_2}/2) = \varphi_{\xi_1}(y_1)\varphi_{\xi_2}(y_2),$$

что и требовалось проверить.

Надо иметь в виду, что гауссовость распределения случайного вектора (ξ, η) не вытекает из гауссовости его компонент: нужна гауссовость всех линейных комбинаций $c_1\xi + c_2\eta$.

Из сферической инвариантности стандартного гауссовского распределения в $\mathbb{R}^n$ сразу следует такой полезный факт: если случайный вектор ξ в $\mathbb{R}^n$ имеет стандартное гауссовское распределение, то для всякого ортогонального линейного оператора U в $\mathbb{R}^n$ случайный вектор $U\xi$ имеет такое же распределение, в частности его компоненты — независимые стандартные гауссовские случайные величины.

Помимо самого гауссовского распределения, в приложениях часто используются квадраты гауссовских случайных величин и квадратичные формы от них, в частности, важную роль играет величина

$$\chi_n^2 := \xi_1^2 + \cdots + \xi_n^2,$$

где $\xi_1, \ldots, \xi_n$ — независимые стандартные гауссовские случайные величины. Распределение χ_n^2 называют *распределением хи-квадрат с n степенями свободы*. Это распределение есть n-кратная свертка распределения ξ_1^2, которое можно выразить через стандартное гауссовское распределение заменой.

Еще одно очень важное в статистике распределение, называемое *распределением Стьюдента с $n-1$ степенью свободы*, задается как распределение случайной величины

$$t_{n-1} := \frac{\xi_n}{\sqrt{\frac{1}{n-1}\chi_{n-1}^2}}.$$

Поскольку ξ_n и χ_n независимы, то также можно найти явное выражение для распределения t_{n-1}. Роль статистики Стьюдента связана со следующим любопытным фактом: для независимых стандартных гауссовских случайных величин $\xi_1, \ldots, \xi_n$ рассмотрим величины

$$\overline{\xi} = \frac{1}{n}\sum_{i=1}^{n}\xi_i, \quad S^2 = \frac{1}{n}\sum_{i=1}^{n}(\xi_i - \overline{\xi})^2,$$

которые можно использовать для оценки среднего и дисперсии по наблюдениям. Оказывается, что величины $\overline{\xi}$ и S^2 независимы, причем первая имеет гауссовское распределение с нулевым средним и дисперсией $1/n$, а вторая имеет такое же распределение, как χ_{n-1}^2/n (фактически распределение Стьюдента и возникает как отношение первого и корня из второго с некоторыми коэффициентами). Утверждение о независимости довольно неожиданно, ведь $\overline{\xi}$ явно входит в выражение для S^2. Для доказательства рассмотрим матрицу U размера $n \times n$, первая строка которой состоит из чисел $1/\sqrt{n}$, а остальные строки подобраны так, чтобы сделать U ортогональной (для этого в этих строках запишем координаты элементов ортогонального базиса в подпространстве, ортогональном вектору первой строки). Как отмечено выше, вектор-столбец $\eta = U\xi$, где $\xi = (\xi_1, \ldots, \xi_n)$, имеет также стандартное гауссовское распределение в $\mathbb{R}^n$, поэтому его компоненты η_i — независимые стандартные гауссовские случайные величины. При этом $\overline{\xi} = \eta_1/\sqrt{n}$, а ортогональность U вместе с определение $\overline{\xi}$ дают равенство

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(\xi_i - \overline{\xi})^2 = \frac{1}{n}\sum_{i=1}^{n}\xi_i^2 - \overline{\xi}^2 = \frac{1}{n}\sum_{i=1}^{n}\eta_i^2 - \frac{1}{n}\eta_1^2 = \frac{1}{n}\sum_{i=2}^{n}\eta_i^2,$$

так что остается воспользоваться независимостью η_1 и $\eta_2, \ldots, \eta_n$.

Обратим внимание на тот (тривиальный, но неожиданный) факт, что для получения из суммы независимых стандартных гауссовских случайных величин такой же величины делить ее надо вовсе не на n, а на $\sqrt{n}$. При делении на n получаем величины, стремящиеся к нулю.

В заключение этой главы приведем без доказательства еще одну предельную теорему, важную для прикладной статистики (применяемую в известном критерии Пирсона). В этой теореме не используются средние и дисперсии. Пусть $\{\xi_n\}$ — последовательность независимых случайных величин с общим распределением P_1. Зафиксируем $k \geqslant 2$ и разобьем прямую на k дизъюнктных борелевских множеств $B_1, \ldots, B_k$. Обычно берут два разнонаправленных луча и делят отрезок между ними на конечное число промежутков. Для каждого $i \leqslant k$ обозначим через $\nu_i^n(\omega)$ количество точек в $\xi_1(\omega), \ldots, \xi_n(\omega)$, попавших в B_i. Положим

$$\pi_n(\omega) := \sum_{i=1}^{k} \frac{(\nu_i^n(\omega) - np_i)^2}{np_i}, \quad p_i = P_1(B_i).$$

2.4.3. Теорема. *Распределения случайных величин π_n слабо сходятся к распределению хи-квадрат с $k-1$ степенью свободы (т. е. χ^2_{k-1}).*

Практические применения аналогичны указанным выше: в предположении, что известна теоретическая функция распределения, при небольшом значении k делят прямую на k промежутков и по эмпирическим данным $x_1, \ldots, x_n$ составляют статистику π_n, используя в качестве ν_i^n количество точек из выборки $x_1, \ldots, x_n$, попавших в i-й промежуток. Затем сравнивают полученное значение с числом R, найденным из уравнения $\Phi_{k-1}(R) = 1 - \alpha$ при заданном малом уровне α, где Φ_{k-1} — функция распределения χ^2_{k-1}.

При рассмотрении случайных величин без математических ожиданий возникают новые явления и в поведении сумм: здесь уже не работают закон больших чисел и центральная предельная теорема, но могут существовать способы нормировки $(\xi_1 + \cdots + \xi_n)/A_n$, дающие нетривиальную сходимость. Вот простой пример: распределение Коши не имеет математического ожидания, но для n независимых копий случайной величины с таким распределением величина $(\xi_1 + \cdots + \xi_n)/n$ имеет то же самое распределение (так что очевидным образом есть слабая сходимость)! В самом деле, имеем $\varphi_{\xi_1}(y) = e^{-|y|}$, откуда

$$\varphi_{(\xi_1+\cdots+\xi_n)/n}(y) = \varphi_{\xi_1+\cdots+\xi_n}(y/n) = \big(\varphi_{\xi_1}(y/n)\big)^n = e^{-|y|}.$$

Есть интересные классы распределений, для которых слабая сходимость к нетривиальным пределам появляется при нормировке $1/n^{\alpha}$, где $\alpha \in (0,2)$.

§ 2.5. Задачи

2.5.1. Доказать, что случайные величины $\xi_1, \dots, \xi_n$ независимы в точности тогда, когда преобразование Фурье случайного вектора с компонентами ξ_i имеет вид

$$\varphi_{\xi_1,\dots,\xi_n}(y_1, \dots, y_n) = \varphi_{\xi_1}(y_1) \dots \varphi_{\xi_n}(y_n).$$

Для этого воспользоваться равенством мер на $\mathbb{R}^n$ с равными преобразованиями Фурье.

2.5.2. Построить пример такой последовательности компактных множеств $B_n \subset [0,1]$ с мерой Лебега не менее некоторого числа $\varepsilon > 0$, что для всякой бесконечной подпоследовательности $\{n_k\}$ пересечение $\bigcap_k^\infty B_{n_k}$ имеет меру нуль.

2.5.3. Показать, что если $\xi_n \to \xi$ по вероятности и $\mathrm{M}|\xi_n|^4 \leqslant C$ при всех n, то $\mathrm{M}|\xi_n - \xi|^2 \to 0$.

2.5.4. Показать, что характеристический функционал распределения Коши равен $e^{-|y|}$.

2.5.5. Найти характеристический функционал равномерного распределения в $[a, b]$.

2.5.6. Найти характеристический функционал распределения Бернулли с параметром p.

2.5.7. Найти характеристический функционал распределения Пуассона с параметром λ.

2.5.8. Найти характеристический функционал экспоненциального распределения с параметром λ.

2.5.9. Пусть $\{\xi_n\}$ — последовательность независимых случайных величин и $\mathcal{E}_n$ — σ-алгебра, порожденная $\xi_{n+1}, \xi_{n+2}, \dots$. Доказать, что для всякого множества E из хвостовой σ-алгебры $\mathcal{E} = \bigcap_n \mathcal{E}_n$ справедлив закон $0-1$: $P(E) = 0$ или $P(E) = 1$. Для этого показать, что при фиксированном n для всякого множества A из σ-алгебры, порожденной $\xi_1, \dots, \xi_n$, верно равенство

$$P(E \cap A) = P(E)P(A)$$

и вывести из этого, что $P(E) = P(E)^2$.

2.5.10. Вывести из предыдущей задачи, что если $\{\xi_n\}$ — последовательность независимых случайных величин, то множество тех ω, для которых ряд $\sum_{n=1}^{\infty} \xi_n(\omega)$ сходится, имеет вероятность 0 или 1. Доказать аналогичные утверждения для множеств точек сходимости последовательностей $(\xi_1 + \cdots + \xi_n)/n$ и $(\xi_1 + \cdots + \xi_n)/\sqrt{n}$.

Глава 3

Некоторые приложения

§ 3.1. Условные математические ожидания

Для дискретных вероятностных пространств нам уже встречалось понятие условного распределения, а с ним естественно связывается и понятие условного математического ожидания. Здесь мы обсудим несколько менее тривиальные непрерывные аналоги этих двух очень полезных объектов.

Пусть $(\Omega, \mathcal{B}, P)$ — вероятностное пространство и $\mathcal{A}$ — некоторая под-σ-алгебра в $\mathcal{B}$. Для интегрируемой случайной величины ξ условным математическим ожиданием ξ относительно σ-алгебры $\mathcal{A}$ называется такая $\mathcal{A}$-измеримая случайная величина $\mathbb{E}^{\mathcal{A}}\xi$, что

$$\mathrm{M}(\xi\eta) = \int_\Omega \xi\eta\,dP = \int_\Omega \mathbb{E}^{\mathcal{A}}\xi\,\eta\,dP \tag{3.1.10}$$

для всех $\mathcal{A}$-измеримых ограниченных случайных величин η.

Ясно, что это тождество равносильно тому, что

$$\int_A \xi\,dP = \int_A \mathbb{E}^{\mathcal{A}}\xi\,dP \quad \forall\, A \in \mathcal{A}.$$

В самом деле, последнее тождество сразу дает (3.1.10) для простых функций, которыми затем можно равномерно приближать всякие ограниченные $\mathcal{A}$-измеримые.

Обоснование существования $\mathbb{E}^{\mathcal{A}}\xi$ несложно провести с помощью теоремы Радона – Никодима, но мы обсудим более наглядный способ, применимый в случае $\xi \in L^2(P)$. В этом случае можно рассмотреть в гильбертовом пространстве $L^2(P)$ линейное подпространство $H^{\mathcal{A}}$, состоящее из всех таких функций ζ, для которых есть $\mathcal{A}$-измеримая функция, почти всюду равная ζ. Нетрудно заметить, что $H^{\mathcal{A}}$ действительно линейно и замкнуто: если $\zeta_n \in H^{\mathcal{A}}$ и $\zeta_n \to \zeta$ в $L^2(P)$, то можно выделить почти всюду сходящуюся подпоследовательность в $\{\zeta_n\}$, причем можно при этом считать, что выбраны $\mathcal{A}$-измеримые представители классов эквивалентности (ведь L^2 состоит из классов эквивалентности, а не из индивидуальных функций). Множество Ω_0 меры 1, на котором последовательность $\zeta_n(\omega)$ сходится, входит в $\mathcal{A}$, а функция $\lim\limits_{n\to\infty} \eta_n(\omega)$ на нем $\mathcal{A}$-измерима, следовательно, она и есть представитель класса из $H^{\mathcal{A}}$, к которому сходится $\{\zeta_n\}$. Следовательно, можно взять ортогональную проекцию ξ на $H^{\mathcal{A}}$; $\mathcal{A}$-измеримый представитель этой проекции и служит условным математическим ожиданием.

Из сказанного ясно, что ξ и $\mathbb{E}^{\mathcal{A}}\xi$ имеют равные математические ожидания (берем $\eta = 1$), а в случае $\xi \in L^2(P)$ получаем из свойств проекций, что

$$\|\mathbb{E}^{\mathcal{A}}\xi\|_{L^2}^2 \leqslant \|\xi\|_{L^2}^2,$$

причем равенство имеет место в точности для элементов из $H^{\mathcal{A}}$. Таким образом, условное математическое ожидание дает наилучшее среднеквадратическое приближение $\mathcal{A}$-измеримыми функциями.

3.1.1. Пример. Если наше вероятностное пространство — квадрат $[0,1]^2$ с борелевской σ-алгеброй и мерой Лебега, а $\mathcal{A}$ — σ-алгебра «полос» вида $B \times [0,1]$, где B — борелевское множество в $[0,1]$, то условное математическое ожидание функции f двух переменных есть ее наилучшее приближение в L^2 функциями одного переменного x. Вычислим это приближение. Для каждой ограниченной борелевской функции $h(x)$ мы имеем

$$\int_{[0,1]^2} f(x,y)h(x)\,dx\,dy = \int_0^1 \biggl(\int_0^1 f(x,y)\,dy\biggr)h(x)\,dx,$$

что показывает, что искомая функция есть

$$g(x) = \int_0^1 f(x,y)\,dy.$$

По теореме Фубини эта функция входит в $L^2[0,1]$, если $f \in L^2([0,1]^2)$. Если f лишь интегрируема, то g также интегрируема.

Отметим, что в подобных вычислениях нередко угадывают ответ, а затем проверяют его правильность.

3.1.2. Пример. Если ξ и η — независимые случайные величины, причем σ-алгебра $\mathcal{A}$ порождена случайной величиной η, то $\mathbb{E}^{\mathcal{A}}\xi = \mathrm{M}\xi$ есть постоянная. В самом деле, как проверено ниже, всякая ограниченная $\mathcal{A}$-измеримая функция имеет вид $f(\eta)$ с некоторой ограниченной борелевской функцией f. Как мы знаем, ξ и $f(\eta)$ тоже независимы, поэтому интеграл от $\xi f(\eta)$ равен произведению интегралов от ξ и $f(\eta)$.

3.1.3. Пример. Если ξ сама $\boldsymbol{\mathcal{A}}$-измерима, то $\mathbb{E}^{\boldsymbol{\mathcal{A}}}\xi = \xi$, а если же $\xi = \xi_1\xi_2$, где ξ_1 ограничена и $\boldsymbol{\mathcal{A}}$-измерима, а ξ_2 входит в L^2, то

$$\mathbb{E}^{\boldsymbol{\mathcal{A}}}\xi = \xi_1\mathbb{E}^{\boldsymbol{\mathcal{A}}}\xi_2.$$

Первое равенство очевидно из определения, а для проверки второго замечаем, что функция $\xi_1\mathbb{E}^{\boldsymbol{\mathcal{A}}}\xi$ является $\boldsymbol{\mathcal{A}}$-измеримой и входит в L^2. При этом для всякой ограниченной $\boldsymbol{\mathcal{A}}$-измеримой функции η интеграл от $\eta\xi_1\mathbb{E}^{\boldsymbol{\mathcal{A}}}\xi$ равен интегралу от $\eta\xi_1\xi_2 = \eta\xi$, ибо функция $\eta\xi_1$ ограничена и $\boldsymbol{\mathcal{A}}$-измерима.

Покажем, что если $\xi_1,\ldots,\xi_n$ — независимые случайные величины, $\boldsymbol{\mathcal{A}}_{n-1}$ — наименьшая σ-алгебра, относительно которой измеримы $\xi_1,\ldots,\xi_{n-1}$, то условное математическое ожидание случайной величины $f(\xi_1,\ldots,\xi_n)$ для ограниченной борелевской функции f задается формулой

$$\mathbb{E}^{\boldsymbol{\mathcal{A}}_{n-1}}f(\omega) = \int_{\mathbb{R}} f\bigl(\xi_1(\omega),\ldots,\xi_{n-1}(\omega),u\bigr)\,P_{\xi_n}(du).$$

Заметим, что множества из σ-алгебра $\boldsymbol{\mathcal{A}}_n$, порожденной случайными величинами $\xi_1,\ldots,\xi_n$, т. е. наименьшей σ-алгебры, относительной которой они измеримы, есть в точности множества A, для которых их индикаторы I_A имеют вид $I_A = \psi(\xi_1,\ldots,\xi_n)$, где ψ — борелевская функция; их можно описать также формулой $A = (\xi_1,\ldots,\xi_n)^{-1}(B)$, где B — борелевское множество в $\mathbb{R}^n$. В самом деле, класс множеств такого вида есть σ-алгебра, причем они входят в $\boldsymbol{\mathcal{A}}_n$.

Теперь для проверки указанного выше равества заметим, что для каждого множества $A \in \boldsymbol{\mathcal{A}}_{n-1}$ интеграл правой части по A есть

$$\int_{\mathbb{R}^n} I_B(x_1,\ldots,x_{n-1})f(x_1,\ldots,x_{n-1},u)\,P_{\xi_1}(dx_1)\cdots P_{\xi_{n-1}}(dx_{n-1})P_{\xi_n}(du),$$

где $A = (\xi_1,\ldots,\xi_{n-1})^{-1}(B)$, $B \in \mathcal{B}(\mathbb{R}^{n-1})$. Этой же величине равен и интеграл от левой части, по определению совпадающий с интегралом от $I_A f(\xi_1,\ldots,\xi_n) = I_B(\xi_1,\ldots,\xi_n)f(\xi_1,\ldots,\xi_n)$. При этом правая часть доказываемого равенства измерима относительно $\boldsymbol{\mathcal{A}}_{n-1}$.

В рассматриваемом круге вопросов полезно следующее утверждение о структуре измеримых функций.

3.1.4. Теорема. (i) *Пусть $\mathcal{A}_n$ — σ-алгебра, порожденная случайными величинами $\xi_1, \ldots, \xi_n$. Тогда $\mathcal{A}_n$-измеримые функции имеют вид*

$$f(\omega) = g\bigl(\xi_1(\omega), \ldots, \xi_n(\omega)\bigr), \tag{3.1.11}$$

где $g\colon \mathbb{R}^n \to \mathbb{R}$ — борелевская функция.

(ii) *Пусть $\mathcal{A}$ — σ-алгебра, порожденная бесконечным набором случайных величин ξ_t. Тогда для всякой $\mathcal{A}$-измеримой функции f найдутся такие счетное множество индексов t_n и функция $g\colon \mathbb{R}^\infty \to \mathbb{R}$ на счетной степени прямой, измеримая относительно σ-алгебры $\mathcal{B}(\mathbb{R}^\infty)$, порожденной координатными функциями, что*

$$f(\omega) = g\bigl(\xi_1(\omega), \ldots, \xi_n(\omega), \ldots\bigr).$$

Доказательство. (i) Как указано выше, множества из $\mathcal{A}_n$ имеют вид $A = (\xi_1, \ldots, \xi_n)^{-1}(B)$, $B \in \mathcal{B}(\mathbb{R}^n)$. Поэтому

$$I_A(\omega) = I_B\bigl(\xi_1(\omega), \ldots, \xi_n(\omega)\bigr).$$

Значит, утверждение (i) верно для индикаторов, а тогда и для f с конечным множеством значений. Общий случай достаточно проверить для ограниченных f (перейдя к $\operatorname{arctg} f$). Найдутся $\mathcal{A}_n$-измеримые простые функции f_j, равномерно сходящиеся к f. По доказанному, имеем $f_j = g_j(\xi_1, \ldots, \xi_n)$, где g_j — ограниченные борелевские функции на $\mathbb{R}^n$. Рассмотрим множество

$$S = \{x \in \mathbb{R}^n \colon \{g_j(x)\} \text{ имеет конченый предел}\}.$$

Нетрудно проверить, что множество S всегда борелевское, причем функция $g(x) = \lim\limits_{j\to\infty} g_j(x)$ на нем измерима по Борелю. Доопределив ее вне S нулем, получим борелевскую функцию g на всем $\mathbb{R}^n$. При этом верно равенство (3.1.11), ибо в силу сходимости $f_j \to f$ мы имеем $\bigl(\xi_1(\omega), \ldots, \xi_n(\omega)\bigr) \in S$ для всех ω.

(ii) Множества из $\mathcal{A}$ имеют вид

$$I_A(\omega) = I_B\bigl(\xi_1(\omega), \ldots, \xi_n(\omega), \ldots,\bigr), \quad B \in \mathcal{B}(\mathbb{R}^\infty),$$

ибо класс таких множеств — σ-алгебра, функции ξ_n измеримы относительно нее, причем множество в правой части входит во всякую σ-алгебру $\mathcal{E}$, относительно которой измеримы все ξ_n. Для доказательства последнего утверждения надо лишь заметить, что класс всех множеств $B \in \mathcal{B}(\mathbb{R}^\infty)$, для которых это верно, сам является σ-алгеброй и очевидным образом содержит множества вида $\{x \in \mathbb{R}^\infty \colon x_n < c\}$ (значит, по определению содержит и $\mathcal{B}(\mathbb{R}^\infty)$ — наименьшую σ-алгебру, содержащую множества $\{x \in \mathbb{R}^\infty \colon x_n < c\}$). Теперь можно применить то же рассуждение, что и в (i). $\square$

Если случайная величина f зависит от независимых случайных векторов (групп случайных величин) ξ и η, то при наличии разложения

$$f(\xi,\eta)=\sum_k f_k(\xi)g_k(\eta)$$

(скажем, сходящегося в L^2) условное математическое ожидание $f(\xi,\eta)$ относительно σ-алгебры $\mathcal{A}$, порожденной η, находится по формуле

$$\mathbb{E}^{\mathcal{A}}f=\sum_k g_k(\eta)\mathrm{M}f_k(\xi).$$

В приложениях часто встречается ситуация, когда задана возрастающая последовательность σ-алгебр $\mathcal{A}_n\subset\mathcal{B}$, причем $\mathcal{B}$ — наименьшая содержащая их σ-алгебра. Примером служат конечные σ-алгебры $\mathcal{A}_n\subset 2^{\mathbb{N}}$, где $\mathcal{A}_n$ состоит из всех подмножеств набора $\{1,\ldots,n\}$ и их дополнений.

Последовательность интегрируемых случайных величин ξ_n называется *мартингалом* относительно последовательности $\mathcal{A}_n$, если ξ_n измерима относительно $\mathcal{A}_n$ и

$$\mathbb{E}^{\mathcal{A}_n}\xi_{n+1}=\xi_n\quad\forall\,n.$$

В случае функций из L^2 это означает, что ξ_n с меньшим номером получается проекцией на $H^{\mathcal{A}_n}$ функций с большими номерами. Поэтому простым (но весьма типичным) примером мартингала является такая последовательность: берем произвольную $\mathcal{B}$-измеримую функцию $\xi\in L^2(P)$ и образуем ее ортогональные проекции $\mathbb{E}^{\mathcal{A}_n}\xi$ на $H^{\mathcal{A}_n}$.

Имеет место важная теорема Дуба о сходимости мартингалов.

3.1.5. Теорема. *Случайные величины $\mathbb{E}^{\mathcal{A}_n}\xi$ сходятся к ξ в $L^2(P)$ и почти наверное, причем для всякого мартингала $\{\xi_n\}$, ограниченного по норме в $L^2(P)$, т. е. с $\sup_n\|\xi_n\|_{L^2(P)}<\infty$, существует такая случайная величина $\xi\in L^2(P)$, что $\xi_n=\mathbb{E}^{\mathcal{A}_n}\xi$ для всех n.*

Таким образом, для ограниченных мартингалов в $L^2(P)$ указанный простой пример является универсальным.

Вот другой классический пример: пусть $S_n=\xi_1+\cdots+\xi_n$, где ξ_i — независимые интегрируемые случайные величины с нулевым математическим ожиданием. Пусть $\mathcal{A}_n$ — σ-алгебра, порожденная $\xi_1,\ldots,\xi_n$. Как показано выше, функции, измеримые относительно $\mathcal{A}_n$, имеют вид $f(\xi_1,\ldots,\xi_n)$, где f — борелевская функция. Тогда $\{S_n\}$ — мартингал относительно $\{\mathcal{A}_n\}$.

Действительно, функция S_n измерима относительно $\mathcal{A}_n$, а разность $\xi_{n+1} = S_{n+1} - S_n$ независима со всякой функцией от $\xi_1, \ldots, \xi_n$, поэтому ортогональная всякой такой функции.

Рассмотрим особо простой (но полезный) случай, когда $\mathcal{A}$ порождается конечным разбиением Ω на дизъюнктные части $A_1, \ldots, A_k \in \mathcal{B}$. В этом случае σ-алгебра $\mathcal{A}$ конечна и состоит из всевозможных объединений множеств A_i и пустого множества. Если $P(A_i) > 0$ при всех i, то

$$\mathbb{E}^{\mathcal{A}}\xi(\omega) = \sum_{i=1}^{n} P(A_i)^{-1} \int_{A_i} \xi \, dP \cdot I_{A_i}(\omega).$$

В самом деле, здесь $\mathcal{A}$-измеримые функции — линейные комбинации индикаторов I_{A_i}. Ясно, что при каждом i функция $I_{A_i}\xi$ имеет такой же интеграл, как произведение I_{A_i} и указанной суммы (этот интеграл есть интеграл от ξ по A_i).

В рассмотренном примере получаем

$$\mathbb{E}^{\mathcal{A}} I_B(\omega) = \sum_{i=1}^{n} P(A_i)^{-1} P(B \cap A_i) \cdot I_{A_i}(\omega)$$

при $B \in \mathcal{B}$. Это можно рассматривать как условную меру $P(B|\mathcal{A})(\omega)$ множества B относительно σ-алгебры $\mathcal{A}$; она в самом деле будет вероятностной мерой по B при фиксированном ω, а интегрирование $P(B|\mathcal{A})(\omega)$ по всякому множеству $A \in \mathcal{A}$ даст $P(A \cap B)$.

Можно попытаться сделать то же самое и в общем случае, положив

$$P(B|\mathcal{A})(\omega) := \mathbb{E}^{\mathcal{A}} I_B(\omega), \quad B \in \mathcal{B}.$$

Однако не всегда можно добиться того, чтобы при фиксированном ω (хотя бы почти всяком) получилась счетно-аддитивная мера по B. Тем не менее при весьма широких условиях (охватывающих все нужные в приложениях случаи) такой выбор счетно-аддитивной по B при почти каждом ω версии возможен.

§ 3.2. Случайные блуждания

В двух важнейших предельных теоремах, рассмотренных выше, речь шла о специальным образом нормированных суммах независимых случайных величин ξ_n. Значительный интерес представляет собой асимптотическое поведение и самих этих сумм

$$S_n = \xi_1 + \cdots + \xi_n$$

без всякой нормировки. Выше мы видели, что для случайных величин с нулевыми математическими ожиданиями $\{S_n\}$ — мартингал.

Известно, что если ξ_n независимы, имеют одно и то же распределение, причем $\mathrm{M}\xi_n = 0$, $\mathrm{M}\xi_n = \sigma^2$, то почти наверное

$$\limsup_{n\to\infty} \frac{S_n}{\sqrt{2n\ln\ln n}} = \sigma, \quad \liminf_{n\to\infty} \frac{S_n}{\sqrt{2n\ln\ln n}} = -\sigma,$$

причем весь отрезок $[-\sigma, \sigma]$ состоит из предельных точек последовательности $S_n(\omega)/\sqrt{2n\ln\ln n}$ почти наверное. Итак, типичные суммы приближенно флуктуируют от $-\sigma\sqrt{2n\ln\ln n}$ до $\sigma\sqrt{2n\ln\ln n}$.

В случае случайных величин со значениями во множестве целых чисел возникает интересная задача о возвратности: последовательность сумм S_n называется возвратной, если

$$P(S_n = 0 \text{ для бесконечно многих } n) = 1.$$

Оказывается, что если ξ_n независимы и имеют общее распределение с нулевым математическим ожиданием, то последовательность $\{S_n\}$ возвратна. Например, так будет, если ξ_n принимает значения -1 и 1 с вероятностью 1. Такой процесс имеет интересный многомерный аналог: случайное блуждание по целочисленной решетке в $\mathbb{R}^d$ со стандартным базисом $e_1, \ldots, e_d$, когда случайный вектор ξ_n с вероятностью $1/(2d)$ принимает значения $e_1, \ldots, e_d$ и им противоположные. Например, на плоскости с вероятностью $1/4$ происходит скачок на 1 по горизонтали или по вертикали. Такое случайное блуждание оказывается возвратным на плоскости и невозвратным при $d > 2$. Более сложным случайным процессом оказывается блуждание по решетке с возможностью перескока не только в соседние вершины, но и в любые другие.

§ 3.3. Цепи Маркова

В предыдущем разделе появились случайные последовательности с дискретными значениями с некоторыми вероятностями переходов из одного состояния в другое. Здесь мы очень кратко обсудим основные понятия и факты, связанные с такими последовательностями с конечным числом возможных значений. Анализ этого весьма важного для приложений класса процессов был начат более ста лет назад А.А. Марковым, в честь которого естественные обобщения процессов такого рода называются марковскими, а процессы с конечным или счетным числом состояний называются марковскими цепями или цепями Маркова. Мы обсудим еще более специальный случай конечной однородной марковской цепи. Пусть дано конечное множество состояний $S = \{s_1, \ldots, s_m\}$ (их можно считать целыми числами, но это несущественно), причем для каждого i заданы «вероятности перехода» p_{ij} из состояния s_i в состояния s_j, т. е. $p_{ij} \geqslant 0$, $\sum_{j=1}^m p_{ij} = 1$.

Например, при двух состояниях 1 и 2 возможен переход из 1 в 2 с вероятностью p_{12}, можно остаться в 1 с вероятностью $p_{11} = 1 - p_{12}$, а из 2 можно перейти в 1 с вероятностью p_{21} или остаться в 2 с вероятностью $1 - p_{21}$. Скажем, сам А.А. Марков с помощью этой модели анализировал чередование гласных и согласных в словах русского алфавита (после гласной может стоять согласная или опять гласная, после согласной может стоять гласная или снова согласная). Предположим, что задано некоторое начальное распределение π_0 на S, т.е. числа $\pi_0(j) \geqslant 0$ с $\sum_{j=1}^{m} \pi_0(j) = 1$. После первого шага начальное распределение трансформируется. Например, в случае двух состояний в 1 можно быть в результате перехода из 2 или в результате того, что остались на месте после первого шага. Таким, образом вероятности нахождения в 1 и 2 теперь равны соответственно $p_{21}\pi_0(2) + p_{11}\pi_0(1)$ и $p_{12}\pi_0(1) + p_{22}\pi_0(2)$. Поэтому возникает новая для нас ситуация: случайная величина ξ_n со значениями в S возникла из случайной величины ξ_0 после n переходов, причем распределение ξ_0 известно (это π_0), а вот распределение π_n величины ξ_n предстоит вычислить. Это одна из возникающих здесь задач — динамика начального распределения. Распределение π_n вычисляется исключительно по предыдущему распределению π_{n-1} и вероятностям перехода p_{ij}. «Марковость» процесса состоит в том, что будущее зависит от прошлого только через настоящее в том смысле, что для знания распределений ξ_k при $k > n$ знание π_n дает столько же, как знание всех π_l с $l \leqslant n$. Можно также сказать, что будущее и прошлое независимы при фиксированном настоящем (это общая концепия «марковости» случайного процесса). В терминах условных вероятностей: при $k \leqslant n, m \geqslant n$ верны равенства

$$P(\xi_k = i, \xi_m = j | \xi_n = s) = P(\xi_k = i | \xi_n = s) P(\xi_m = j | \xi_n = s).$$

Найти π_{n+1} по π_n и матрице $P = (p_{ij})_{i,j \leqslant m}$, называемой матрицей переходов, легко: вероятность нахождения в s_i в момент $n+1$ есть

$$\sum_{j=1}^{m} p_{ji} \pi_n(j),$$

ибо в s_i можно попасть из одной из точек s_j, в которых процесс мог быть в момент n. Фактически здесь речь идет даже не о самой случайной величине ξ_{n+1}, а лишь о ее распределении, задаваемом указанной выше формулой. Поэтому доказательство существования последовательности с такими распределениями нуждается в обосновании (его можно дать с помощью теоремы Колмогорова). Если рассмотреть вектор-строки с компонентами $\pi_n(1), \dots, \pi_n(m)$ и обозначить их теми же символами π_n, что и вероятности, то получим простую формулу

$$\pi_{n+1} = \pi_n P,$$

означающую действие матрицы $P = (\pi_{ij})$ на вектор-строку π_n справа (вектор-строка умножается на первый столбец, затем на второй и т. д.). Следовательно,

$$\pi_n = \pi_0 P^n.$$

Среди целого ряда интересных задач, возникающих при изучении описанной марковской динамики, упомянем две: существует ли предел π_n и существует ли такое начальное распределение π_0, которое не будет меняться со временем (*стационарное распределение*), т. е. $\pi_0 = \pi_0 P$?

На последний вопрос можно дать положительный ответ, так как множество векторов $\pi \in \mathbb{R}^m$ с неотрицательными компонентами, в сумме дающими 1, является выпуклым компактом, который оператор P переводит в него же, так что можно воспользоваться известной теоремой Боля – Брауэра о неподвижной точке. С точки зрения теории матриц это тоже довольно полезный факт: он означает, что стохастическая матрица (матрица с неотрицательными элементами, сумма которых в каждой строке равна 1) имеет собственное число 1.

Поведение $\pi_0 P^n$ может быть сложным. Если матрица P симметрична, то она имеет собственный ортогональный базис, в котором просто записываются и матрицы P^n. Однако стохастичность не влечет симметричность. Все же некоторая специфика появляется. Если записать P в жордановой форме, то появится некоторое число диагональных элементов и некоторое число нетривиальных жордановых клеток, в которых стоит на диагонали некоторое число λ, над диагональю стоят единицы, остальные элементы нулевые. Такая клетка имеет вид $\lambda \cdot \mathrm{I} + B$, где в матрице B все элементы нулевые, кроме единиц над главной диагональю. При возведении P в степень n диагональные элементы возводятся в степень n, блоки возводятся в степень по правилу

$$(\lambda \cdot \mathrm{I} + B)^n = \sum_k C_n^k \lambda^k B^{n-k},$$

причем $B^{n-k} = 0$ при $n - k \geqslant d$, где d — порядок рассматриваемого блока. Поскольку матрицы P^n остаются стохастическими, то ясно, что не может быть собственных чисел, по модулю больше 1. Кроме того, собственным числам λ с $|\lambda| = 1$ не может отвечать нетривиальная клетка. В противном случае также возникает противоречие со стохастичностью P^n, ибо при $d > 1$ указанная выше сумма для $(\lambda \cdot \mathrm{I} + B)^n$ из d элементов растет по модулю как C_n^{d-1}. Именно собственные числа P с модулем 1 и играют основную роль в поведении P^n при $n \to \infty$.

Обратим внимание, что 1 стоит в жордановой форме матрицы P, но ее может не быть в самой матрице (так заведомо будет, если все элементы P положительны).

Конечная марковская цепь называется *эргодической*, если существует такое $n_0 \geqslant 1$, что элементы матрицы P^{n_0} положительны. Например, это условие выполнено, если элементы самой матрицы P положительны. Важный результат теории марковских цепей состоит в следующем.

3.3.1. Теорема. *Для эргодической цепи Маркова стационарное распределение π единственно, $\pi(s_i) > 0$ для всех i, причем*

$$\lim_{n\to\infty} p_{ij}^{(n)} = \pi(s_j) \quad \text{\textit{для всех} } i, j, \text{ \textit{где} } P^n = (p_{ij}^{(n)}).$$

3.3.2. Пример. Найдем стационарное распределение цепи Маркова с двумя состояниями и матрицей перехода с элементами p, $1-p$, q, $1-q$. Для собственного вектора с компонентами $\pi_1, 1-\pi_1$, отвечающего собственному числу 1, получаем

$$p\pi_1 + q(1-\pi_1) = \pi_1,$$

откуда $\pi_1 = q/(1-p+q)$. Например, при $p = q = 1/2$ это дает $1/2, 1/2$.

Можно было бы иметь дело и с вектор-столбцами вероятностей, но тогда умножать их надо слева на транспонированную матрицу к P.

Аналогично вводятся цепи Маркова со счетным числом состояний, тогда рассмотренные выше случайные блуждания попадают в этот класс. При этом вообще можно определять случайное блуждание по целочисленной решетке в $\mathbb{Z}^d$ как цепь Маркова с пространством состояний $\mathbb{Z}^d$.

§ 3.4. Задачи

3.4.1. Пусть случайные величины ξ и η независимы, причем существует $\mathrm{M}\xi$. Доказать, что $\mathrm{M}\xi$ совпадает с условным математическим ожиданием ξ относительно σ-алгебры, порожденной η.

3.4.2. Привести пример эргодической цепи Маркова, для которой матрица перехода P имеет нулевой элемент.

3.4.3. Привести пример конечной цепи Маркова с неединственным стационарным распределением.

Программа экзамена

1. Конечные вероятностные пространства. События. Независимость событий. Доказательство формулы полной вероятности и формулы Байеса.

2. Общее понятие вероятностного пространства, σ-алгебра событий, борелевская σ-алгебра на прямой, вероятностные меры.

3. Случайные величины и их основные свойства (операции над ними и пределы). Доказательство равносильности двух определений измеримой функции.

4. Произведения вероятностных пространств (конечные и счетные). Независимые случайные величины. Характеризация независимости в терминах совместного распределения.

5. Математическое ожидание (интеграл Лебега) и его основные свойства. Дисперсия. Доказательство формулы замены переменных для образа меры при измеримом отображении.

6. Доказательство неравенства Чебышёва. Сходимость почти наверное и по вероятности. Сходимость в среднем и среднем квадратическом. Связь между разными видами сходимости.

7. Функция распределения случайной величины. Медиана. Плотность распределения. Выражение математического ожидания и дисперсии с помощью плотности распределения (вывод из формулы замены переменных).

8. Основные дискретные и непрерывные распределения: бернуллиевское, пуассоновское, равномерное, гауссовское, экспоненциальное, распределение Коши.

9. Равенство $\mathrm{M}(\xi\eta) = \mathrm{M}\xi\mathrm{M}\eta$ для независимых случайных величин с конечными математическими ожиданиями.

10. Закон больших чисел (условия сходимости по вероятности и почти наверное).

11. Слабая сходимость распределений. Характеризация через сходимость функций распределения.

12. Характеристический функционал (преобразование Фурье) распределения. Описание слабой сходимости в терминах преобразований Фурье.

13. Центральная предельная теорема для одинаково распределенных независимых случайных величин с конечной дисперсией.

14. Условное математическое ожидание (случай квадратично интегрируемых случайных величин). Основные свойства и примеры.

15. Случайные блуждания и цепи Маркова (основные понятия и факты).