

ВВЕДЕНИЕ В ТЕОРИЮ ВЕРОЯТНОСТЕЙ:
КОНСПЕКТ ЛЕКЦИЙ

В.И. БОГАЧЕВ

мат. фак. ВШЭ, 3 курс, осень 2017

Оглавление

Предисловие	5
Глава 1. Вероятностные пространства и случайные величины	9
§ 1.1. Вероятностные пространства	10
§ 1.2. Случайные величины	14
§ 1.3. Математическое ожидание и дисперсия	17
§ 1.4. Сходимость случайных величин	25
§ 1.5. Некоторые классы случайных величин	26
§ 1.6. Независимые случайные величины	27
§ 1.7. Задачи	33
Программа коллоквиума	37

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. Общеизвестное 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 не дискретной 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 практически 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

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

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

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

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 $\mathcal{B}$ вероятностной меры допускала счетные операции объединения, пересечения и дополнения, т. е. была σ -алгеброй (что включает еще естественное требование, что $\Omega, \emptyset \in \mathcal{B}$), а сама вероятностная мера $P: \mathcal{B} \rightarrow [0, 1]$ 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 $B_n \in \mathcal{B}$. Название «вероятностная» включает 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 $\mathcal{A}$ вместо σ -алгебры (в этом случае нужное равенство требуется для таких счетных объединений дизъюнктивных множеств из $\mathcal{A}$, что и их объединение лежит в $\mathcal{A}$). 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 $\sigma(\mathcal{A})$, содержащую $\mathcal{A}$. Эта минимальная σ -алгебра формально допускает очень простое описание: надо пересечь все вообще σ -алгебры множеств отрезка, содержащие $\mathcal{A}$ (такие есть, например, класс всех множеств). Однако конструктивного описания множеств из $\sigma(\mathcal{A})$ 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

$$P^*(B) = \inf \left\{ \sum_{n=1}^{\infty} P(A_n) : A_n \in \mathcal{A}, B \subset \bigcup_{n=1}^{\infty} A_n \right\},$$

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: водится класс множеств $\mathcal{L}$ с тем свойством, что

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

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 не дискретным 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 $f: \Omega \rightarrow \mathbb{R}$ на пространстве Ω , наделенном σ -алгеброй $\mathcal{B}$, называется измеримой относительно $\mathcal{B}$ (или $\mathcal{B}$ -измеримой), если множество $\{\omega: f(\omega) < c\}$ 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 $\mathcal{B}_1 = \mathcal{B}(\mathbb{R})$. $\square$

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 $\{\omega : f(\omega) < c\}$ есть объединение по натуральным k множеств $\Omega_k = \bigcup_{N=1}^{\infty} \bigcap_{n \geq N} \{\omega : f_n(\omega) < c - 1/k\}$, 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.). $\square$

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

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

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

1.2.4. Определение. *Функция распределения с.в. ξ задается равенством*

$$F_\xi(t) = P(\omega: \xi(\omega) < t), \quad t \in \mathbb{R}.$$

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 $t \leq 0$ и $F_\xi(t) = 1$ при $t > 0$. Если ξ принимает конечное число разных значений $c_1, \dots, c_n$, 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 $\inf\{s: F_\xi(s) \geq 1/2\}$.

Отметим, что медиану бывает полезно вводить и для конечных наборов чисел $c_1, \dots, c_n$, необязательно связанных с вероятностными распределениями. Для этого такой набор упорядочивают по возрастанию и получают числа $b_1 \leq \dots \leq b_n$ (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

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

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 $\xi_n(\omega) \rightarrow \xi(\omega)$ 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

$$\lim_{n \rightarrow \infty} P(\omega: |\xi_n(\omega) - \xi(\omega)| \geq \varepsilon) = 0.$$

Отметим, что при сходимости по вероятности возможна ситуация, когда $P(\omega: |\xi_n(\omega) - \xi(\omega)| > 0) = 1$, 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 $c_1 I_{A_1} + \dots + c_n I_{A_n}$ положим

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

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

Если $\xi \geq 0$, то положим

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

где S_{ξ} — множество всех простых случайных величин η , для которых $\eta \leq \xi$ 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

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

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

$$M\xi := M\xi^+ - M\xi^-.$$

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

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

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

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 являются функции, равные почти всюду. Тогда $L^1(P)$ оказывается не только линейным пространством, но еще и нормированным, если в качестве нормы взять $M|\xi|$. 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 $\xi \leq \eta$ п.н. и обе с.в. интегрируемы, то $M\xi \leq M\eta$. Наконец, если с.в. ξ и η таковы, что $|\xi| \leq |\eta|$ 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

$$D\xi := M(\xi - M\xi)^2 = M\xi^2 - (M\xi)^2$$

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 $M|\xi|^k$).

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

$$P(\omega: |\xi(\omega)| \geq R) \leq \frac{M|\xi|}{R}.$$

ДОКАЗАТЕЛЬСТВО. Пусть Ω_R — множество в левой части этого неравенства. Тогда $R I_{\Omega_R} \leq |\xi|$. $\square$

1.3.3. Следствие. Для всякой случайной величины ξ с конечной дисперсией имеем

$$P(|\xi - M\xi| \geq R) \leq \frac{D\xi}{R^2}.$$

Из этой оценки следует, что $D\xi = 0$ 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

$$P(|\xi - M\xi| \geq 3\sigma) \leq \frac{D\xi}{9\sigma^2} = \frac{\sigma^2}{9\sigma^2} = \frac{1}{9}.$$

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

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

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

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

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

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

1.3.6. Теорема. (ТЕОРЕМА ЛЕБЕГА О МАЖОРИРУЕМОЙ СХОДИМОСТИ) Пусть существует $\Phi \in \mathcal{L}^1(P)$ с

$$|\xi_n(\omega)| \leq \Phi(\omega) \quad \text{п.н.}$$

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

$$M\xi = \lim_{n \rightarrow \infty} M\xi_n. \quad (1.3.1)$$

Кроме того, $\lim_{n \rightarrow \infty} M|\xi - \xi_n| = 0$.

1.3.7. Теорема. (ТЕОРЕМА БЕППО ЛЕВИ О МОНОТОННОЙ СХОДИМОСТИ) Пусть

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

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

$$\xi_n \geq 0 \quad \text{и} \quad \sup_n M\xi_n < \infty.$$

Тогда $\xi \in \mathcal{L}^1(P)$ и

$$M\xi \leq \liminf_{n \rightarrow \infty} M\xi_n.$$

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:

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

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

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

Все это можно проверить по определению, но проще воспользоваться теоремами о сходимости: считая, что $\xi \geq 0$ (к этому по определению сводится общий случай), вводим случайные величины ξ_n , совпадающие с ξ на $\{\omega_1, \dots, \omega_n\}$ 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

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

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

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

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

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

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 дизъюнктивных интервалов (a_i, b_i) суммарной длины не более δ верно неравенство $\sum_{i=1}^n |F_\xi(b_i) - F_\xi(a_i)| < \varepsilon$. 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 $(\Omega, \mathfrak{B}, P)$. Тогда для каждой ограниченной борелевской функции f верно равенство

$$Mf(\xi) = \int_{\Omega} f(\xi(\omega)) P(d\omega) = \int_{\mathbb{R}} f(t) P_{\xi}(dt). \quad (1.3.2)$$

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 $B \in \mathfrak{B}(\mathbb{R})$, 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 $|f| \leq M$, то для заданного $\varepsilon > 0$ можно разделить $[-M, M]$ на конечное число промежутков $J_1, \dots, J_n$ 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 $|f(x) - g(x)| \leq \varepsilon$ при всех x , ибо это верно на каждом из множеств $f^{-1}(J_k)$, на которые разбивается все пространство.

Если $f \geq 0$ 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 $N \rightarrow \infty$ 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. $\square$

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

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

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 $\xi: \mathbb{R}^d \rightarrow \mathbb{R}^d$ 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

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

что при наличии плотности равно

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

Аналогично для дисперсии:

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

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

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

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

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

$$Mf(\xi) = \int_{\mathbb{R}^d} f(x_1, \dots, x_d) P_\xi(dx)$$

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

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

§ 1.4. Сходимость случайных величин

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

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

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

убывают, а их пересечение имеет меру нуль по условию. Значит, мы имеем $P(A_n) \rightarrow 0$, 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 $f_{3,1}, \dots, f_{3,4}$ — 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

$$M|\xi_n - \xi| \rightarrow 0.$$

Из него следует сходимость по мере в силу неравенства Чебышёва:

$$P(|\xi - \xi_n| \geq \varepsilon) \leq \varepsilon^{-1} M|\xi - \xi_n|, \quad \varepsilon > 0.$$

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 $M|\xi_n - \xi|^p \rightarrow 0$. 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 $k = 0, 1, 2, \dots$ 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 $\sigma \geq 0$ задается так. Если $\sigma = 0$, то это мера Дирака в точке a . Если $\sigma > 0$, то это мера с плотностью

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

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 ,

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

где $\alpha > 1$ — некоторый параметр.

Ясно, что распределение Коши и распределение Парето при $\alpha \leq 2$ 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 $\xi: (0, 1) \rightarrow \mathbb{R}$, 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 $\xi_1, \dots, \xi_n$ на общем вероятностном пространстве $(\Omega, \mathcal{B}, P)$ называют независимыми, если*

$$\begin{aligned} P(\omega: \xi_1(\omega) < c_1, \dots, \xi_n(\omega) < c_n) = \\ = P(\omega: \xi_1(\omega) < c_1) \cdots P(\omega: \xi_n(\omega) < c_n) \quad \forall c_1, \dots, c_n \in \mathbb{R}. \end{aligned}$$

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 $(\xi_1, \dots, \xi_n)$ 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

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

Верна также симметричная формула

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

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

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

Частью теоремы Фубини является утверждение о том, что внутренние интегралы по одному переменному автоматически существуют 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 $\mathcal{B}_1 \otimes \cdots \otimes \mathcal{B}_n$, 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 $\{x: x_1 < c_1, \dots, x_n < c_n\}$.

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

$$\begin{aligned} P(\omega: \xi_1(\omega) \in B_1, \dots, \xi_n(\omega) \in B_n) = \\ = P(\omega: \xi_1(\omega) \in B_1) \cdots P(\omega: \xi_n(\omega) \in B_n) \quad \forall B_1, \dots, B_n \in \mathcal{B}(\mathbb{R}). \end{aligned}$$

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 $0 < P(\eta^{-1}(B_2)) < 1$; его существование как раз и означает, что $f(\xi)$ имеет отличное от дираковского распределение. Тогда число $P(\eta^{-1}(B_2))$ отлично от своего квадрата, поэтому

$$\begin{aligned} P(\omega: \xi(\omega) \in f^{-1}(B_2), \eta(\omega) \in B_2) \neq \\ \neq P(\omega: \xi(\omega) \in f^{-1}(B_2)) P(\omega: \eta(\omega) \in B_2), \end{aligned}$$

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 $\eta_1, \dots, \eta_n$ 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*

$$M(\xi\eta) = M\xi M\eta.$$

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 $M(\xi\eta)$ 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 $M\xi$ и $M\eta$. $\square$

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

$$M(\xi_1 \cdots \xi_n) = M\xi_1 \cdots M\xi_n.$$

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 $\xi_1, \dots, \xi_n$ имеют конечные дисперсии, то такова и их сумма, причем*

$$D(\xi_1 + \cdots + \xi_n) = D\xi_1 + \cdots + D\xi_n,$$

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

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

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

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

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

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

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

$$P(A_{i_1} \cap \dots \cap A_{i_k}) = P(A_{i_1}) \dots P(A_{i_k}) \quad (1.6.3)$$

для каждого поднабора $\{i_1, \dots, i_k\} \subset \{1, \dots, n\}$ 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 $\{\omega: I_{A_1}(\omega) \in B_1, \dots, I_{A_n}(\omega) \in B_n\}$ 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

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

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 дизъюнктивностью, которая в нетривиальных случаях как раз несовместима с независимостью (при $A \cap B = \emptyset$ имеем $P(A \cap B) = 0$, что не равно $P(A)P(B)$ 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 $I_1, \dots, I_4$ 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 $P(\bigcup_{n=k}^{\infty} B_n) = 1$. Для всякого $N \geq k$ имеем

$$1 - P\left(\bigcup_{n=k}^{\infty} B_n\right) \leq 1 - P\left(\bigcup_{n=k}^N B_n\right) = P\left(\bigcap_{n=k}^N (\Omega \setminus B_n)\right) = \prod_{n=k}^N (1 - P(B_n)),$$

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

Разумеется, здесь важна независимость B_n . Для произвольных событий B_n при условии $P(B_n) \geq c > 0$ 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 $1, \dots, 100$ 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 $1, \dots, N$.

§ 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 лотереи, а две трети на акции ЕЕЕ «Наколково». Какова вероятность полной потери вложенных средств, если в лотерее 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 величины ξ и η имеют конечные дисперсии $D\xi$ и $D\eta$, причем $\xi \leq \eta$. Верно ли, что $D\xi \leq D\eta$?

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

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

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

Программа коллоквиума

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 $M(\xi\eta) = M\xi M\eta$ для независимых случайных величин с конечными математическими ожиданиями.