Теория Вероятностей - 24, МатФак ВШЭ

Андрей В. Дымов и Андрей В. Хохлов

4 июля 2025 г.

Содержание

1 Предварительно о теории вероятностей **4**

2 Колмогоровская аксиоматика теории вероятностей **7**
2.1 Множество элементарных исходов Ω . . . 7
2.2 Вероятностное пространство в случае не более, чем счетного множества Ω . 9
2.3 Вероятностное пространство: общий случай . . . 12

3 Условная вероятность и независимость **13**
3.1 Условная вероятность . . . 13
3.2 Несколько очень полезных соотношений . . . 16
3.3 Независимость . . . 18
3.4 Вероятностная модель для независимых испытаний . . . 19
3.5 Леммы Бореля-Кантелли . . . 22

4 Случайные величины **23**
4.1 Случайные величины и их распределения . . . 23
4.2 Типы случайных величин и векторов . . . 25
4.3 Функции распределения . . . 26
4.4 Независимость случайных величин . . . 30

5 Характеристики случайных величин **33**
5.1 Математическое ожидание . . . 33
5.2 Дисперсия . . . 35
5.3 Пространства L_p случайных величин и высшие моменты . . . 36
5.4 Ковариация и коэффициент корреляции . . . 37
5.5 Производящая функция случайной величины . . . 38
5.6 Медиана и квантили . . . 40
5.7 Неравенства для моментов . . . 41

6 Важнейшие дискретные распределения **42**
6.1 Распределние Бернулли (индикатор) . . . 42
6.2 Биномиальное распределение . . . 42
6.3 Геометрическое распределение . . . 43
6.4 Распределение Паскаля . . . 44
6.5 Распределение Пуассона . . . 45

7 Важнейшие непрерывные распределения **46**
7.1 Равномерные распределение . . . 46
7.2 Гауссовское или нормальное распределение . . . 47
7.3 Показательное или экспоненциальное распределение . . . 47
7.4 Логнормальное распределение . . . 47
7.5 Семейства распределений . . . 48

8 Вокруг закона больших чисел и центральной предельной теоремы **49**
8.1 Неравенства Маркова и Чебышёва . . . 49
8.2 Закон больших чисел в форме Чебышёва . . . 50
8.3 Некоторые применения закона больших чисел . . . 51
8.4 Центральная предельная теорема . . . 54

9 Связи вероятностной теории с естествознанием **56**
9.1 Содержательный пример: модель случайного блуждания . . . 57

10 Другие подходы в теории вероятностей. Неколмогоровские теории **60**
10.1 Частотный подход фон Мизеса . . . 60
10.2 Состояния и квантовая схема вероятностной модели. . . . 61
10.3 Байесовский подход . . . 63

11 Метод характеристических функци **65**
11.1 Определение и первые свойства . . . 65
11.2 Теорема обращения . . . 70
11.3 Характеристическая функция случайного вектора . . . 72

12 Сходимости случайных величин **73**
12.1 Слабая сходимость мер . . . 73
12.2 Сходимость случайных величин по распределению . . . 75
12.3 Связь сходимости по распределению с характеристическими функциями . . . 77

13 Закон больших чисел в форме Хинчина и центральная предельная теорема: доказательства **79**
13.0.1 Объяснение ЦПТ с точки зрения теории динамических систем . . . 80

14 Виды сходимостей случайных величин и их связи **80**
14.0.1 Доказательство импликации $\Leftarrow$ в теореме 12.11 . . . 83

15 Усиленный ЗБЧ, а также ЦПТ для разнородных независимых случайных величин **84**

16 Многомерное нормальное распределение **85**

A Напоминание: комбинаторика и простейшие формулы **88**
A.1 Простейшие базовые формулы . . . 88
A.2 Некоторые примеры . . . 88
A.2.1 Разложения шаров по ящикам. Различимые и неразличимые объекты 89
A.3 Методы вычислений . . . 90
A.3.1 Рекуррентные соотношения. Треугольник Паскаля . . . 90
A.4 Асимптотики и формула Стирлинга . . . 92

B Базовые объекты и конструкции теории меры **93**
B.1 Алгебры и сигма-алгебры . . . 93
B.2 Мера . . . 93
B.3 Прямое произведение пространств с мерой . . . 95
B.4 Образ меры при отображении . . . 96
B.5 Типы мер . . . 96
B.5.1 Сингулярные меры . . . 96
B.5.2 Абсолютно непрерывные меры . . . 97
B.5.3 Теорема о разложении меры . . . 97
B.6 Интеграл Лебега . . . 97
B.6.1 Построение интеграла . . . 98
B.6.2 Свойства интеграла Лебега . . . 99
B.6.3 Замена переменной в интеграле Лебега . . . 99
B.6.4 Нужно прокинуть предел внутрь матожидания? Вам сюда . . . 100
B.6.5 Пункт обмена интегралов местами . . . 100

C Список учебников **102**

1 Предварительно о теории вероятностей

Имеется значительное число учебников по Теории Вероятностей и никак нельзя сказать, что все они копируют друг друга, хотя набор основных тем обычно практически неизменен. Как правило в них в начале представлены сюжеты элементарной (конечной и дискретной) теории с (как бы взятыми из реальности) интерпретациями ее применений, далее обсуждают предельные случаи некоторых конечных моделей, теоретико-множественные конструкции— алгебры подмножеств и теорию меры, теорию интеграла Лебега на пространствах с мерой, разбор основных примеров сходимостей последовательностей случайных моделей[1]. При начальном изучении разница в подходах, может спровоцировать восприятие теории вероятностей как чисто-абстрактного, формального знания или, напротив, как набора эмпирически-эвристических рецептов, которые непонятно как выстраивать логически. При этом, в Теорию Вероятностей часто включают и чисто-практические фрагменты Математической Статистики.

Поэтому представляется разумным в самом начале сказать несколько слов, проясняющих хотя бы некоторые существующие расхождения во взглядах на предмет. В целом от теории вероятностей (по исторической традиции) ждут содержательных и понятных интерпретаций в разных областях человеческой деятельности с одной стороны, и в то же время строгого в рамках традиционных требований к математическим сюжетам изложения. При рассмотрении конечных и дискретных моделей сделать это было сравнительно просто, но оказалось невозможным одними только конечными вероятностными моделями представить, например, результаты измерений реальных физических процессов. Переход от конечных моделей к использующим актуальные бесконечности оказался весьма трудоемким и некоторый прогресс был достигнут лишь к 30-м годам XX века.

Неоднозначность вероятностной модели мира

Термин «вероятность» часто употребим при описании наблюдений за окружающим миром, но при этом вопрос можно ли *непосредственно* наблюдать вероятность обычно даже не формулируется. В лучшем случае предлагается приблизительная трактовка того, какие именно действия надо предпринять, чтобы *приблизительно(!)* сосчитать некую величину между 0 и 1. Но при этом в естественнонаучных теориях понятие вероятности (и в квантовой теории понятие амплитуды вероятности) используется уже в *точных* формулах — эта традиция несоответствия, похоже, не имеет признаков к изменению. По-видимому, предполагается, что Теория Вероятностей как математическая абстрактная дисциплина отвечает за все.

Если мы посмотрим чуть внимательней на проблему измерений реальных физических процессов, то легко увидим обстоятельства, которые послужили причиной тому, что в настоящее время существуют *различные* теории вероятностей, причем количественные выводы в каждой из этих теорий иногда не вполне ясно как сравнивать.

Во-первых, возможные разночтения состоят в формализации самого понятия **вероятности**: относится ли оно к единичному измерению или имеет смысл лишь для потенциально бесконечной *последовательности* измерений. Для описания свойств последовательностей измерений необходимо придать строгий смысл другому важному общеупотребительному термину – **независимости** измерений.

[1] В зависимости от вкусов автора учебника упор в изложении делается либо на интерпретациях (и таким образом курс приобретает как бы инженерно-прикладной характер – см. например, известный курс для слушателей Академии им. Жуковского Е.С.Вентцель "Теория Вероятностей"), либо на абстрактных, чисто математических утверждениях о пространствах с мерой (см. например, А.А.Боровков "Теория Вероятностей")

Второе обстоятельство состоит в том, что в окружающем мире для физических теорий всегда приняты масштабные ограничения — так, например, почти все согласны, что физические описания микромира специфичны и эти отличия не удается описать как простой дополнительный набор настраиваемых параметров в единой общей теории. Возникает следующий вопрос о строении теории: а можно ли надеяться, что соответствующие математические определения вероятности и независимости должны быть одинаковыми во всех случаях приложений теории, не зависеть от масштаба описываемого реального эксперимента?

Пока что просто перечислим основные современные направления, выросшие из исторического общеинтуитивного подхода в области конечных моделей:

- аксиоматический подход А.Н.Колмогорова (который единственно и будет в центре нашего внимания в дальнейшем), в рамках которого вероятность понимается как неотрицательная вещественнозначная мера $0 \leqslant \mathbb{P} \leqslant 1$ на подходящем множестве, а независимость определяется как соотношение на возникающие числовые значения.

- фриквентистский (частотный) подход Рихарда фон Мизеса, в рамках которого вероятность относится к повторениям измерений и определяется как свойство числовых рядов, этот подход (хотя и неявно) используется в Математической статистике. Заметим, что некоторые теоремы колмогоровской теории являются аксиомами для фриквентистов. Расхождение фриквентистов с колмогоровской теорией в общих чертах касается вопроса, к чему должна относиться вещественнозначная мера $0 \leqslant \mathbb{P} \leqslant 1$: к повторениям эксперимента или к единичному эксперименту.

- эвристический и свободный от какой-либо теоретико-множественной схемы байесианский подход[2], который однако же в конкретных задачах математической статистики иногда быстро приводит к ответам.

- квантовая теория вероятностей, которая фактически (но неявно) изначально использовалась в приложениях к задачам микромира, но, как оказалось, эта теория не сводима к колмогоровской аксиоматике и по существу ей противоречит. Специфика данной теории в том, что в ней присутствует линейная структура на рассматриваемых множествах.

- экзотические неколмогоровские теории, в которых рассматриваемая мера $\mathbb{P}$ не обязательно вещественнозначная, а например, комплекснозначная или p-адическая.

Пока что скажем, что колмогоровская теория вероятностей содержательно состоит из математических утверждений теории меры $\mathbb{P}$ применительно к аксиоматически введенному понятию независимости мер.

Необходимость теории

Само по себе понятие вероятности исторически казалось интуитивно ясным и использовалось для характеристики разных возможных ситуаций — это так называемая «оценка шансов». Однако желание иметь такую характеристику в численном выражении, с самого начала порождает трудности, известные под названием парадоксов теории вероятностей. Проиллюстрируем такую ситуацию следующим интуитивно понятным (но по вариантам

[2] В котором допустимо делать количественные вероятностные утверждения не только о данных в измерениях, подверженных случайным воздействиям. Например, фразы типа «Вероятность того, что Альберт Эйнштейн выпил чашку чая 1 августа, 1948 г. составляет 0.35» байесианцами считаются осмысленными, так как отражают частную силу веры в истинность предложения.

ответов парадоксальным) вопросом: *В двух карманах случайно лежат две монеты, какова вероятность что монет в карманах поровну*

Подсчет всех шансов в эксперименте «монеты в карманах» указывает как минимум на два разных подхода: см. левую и правую картинку на Рис.1.

Рис. 1: Два способа подсчета шансов в эксперименте «монеты в карманах»

Интуитивный подход к оценке «вероятности того, что монет поровну» связан со сравнением количества шансов, составляющих «благоприятное» событие, с количеством всех «шансов» . Однако здесь кроется явная неоднозначность численного ответа: априори неясно, какую именно модель «всех шансов» следует избрать для приведенной словесной формулировки, да и роль симметрии разных шансов совершенно неясна. Важный вывод: *численные оценки вероятности имеют смысл только в связке с конкретно выбранной моделью эксперимента*, а без выбора модели само по себе утверждение о численном значении вероятности представляется лишенным смысла.

Поскольку задача количественного описания решается только с предъявлением модели вычисления вероятности (например, множества исходов, если мы хотим оставаться в рамках языка теории множеств), то для решения прикладных задач необходимо правило создания вероятностных моделей. Такого единого правила нет, и существующие разные версии теории вероятностей возникли отчасти поэтому. Зафиксированное множество «шансов» принято называть **пространством элементарных событий данного эксперимента**, в его выборе обычно имеется некоторый произвол (сравните с известным вопросом какой стороной выпадает монета при подбрасывании: шутники предлагают рассматривать случаи вставания монеты на ребро и зависания монеты в воздухе).

Небольшой модификацией вопроса про карманы и монеты можно пояснить базовую идею байесовского подхода: пусть карманы различаются как правый и левый и вопрос состоит в том, как количественный ответ зависит от априорного знания «правый карман непуст». Байесовский подход вообще избегает предъявления пространства элементарных событий, вместо этого изучаются зависимости ответа от описания эксперимента.

Интерпретации теории

Исторически так уж сложилось, что утверждения теории вероятностей широко используются как в самых разных приложениях так и в теоретической физике. Но надо признать, что в силу той же традиции описания моделей приводятся не всегда[3]. При практических

[3]Например, современные попытки аксиоматически сформулировать принципы квантовой механики или квантовой теории поля включают аксиомы, в которые вероятность входит как готовый параметр

рассмотрениях подбирают подходящую математическую модель, то есть пространство с мерой такое, что возможны количественные оценки вероятностей результатов экспериментов. Вопрос об оптимальном подборе такой модели не является частью теории, но зато *уже в рамках выбранной модели можно ставить вопрос о количественном соответствии реальности и вычисляемой в модели вероятности*: вообще говоря, процедура соответствия основана на повторениях экспериментов (измерений), методы проверки соответствия составляют отдельную дисциплину, которая называется Математической Статистикой.

При изложении любой теории возникает список основных понятий. План этого вводного курса теории вероятностей состоит в том, чтобы проследить за формулировками, начав с достаточно простых интуитивно ясных моделей и дополняя соответствующие построения до более общих случаев. Можно надеяться, что на этом пути связь аксиоматически вводимых понятий с интуитивным их пониманием (возникшем еще до всяческих теорий в процессе практической деятельности) станет более понятна.

2 Колмогоровская аксиоматика теории вероятностей

С этого момента и до конца нашего курса мы будем заниматься колмогоровской аксиоматической теорией вероятностей. Однако следует помнить, что помимо колмогоровского принципа используются и другие (неколмогоровские) теории, причем их использование весьма распространено — мы коротко поговорим о них чуть позже.

Пусть имеется эксперимент, исходы которого мы хотели бы интерпретировать как случайные. Наша цель — вычислить какие-нибудь вероятности, связанные с этими исходами. Например, мы бросаем два игральных кубика, на гранях каждого из которых написаны числа от 1 до 6, и хотим найти вероятность того, что сумма результатов двух бросков делится на 3. При этом заметим, что понятие "вероятность" пока не определено.

Возникает две задачи. Первая — формулировка математической модели, описывающей этот эксперимент и, в частности, определение понятия вероятность. Вторая — непосредственное вычисление интересующих нас вероятностей в рамках построенной модели.

В колмогоровской аксиоматике математическая модель состоит из трех компонент: множество исходов Ω нашего эксперимента (выше мы его называли множество "шансов"), сигма-алгебра его подмножеств $\mathcal{F}$ (вероятность только этих подмножеств и окажется возможным измерить), и вероятность $\mathbb{P}$, являющаяся мерой на измеримом пространстве $(\Omega, \mathcal{F})$. При этом в случае, когда множество Ω не более, чем счетно, можно выбрать $\mathcal{F} = 2^{\Omega}$ и исключить сигма-алгебру $\mathcal{F}$ из рассмотрения. Далее мы обсуждаем обозначенные компоненты детально.

2.1 Множество элементарных исходов Ω

Колмогоровский подход предполагает, что любой эксперимент можно описать в терминах набора альтернатив ω, называемых *элементарными исходами*. Их объединение составляет некоторое множество Ω, называемое *множеством элементарных исходов*. Для каждого эксперимента множество элементарных исходов Ω выбирается индивидуально, причем этот выбор зависит от того, что именно вас интересует в рассматриваемом эксперименте, и основан на вашем личном жизненном опыте и вашей интуиции. Математика же начинается *только* с того момента, когда множество Ω зафиксировано.

Несмотря на то, что Ω также называют *пространством элементарных исходов*, на Ω *не* предполагается наличия никакой структуры, в частности не требуется структуры линейного пространства.

теории, что это означает с точки зрения постановки эксперимента часто оставляют без разъяснений.

Пример 2.1. Эксперимент состоит в подбрасывании одной монеты $n \geq 1$ раз. Нас интересует, что выпадает при каждом броске, орел или решка. Разумно выбрать

$$\Omega = \{"\text{О}" ,"\text{Р}" \}^n = \big\{(i_1, \dots, i_n) : \; i_j = "\text{О}" \text{ или } "\text{Р}" \;, \; j = 1, \dots, n\big\}. \tag{2.1}$$

Таким образом, элементарными исходами нашего эксперимента являются всевозможные вектора $(i_1, \dots, i_n) \in \Omega$, где i_j ="О" , если на j-ом броске выпал орел, и i_j ="Р" в противном случае.

Но, если вы ни разу в жизни не видели, чтобы при подбрасывании монеты выпадала решка, и вы не верите людям, которые говорят, что решка иногда таки выпадает, вам было бы разумно выбрать

$$\Omega = \{0\}. \tag{2.2}$$

Здесь $\omega = 0$ отвечает единственному рассматриваемому вами исходу эксперимента, когда выпало n орлов подряд. [4]

Пример 2.2. Множество элементарных исходов (2.1) также описывает эксперимент, состоящий в одновременном подбрасывании n *пронумерованных* монет. Так как монеты пронумерованы, они *различимы*. В частности, при $n = 2$ построенное множество элементарных исходов можно записать в виде

$$\Omega = \{"\text{ОО}" \;, \; "\text{ОР}" \;, \; "\text{РО}" \;, \; "\text{РР}" \;\}. \tag{2.3}$$

Такое множество Ω схематически изображено на правой части рис. 1.

В ситуации же, когда монеты *неразличимы*, изображенной на левой части рис. 1, разумно выбрать

$$\Omega = \{"\text{ОО}" \;, \; "\text{ОР}" \;, \; "\text{РР}" \;\}.$$

При такой модели эксперимента мы можем говорить только о количестве выпавших орлов и решек, но не о последовательности в которой они выпадали, так что элементарные исходы "ОР" и "РО" из (2.3) в этом случае отождествляются.

Пример 2.3. Рассмотрим упоминавшийся выше эксперимент с бросанием двух кубиков. Если считать кубики различимыми, то разумно выбрать

$$\Omega = \big\{(i, j) : \; 1 \leq i, j \leq 6\big\}, \tag{2.4}$$

где i — число, выпавшее на первом кубике, а j — на втором. Если считать их неразличимыми, то

$$\Omega = \big\{(i, j) : \; 1 \leq i \leq j \leq 6\big\}, \tag{2.5}$$

Если же теперь вспомнить, что нас интересовала лишь сумма выпавших очков, при этом нам не важно как именно сумма разбита по двум кубикам, мы можем выбрать

$$\Omega = \big\{2, 3, \dots, 11, 12\big\}. \tag{2.6}$$

Пример 2.4. Вы знаете, что к вам зайдет знакомый между 8 и 9 часами вечера, и никакой дополнительной информации у вас нет. Разумно считать время его визита случайным числом между 8 и 9 часами. Выбора $\Omega = [0, 60] \subset \mathbb{R}$ хватит для любой точности измерений. Однако, если мы округляем время до минуты, то можно положить $\Omega = \{0, 1, \dots, 60\}$. А можно и оставить прежний выбор $\Omega = [0, 60]$, просто нецелые элементарные исходы $\omega \in \Omega$ никогда не реализуются. Заметим, что обычно в подобных задачах рассматриваемый отрезок сжимают до отрезка единичной длины, выбирая $\Omega = [0, 1]$.

[4] Несмотря на все "реверансы" , которые мы делаем и будем делать по поводу субъективности выбора математической модели рассматриваемого эксперимента, на контрольных мероприятиях выбор (2.2) мы вряд ли оценим, хотя с *математической* точки зрения он вполне корректен.

2.2 Вероятностное пространство в случае не более, чем счетного множества Ω

В примере 2.4 предлагается рассматривать континуальное множество элементарных исходов Ω. Этот случай заметно сложнее, чем ситуация, рассмотренная во всех остальных примерах, в которых множество Ω было не более, чем счетным. Вскоре мы вернемся к общему случаю, а пока, для упрощения восприятия, сделаем

Предположение 2.5. Множество элементарных исходов Ω конечно либо счетно.

Рассмотрим функцию $p : \Omega \mapsto [0, 1]$, удовлетворяющую свойству

$$\sum_{\omega \in \Omega} p(\omega) = 1. \tag{2.7}$$

Число $p(\omega)$ будем называть *вероятностью элементарного исхода* ω.

Событием или *случайный событием* мы будем называть произвольное подмножество $A \subset \Omega$. Эта терминология довольно естественна: событие состоит из набора элементарных исходов. Так, в примере 2.2 можно рассмотреть событие $A = \{$"OP" , "PO" $\}$, состоящее в том, что выпали один орел и одна решка.

Наконец, определим функцию множеств $\mathbb{P} : 2^{\Omega} \mapsto [0, 1]$, сопоставляющую каждому событию $A \subset \Omega$ число по правилу [5]

$$\mathbb{P}(A) = \sum_{\omega \in A} p(\omega). \tag{2.8}$$

Функция $\mathbb{P}$ называется *распределением вероятностей*, или *вероятностной мерой*, или просто *вероятностью*, а число $\mathbb{P}(A)$ — *вероятностью события* A. Отметим, что $\mathbb{P}$ действительно является счетно-аддитивной мерой.

Из определения сразу вытекает следующий результат (докажите его!):

Лемма 2.6. *1)* $\mathbb{P}(\emptyset) = 0$*;*

2) $\mathbb{P}(\Omega) = 1$*;*

3) $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$ *для любого события* $A \subset \Omega$ *и его дополнения* $A^c := \Omega \setminus A$*;*

3) Для любых событий $A, B \subset \Omega$*,*

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B).$$

В частности, если $A \cap B = \emptyset$*,*

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B).$$

Традиционно выделяют ситуацию, особенно часто встречающуюся в приложениях, когда вероятности $p(\omega)$ всех элементарных исходов $\omega \in \Omega$ равны. Тогда, согласно требованию (2.7), находим

$$p(\omega) = \frac{1}{|\Omega|} \quad \forall \omega \in \Omega \quad \text{и} \quad \mathbb{P}(A) = \frac{|A|}{|\Omega|} \quad \forall A \subset \Omega.$$

Такое определение вероятности называют *классическим*.

Наконец, мы готовы дать основное определение, которое однако работает *только в рамках предположения 2.5.*

[5] Так как Ω не более, чем счетно и $p(\omega) \geq 0$, сумма ряда (2.8) хорошо определена, то есть конечна и не зависит от порядка суммирования.

Определение 2.7 (случай не более, чем счетного Ω)**.** Пара $(\Omega, \mathbb{P})$ называется *вероятностным пространством.*

! Выбрать вероятностное пространство и означает построить математическую модель вашего эксперимента **!**

Важно очень хорошо понимать, что выбор вероятностного пространства субъективен, связан с вашей "физической" интуицией, и не может быть правильным или неправильным, потому что критерий правильности не определен. Интересуясь вопросом "с какой вероятностью во время утренней пробежки я встречу динозавра" , вполне можно положить $\Omega = \{$"Да" , "Нет" $\}$ с вероятностями элементарных исходов $p($"Да" $) = p($"Нет" $) = 1/2$. Нельзя утверждать, что этот выбор неправильный, но можно лишь сказать, что он плохо соответствует жизненному опыту большинства из нас. Выбор вероятностного пространства — это задание аксиоматики, в рамках которой вы планируете работать.

Легко провести аналогию с геометрией: пока вы не задали аксиоматику, нельзя ответить на вопрос "верно ли, что через точку, не лежащую на данной прямой, можно провести лишь одну прямую, параллельную данной?" . Действительно, в евклидовой геометрии, к которой большинство из нас приучил жизненный опыт, ответ на этот вопрос положителен, а в геометрии Лобачевского — отрицателен. Одну из этих геометрий *разумно* применять в одних задачах, другую — в других. *Математика* же начинается *только* после того, как задана аксиоматика. В колмогоровской теории вероятностей — как только введено вероятностное пространство $(\Omega, \mathbb{P})$. И только после этого утверждения "правильно" или "неправильно" обретают смысл.

Итак, решение любой "текстовой" задачи по теории вероятностей состоит из двух этапов. Первый этап не имеет прямого отношения к математике и состоит в выборе вероятностного пространства; это может быть сделано разными способами. Второй этап состоит в решении *математической* задачи в рамках выбранной модели.

Пример 2.8. В примере 2.1 имеем $|\Omega| = 2^n$. Следовательно, классическое определение вероятности дает

$$p(\omega) = \frac{1}{2^n} \quad \forall \omega \in \Omega \quad \text{и} \quad \mathbb{P}(A) = \frac{|A|}{2^n} \quad \forall A \subset \Omega.$$

Такой выбор разумен, если вы считаете, что ваша монетка симметрична. А теперь обдумайте как определить вероятностное пространство в случае, когда монетка, которую мы бросаем, такова, что вероятности выпадения орла и решки при единичном броске не равны. Пока вы размышляете, мы двинемся дальше, но обязательно вернемся к этому сюжету позже, когда введем ключевое понятие *независимости событий.* Отметим только некорректность предложенного вопроса: мы начали говорить о вероятности выпадения орла либо решки *до* определения вероятностного пространства.

Пример 2.9. Вернемся к задаче о подбрасывании двух кубиков, которую мы начали обсуждать в начале параграфа 2 и продолжили в примере 2.3. Напомним, что нас интересует вероятность события

$$A = \{\text{"сумма результатов двух бросков кратна 3" }\}. \tag{2.9}$$

а) В случае различимых кубиков, когда множество элементарных исходов выбрано в виде (2.4), получаем, что событие A состоит из элементарных исходов $(1, 2), (1, 5), (2, 4), (3, 6), (4, 5)$, симметричных им исходов $(2, 1), (5, 1), (4, 2), (6, 3), (5, 4)$, а также исходов $(3, 3), (6, 6)$. Таким образом, $|A| = 12$, а $|\Omega| = 36$, так что классическое определение вероятности дает

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|} = \frac{1}{3}. \tag{2.10}$$

б) В случае, когда кубики считаются неразличимыми, то есть множество элементарных исходов имеет вид (2.5), получаем $A = \{(1,2),(1,5),(2,4),(3,6),(4,5),(3,3),(6,6)\}$, так что $|A| = 7$. Так как $|\Omega| = 21$, классическое определение вероятности дает

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|} = \frac{1}{3}. \tag{2.11}$$

Нам повезло — вероятность, вычисленная для неразличимых кубиков совпала с вероятностью, вычисленной для различимых. Вообще говоря, это далеко не всегда так.

в) В ситуации, когда множество элементарных исходов выбрано в виде (2.6), $|\Omega| = 11$, $A = \{3, 6, 9, 12\}$, так что классическое определение вероятности дает

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|} = \frac{4}{11}. \tag{2.12}$$

Разные выборы моделей ведут к разным вероятностям одного и того же события. Возможно, в этом месте читатель захочет закрыть конспект со словами "нет, ну это чушь, конечно единственный правильный выбор вероятностного пространства сделан в пункте (а), и вообще не понятно, где может возникнуть ситуация неразличимых объектов" . Действительно, если у нас есть два совершенно одинаковых, неразличимых кубика, можно один покрасить в синий цвет, а второй — в красный, и конечно вероятность события (2.9) от этого не изменится. Ну, а модель (в) не имеет вообще никакого отношения к реальности, так как, конечно, сумма будет равна 6 гораздо чаще, чем 12. Однако, все эти разговоры имеют отношение лишь к соответствию выбранной модели окружающей нас действительности и нашей интуиции в отношении слова "вероятность" , в то время как с точки зрения математики все три модели *совершенно* корректны. Более того, к интуиции нужно относиться очень аккуратно. Вспоминая известный феномен кота Шрёдингера, мы убеждаемся, что даже наблюдение за системой, уж не говоря о раскраске ее в разные цвета, далеко не невинное действие. А рассуждая о соответствии действительности модели пункта (в), мы отождествили частоту встречаемости события при повторном эксперименте с его вероятностью при единичном, но это совершенно разные объекты. Результат же, утверждающий, что при увеличении числа независимых повторений эксперимента частота, с которой происходит данное событие, стремится к его вероятности при единичном эксперименте, действительно имеет место и называется *закон больших чисел (ЗБЧ)*. Это один из центральных результатов нашего курса, он требует уточнения и доказательства (в отличие от фриквентистского подхода, в котором ЗБЧ постулируется и фактически является определением вероятности).

Завершим это обсуждение следующим примером, подтверждающим наш скепсис относительно интуиции, особенно когда дело касается непривычных нам условий; в данном случае — непривычных нам масштабов микромира.

Пример 2.10. Распределение n бозонов по m уровням энергии [6] управляется так называемой статистикой Бозе-Эйнштейна. Она соответствует классическому определению вероятности в задаче о размещении n *неразличимых* частиц по m *различимым* ячейкам. В этом случае множество элементарных исходов можно выбрать в виде

$$\Omega = \{\omega = (n_1, \ldots, n_m) : n_1 + \cdots + n_m = n\}.$$

Событие $A_{i,k}$, состоящее в том, что в ячейке номер i содержится ровно k частиц, имеет вид $A_{i,k} = \{\omega : n_i = k\}$. С помощью стандартного метода "шаров и перегородок" (см. аппендикс A.2.1) можно проверить, что $|A_{i,k}| = C^n_{n+m-1}$ и $|\Omega| = C^{n-k}_{n+m-k-2}$, так что

$$\mathbb{P}(A) = \frac{C^n_{n+m-1}}{C^{n-k}_{n+m-k-2}}.$$

[6] При микроканоническом ансамбле.

Резюмируем. Пусть вам поставили следующую задачу: имеется 6 ящиков, в которые случайным образом разбросали 15 шаров. Требуется найти вероятность, что останутся пустые ящики. Прежде, чем ее решать, необходимо выяснить, предполагаются ли шары различимыми, и предполагаются ли ящики различимыми. Очевидно, имеется 4 варианта, каждый из которых требует своего решения задачи, см. Аппендикс A.2.1.

2.3 Вероятностное пространство: общий случай

В случае, когда множество элементарных исходов Ω несчетно (как в примере 2.4), определение вероятности, данное в (2.7), (2.8), очевидно не работает. Поэтому предлагается его нижеследующее обобщение, основанное на теории меры. См. Аппендикс B, в котором мы напоминаем некоторые базовые факты теории меры.

Предположим, что на множестве Ω задана сигма-алгебра его подмножеств $\mathcal{F}$.

Определение 2.11. Неотрицательная мера $\mathbb{P}$ на измеримом пространстве $(\Omega, \mathcal{F})$ называется *вероятностной мерой* или *вероятностным распределением*, если $\mathbb{P}(\Omega) = 1$.

Следующий результат немедленно следует из стандартных свойств меры и тождества $\mathbb{P}(\Omega) = 1$ (проверьте!).

Лемма 2.12. *Верны утверждения леммы 2.6, где $A, B \in \mathcal{F}$.*

Теперь дадим основное определение вероятностного пространства, работающее для произвольного множества элементарных исходов Ω (в отличие от определения 2.7). Если вас просят дать определение вероятностного пространства, нужно приводить именно его.

Определение 2.13. *Вероятностным пространством* называется тройка $(\Omega, \mathcal{F}, \mathbb{P})$, где Ω — множество элементарных исходов, $\mathcal{F}$ — σ-алгебра его подмножеств, а $\mathbb{P}$ — вероятностная мера на измеримом пространстве $(\Omega, \mathcal{F})$.

Определение 2.7 является частным случаем определения 2.13. Действительно, легко проверить, что функция множеств (2.8) задает вероятностную меру на измеримом пространстве $(\Omega, \mathcal{F})$, где $\mathcal{F} = 2^{\Omega}$ — σ-алгебра всех подмножеств (не более, чем счетного) множества Ω.

Пример 2.14. Рассмотрим вероятностное пространство $(\Omega, \mathcal{F}, \mathbb{P}) = ([0,1], \mathcal{B}[0,1], Leb)$, где $\mathcal{B}[0,1]$ обозначает σ-алгебру борелевских подмножеств отрезка $[0,1]$, а Leb обозначает меру Лебега на $([0,1], \mathcal{B}[0,1])$. Такое вероятностное распределение $\mathbb{P}$ называется *равномерным распределением* на отрезке $[0,1]$. Предложенная вероятностное пространство разумно для примера (2.4).

Отметим, что в примере 2.14 $\mathbb{P}\{\omega\} = 0$ для каждого элементарного исхода $\omega \in \Omega$. Это характерно для многих вероятностных моделей с несчетным множеством элементарных исходов (в отличие от случая счетного множества Ω, см. (2.7)).

Подчеркнем появление σ-алгебры $\mathcal{F}$ в определении 2.7, в отличие от определения 2.13. Оно обусловлено тем, что область определения меры $\mathbb{P}$ не обязана содержать все подмножества множества Ω. Уже в примере 2.14 это не так: известно, что для меры Лебега существуют неизмеримые подмножества, поэтому приходится ограничить область определения $\mathbb{P}$ сигма-алгеброй $\mathcal{F} = \mathcal{B}[0,1]$. [7] Таким образом, интерпретация σ-алгебры $\mathcal{F}$ такова: она задает набор множеств, вероятность которых мы можем измерить.

[7] В этом примере мы могли бы выбрать вместо борелевской сигма-алгебры более широкую σ-алгебру подмножеств, измеримых по Лебегу. Но для целей теории вероятностей обычно хватает выбора $\mathcal{F} = \mathcal{B}[0,1]$.

Отметим естественность требования, что область определения $\mathcal{F}$ вероятности является алгеброй. К примеру, если имеются два события $A, B \in \mathcal{F}$, естественно задать вопросы "с какой вероятностью произойдет хотя бы одно из событий A и B" и "с какой вероятностью произойдут оба события A и B" . Другими словами, требуется найти вероятности $\mathbb{P}(A \cup B)$ и $\mathbb{P}(A \cap B)$.

Свойство $\mathcal{F}$ быть не только алгеброй, но и σ-алгеброй, а вероятности $\mathbb{P}$ быть σ-аддитивной, также оказывается весьма полезным. В частности, лемма B.9(4) утверждает следующий естественный факт (неверный без свойства σ-аддитивности меры $\mathbb{P}$!): если имеется последовательность "уменьшающихся" событий $C_i \in \mathcal{F}$, то есть таких, что $C_{i+1} \subset C_i$, и их вероятности $\mathbb{P}(C_i)$ стремятся к нулю при $i \to \infty$, то вероятность того, что они все произойдут сразу равна нулю:

$$\mathbb{P}(\bigcap_i C_i) = 0.$$

Это свойство называется *непрерывностью вероятности и нуле.*

Задача 2.15. 1. Проинтерпретируйте пункты (2,3) леммы B.9 с точки зрения теории вероятностей, аналогично тому, как это сделано выше для пункта (4).
2. Пусть μ — неотрицательная аддитивная функция множеств на измеримом пространстве $(\Omega, \mathcal{F})$. Докажите, что для любых не пересекающихся друг с другом множеств $C_i \in \mathcal{F}$, $i \geq 1$, выполнено

$$\mu(\bigcup_{i=1}^{\infty} C_i) \geq \sum_{i=1}^{\infty} \mu(C_i).$$

3 Условная вероятность и независимость

3.1 Условная вероятность

Рассмотрим следующую задачу.

Пример 3.1. В семье имеется два (различимых!) ребенка. Какова вероятность того, что оба — девочки? Разумным выбором вероятностного пространства для этой задачи представляется $(\Omega, \mathcal{F}, \mathbb{P})$, где

$$\Omega = \{\text{ММ, МД, ДМ, ДД}\}, \qquad \mathbb{P}\{\omega\} = \frac{1}{|\Omega|} = \frac{1}{4} \quad \forall \omega \in \Omega, \qquad \mathcal{F} = 2^{\Omega}. \tag{3.1}$$

Тогда интересующее нас событие имеет вид $A = \{\text{ДД}\}$, так что $\mathbb{P}(A) = 1/4$.

Зададимся теперь вопросом, какова вероятность того, что оба ребенка — девочки, при условии, что хотя бы один ребенок — девочка? [8] Введенная в предыдущем параграфе аксиоматика не дает никакого рецепта как можно было бы учесть это условие в рамках предложенной в примере 3.1 модели. Разумным представляется построить новое вероятностное пространство $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{\mathbb{P}})$, учитывающее это условие:

$$\widetilde{\Omega} = \{\text{МД, ДМ, ДД}\}, \qquad \widetilde{\mathbb{P}}\{\omega\} = \frac{1}{|\Omega|} = \frac{1}{3} \quad \forall \omega \in \Omega, \qquad \mathcal{F} = 2^{\widetilde{\Omega}}. \tag{3.2}$$

В этом случае интересующее нас событие $A = \{\text{ДД}\}$ имеет вероятность $\mathbb{P}(A) = 1/3$.

Однако, такой подход неудобен. Если мы захотим поменять рассматриваемое условие на другое (например, "второй ребенок — мальчик"), нам придется строить новое вероятностное пространство. В приложениях же (и математических тоже) регулярно требуется

[8] Другими словами: если известно, что хотя бы один ребенок — девочка.

вычислять вероятность одного и того же события при множестве различных предположений. Например, страховую компанию может интересовать вероятность события A_I, что автомобиль попадет в ДТП с ущербом, лежащем в интервале $I \subset \mathbb{R}$ рублей. При этом разумно оценивать не только (безусловную) вероятность $\mathbb{P}(A_I)$, но и вероятности события A при условии, что автомобиль был той или иной марки, отдельно для каждой марки автомобиля. Следующее определение позволяет обойтись одним и тем же вероятностным пространством при вычислении всевозможных вероятностей со всевозможными условиями.

Рассмотрим произвольное вероятностное пространство $(\Omega, \mathcal{F}, \mathbb{P})$ и события $A, H \in \mathcal{F}$, причем $\mathbb{P}(H) > 0$.

Определение 3.2. *Условной вероятностью события A при условии события H*, $\mathbb{P}(H) > 0$, называется число [9]

$$\mathbb{P}(A|H) = \frac{\mathbb{P}(A \cap H)}{\mathbb{P}(H)}. \tag{3.3}$$

Если $\mathbb{P}(H) = 0$, то условная вероятность $\mathbb{P}(A|H)$ не определена. Общественный договор таков: если вас просят вычислить вероятность какого-либо события A при зафиксированных дополнительных данных H, всегда подразумевается условная вероятность в смысле определения 3.2. При этом формулировки могут быть совершенно различны. Вместо того, чтобы говорить об «условной вероятности того, что случайно выбранное лицо страдает дальтонизмом при условии, что это лицо — женщина», можно говорить о вероятности того, что женщина страдает дальтонизмом. Часто слова «при условии H» заменяют словами, «если известно, что H произошло». Короче говоря, наши формулы и символы не допускают никакой двусмысленности, но словесные выражения часто недостаточно четки и требуют точного истолкования. В противоположность условной вероятности $\mathbb{P}(A|H)$ говорят для ясности о безусловной вероятности $\mathbb{P}(A)$. Строго говоря, эпитет «безусловная» является лишним, и его можно опускать.

Убедимся, что определенная таким образом условная вероятность отвечает нашей интуиции.

Пример 3.3. Пусть в примере 3.1 требуется вычислить вероятность того, что оба ребенка — девочки, при условии, что хотя бы один — девочка. Другими словами, требуется найти условную вероятность $\mathbb{P}(A|H)$, где $H = \{\text{МД, ДМ, ДД}\}$. По определению 3.2,

$$\mathbb{P}(A|H) = \frac{\mathbb{P}(A \cap H)}{\mathbb{P}(H)} = \frac{\mathbb{P}\{\text{ДД}\}}{\mathbb{P}(H)} = \frac{1/4}{3/4} = \frac{1}{3}.$$

Мы видим, что ответ совпадает с результатом, полученным ранее в рамках модели (3.2).

Пример 3.4. Рассмотрим дискретное вероятностное пространство $(\Omega, \mathcal{F}, \mathbb{P})$, где Ω — не более, чем счетное множество, вероятность $\mathbb{P}$ задана формулами (2.7)-(2.8), а $\mathcal{F} = 2^{\Omega}$. Пусть $H \subset \Omega$. Тогда, согласно формуле (3.3), $\forall A \in \Omega$,

$$\mathbb{P}(A|H) = \sum_{\omega \in A} p_H(\omega), \quad \text{где} \quad p_H(\omega) = \begin{cases} \frac{p(\omega)}{\mathbb{P}(H)}, & \text{если} \quad \omega \in H, \\ 0, & \text{если} \quad \omega \notin H. \end{cases}$$

Заметим, что $\mathbb{P}(\Omega|H) = \sum_{\omega \in \Omega} p_H(\omega) = 1$.

Пример (3.4) проясняет смысл происходящего. В случае дискретного вероятностного пространства условная вероятность устроена точно так же, как и безусловная, то есть

[9] Нередко событие H называют *гипотезой*.

задается формулами (2.7)-(2.8), но вероятности элементарных исходов $p(\omega)$ модифицированы. А именно, вероятность исходов ω, противоречащих событию H (то есть, $\omega \notin H$), обнуляются, а вероятности остальных ω нормируются так, чтобы их сумма была равна 1.

В общем случае (3.3) ситуация аналогична: вероятность куска $A \setminus H$ события A, противоречащего H, обнуляется, а вероятность оставшегося куска $A \cap H$ нормируется мультипликативной константой так, чтобы $\mathbb{P}(\Omega|H) = 1$.

Теперь обсудим свойства условных вероятностей. Основная их часть следует из следующей тривиальной леммы.

Лемма 3.5. *Условная вероятность* [10] $\mathbb{P}(\cdot|H)$ *является вероятностной мерой на измеримом пространстве* $(\Omega, \mathcal{F})$.

Доказательство. Из определения 3.2 следует

- Неотрицательность: $\mathbb{P}(A|H) \geq 0 \ \forall A \in \mathcal{F}$,
- Нормировка: $\mathbb{P}(\Omega|H) = \dfrac{\mathbb{P}(H)}{\mathbb{P}(H)} = 1$,
- σ-аддитивность:

$$\mathbb{P}(\cup_{i=1}^{\infty} A_i|H) = \frac{\mathbb{P}\big(\cup_{i=1}^{\infty} (A_i \cap H)\big)}{\mathbb{P}(H)} = \frac{\sum_{i=1}^{\infty} \mathbb{P}(A_i \cap H)}{\mathbb{P}(H)} = \sum_{i=1}^{\infty} \mathbb{P}(A_i|H),$$

  для любого набора событий $A_i \in \mathcal{F}$, таких что $A_i \cap A_j = \emptyset$ при $i \neq j$.

□

В частности, каждое событие $H \in \mathcal{F}$ естественным образом порождает новое вероятностное пространство $(\Omega, \mathcal{F}, \mathbb{P}(\cdot|H))$, отличающееся от старого выбором вероятности.

Согласно лемме 3.5, условная вероятность $\mathbb{P}(\cdot|H)$ удовлетворяет всем утверждениям, касающимся безусловной вероятности $\mathbb{P}(\cdot)$. Например, для любых событий $A, B \in \mathcal{F}$,

$$\mathbb{P}(A \cup B|H) = \mathbb{P}(A|H) + \mathbb{P}(B|H) - \mathbb{P}(A \cap B|H).$$

Еще раз подчеркнем, что это равенство не требует отдельного доказательства, а следует из лемм 3.5 и 2.12. Отметим, однако, что сказанное выше касается лишь функции $\mathbb{P}(\cdot|H)$ при фиксированном множестве H.

Задача 3.6. Постройте пример вероятностного пространства $(\Omega, \mathcal{F}, \mathbb{P})$ и событий $A, H \in \mathcal{F}$, $\mathbb{P}(H) \neq 0$, таких что
а) $\mathbb{P}(A|H) + \mathbb{P}(A|H^c) \neq 1$,
б) $\mathbb{P}(A|H) + \mathbb{P}(A^c|H^c) \neq 1$.

Все конструкции, построенные для безусловной вероятности, можно применить и к условной. В частности, конструкцию условной вероятности:

Задача 3.7. Пусть $H_1, H_2 \in \mathcal{F}$. Обозначим $\mathbb{P}_{H_i} := \mathbb{P}(\cdot|H_i)$, $i = 1, 2$. Докажите, что

$$\mathbb{P}_{H_1}(\cdot|H_2) = \mathbb{P}(\cdot|H_1 \cap H_2) = \mathbb{P}_{H_2}(\cdot|H_1).$$

То есть, для любого $A \in \mathcal{F}$,

$$\mathbb{P}_{H_1}(A|H_2) = \mathbb{P}(A|H_1 \cap H_2) = \mathbb{P}_{H_2}(A|H_1).$$

[10] Значок $\cdot$ в формуле $\mathbb{P}(\cdot|H)$ указывает на место, куда нужно подставить аргумент при вычислении значений рассматриваемой функции. То есть, $\mathbb{P}(\cdot|H)$ — функция, сопоставляющая множествам $A \in \mathcal{F}$ числа $\mathbb{P}(A|H)$.

3.2 Несколько очень полезных соотношений

В этом параграфе для пересечения событий $A_1 \ldots, A_n \in \mathcal{F}$ мы будем использовать обозначение

$$A_1 \ldots A_n := A_1 \cap \cdots \cap A_n.$$

Все результаты этого параграфа совершенно тривиальны с математической точки зрения, но удивительно эффективны в приложениях.

Формула умножения вероятностей

Пусть $A, B \in \mathcal{F}$, причем $\mathbb{P}(A) \neq 0$. Тогда из определения условной вероятности (3.3) сразу следует соотношение

$$\mathbb{P}(AB) = \mathbb{P}(B|A)\mathbb{P}(A). \tag{3.4}$$

По индукции получаем *формулу умножения вероятностей*:

$$\mathbb{P}(A_1 \ldots A_n) = \mathbb{P}(A_1)\mathbb{P}(A_2|A_1) \cdots \mathbb{P}(A_n|A_1 \ldots A_{n-1}).$$

Формула полной вероятности

Будем говорить, что события $H_1, \ldots, H_n$ задают *разбиение* события A, если $H_i \cap H_j = \emptyset$ при $i \neq j$ и $A \subset \cup_{i=1}^n H_i$.

Следующая формула является основным средством вычисления вероятности событий, естественным образом разбивающихся на несколько кусков.

Предложение 3.8. *Пусть события $H_1, \ldots, H_n$ задают разбиение события A и имеют положительные вероятности $\mathbb{P}(H_i) > 0$. Тогда*

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

Доказательство. Первое равенство следует из аддитивности меры и соотношения $A = (AH_1) \cup \cdots \cup (AH_n)$, $(AH_i) \cap (AH_j) = \emptyset$ для $i \neq j$. Второе равенство следует из первого и формулы (3.4). □

На практике события H_i обычно выбираются так, чтобы $\cup_{i=1}^n H_i = \Omega$. В этом случае говорят, что семейство H_i образует *полную группу событий*.

Пример 3.9. Имеется две урны, в одной 1 черный шар и 2 белых, а во второй по одному черному и белому шару. Вы наугад выбираете урну и вытаскиваете оттуда шар. Какова вероятность вытащить белый?

Сперва нужно построить вероятностное пространство, отвечающее этому эксперименту. Шары будем считать пронумерованными, так что множество элементарных исходов разумно выбрать в виде

$$\Omega = \{U_1B, U_1W_1, U_1W_2, U_2B, U_2W\},$$

где префикс "U_j" отвечает выбору j-ой урны, а B и W обозначают цвет выбранного шара: "black" или "white" .

Теперь нужно определить вероятности элементарных исходов. Условие задачи разумно интерпретировать следующим образом:

$$\mathbb{P}(U_1) = \mathbb{P}(U_2) = 1/2, \tag{3.5}$$

где $U_1 = \{U_1B, U_1W_1, U_1W_2\}$ — событие, что мы выбрали первую урну, а U_2 определено аналогично. Кроме того, "внутри" каждой урны вероятности распределены классическим образом:

$$\mathbb{P}(U_1B|U_1) = \mathbb{P}(U_1W_1|U_1) = \mathbb{P}(U_1W_2|U_1) = 1/3, \quad \mathbb{P}(U_2B|U_2) = \mathbb{P}(U_2W|U_2) = 1/2. \tag{3.6}$$

Согласно формуле умножения вероятностей (3.4), эти условия однозначно определяют вероятности элементарных исходов:

$$\mathbb{P}(U_1B) = \mathbb{P}(U_1W_1) = \mathbb{P}(U_1W_2) = 1/6, \quad \mathbb{P}(U_2B) = \mathbb{P}(U_2W) = 1/4.$$

Теперь конструкция, предложенная в параграфе 2.2, дает желаемое вероятностное пространство, и мы можем приступить к вычислению требуемой вероятности события

$$White := \{U_1W_1, U_1W_2, U_2W\}.$$

Это можно сделать просто сложив вероятности соответствующих элементарных исходов:

$$\mathbb{P}(White) = 1/6 + 1/6 + 1/4 = 7/12.$$

Однако, на практике все хорошо понимают, [11] что условия (3.5) и (3.6) однозначно задают вероятностное распределение, поэтому вопрос "существования" не возникает, и вероятностное пространство обычно явно не строят. Вместо этого напрямую вычисляют вероятность требуемого события, используя формулу полной вероятности:

$$\mathbb{P}(White) = \mathbb{P}(White|U_1)\mathbb{P}(U_1) + \mathbb{P}(White|U_2)\mathbb{P}(U_2) = 2/3 \cdot 1/2 + 1/2 \cdot 1/2 = 7/12.$$

Формула Байеса

Пусть события $A, B \in \mathcal{F}$ имеют положительные вероятности. Тогда из формулы умножения вероятностей тривиальным образом следует *формула Байеса*:

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)}. \tag{3.7}$$

Она имеет следующее развитие.

Предложение 3.10 (Теорема Байеса)**.** *Если события $H_1, \dots, H_n$ задают разбиение множества A и $\mathbb{P}(H_j) > 0\,\forall j$, $\mathbb{P}(A) > 0$, то*

$$\mathbb{P}(H_k|A) = \frac{\mathbb{P}(A|H_k)\mathbb{P}(H_k)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A|H_k)\mathbb{P}(H_k)}{\sum_{j=1}^{n}\mathbb{P}(A|H_j)\mathbb{P}(H_j)}, \quad \forall k.$$

Доказательство: Первое равенство есть в точности формула Байеса (3.7). Чтобы получить второе равенство, мы применили формулу полной вероятности к знаменателю $\mathbb{P}(A)$. □

Несмотря на возможное ощущение пустого жонглирования формулами, теорема Байеса имеет глубокую интерпретацию. Имеется n взаимоисключающих гипотез $H_1, \dots, H_n$. Мы знаем их *априорные*[12] вероятности $\mathbb{P}(H_i)$. Мы проводим эксперимент и наблюдаем событие A. Спрашивается: какова *апостериорная*[13] вероятность $\mathbb{P}(H_i|A)$? Теорема Байеса позволяет найти эти апостериорные вероятности, если дополнительно известны вероятности $\mathbb{P}(A|H_i)$ наблюдать событие A при условии верности гипотез H_i.

[11] "Блажен, кто верует"

[12] "Априорный" означает "до проведения эксперимента" .

[13] "Апостериорный" значит "после проведения эксперимента" , когда мы получили дополнительные данные — в нашем случае они состоят в том, что произошло событие A.

Пример 3.11. Пусть в задаче из примера 3.9 известно, что вы вытащили белый шар. Какова вероятность того, что вы выбрали первую урну?

Роль гипотез здесь играют события U_1 и U_2, их вероятности нам известны по условию (3.5). Условные вероятности наблюдать событие $White$, состоящее в том, что вы вытянули белый шар, нам известны из (3.6):

$$\mathbb{P}(White|U_1) = 2/3, \quad \mathbb{P}(White|U_2) = 1/2.$$

Тогда, согласно формуле Байеса,

$$\mathbb{P}(U_1|White) = \frac{2/3 \cdot 1/2}{2/3 \cdot 1/2 + 1/2 \cdot 1/2} = 4/7.$$

3.3 Независимость

Независимость — ключевое понятие теории вероятностей, выделяющее ее из абстрактной теории меры: в теории меры такого понятия не существует, потому что отсутствует естественная мотивация его вводить.

Как и прежде, мы живем на вероятностном пространстве $(\Omega, \mathcal{F}, \mathbb{P})$.

Определение 3.12. События $A, B \in \mathcal{F}$ называются *независимыми*, если

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B). \tag{3.8}$$

Легко видеть, что в случае, когда $\mathbb{P}(B) \neq 0$, определение (3.8) эквивалентно соотношению

$$\mathbb{P}(A|B) = \mathbb{P}(A), \tag{3.9}$$

означающему, что от наложения условия B вероятность события A не меняется. Это наблюдение можно рассматривать как мотивировку определения 3.12. Несмотря на более прозрачную интерпретацию формулы (3.9) в сравнении с формулой (3.8), в качестве определения обычно используют последнюю, потому что она удобнее для непосредственной проверки, и в ней отсутствует ограничение $\mathbb{P}(B) \neq 0$. Тут же заметим, что

Задача 3.13. Если $\mathbb{P}(B) = 0$, то событие B независимо с любым событием $A \in \mathcal{F}$.

Сформулируем в виде задачи следующую несложную, но очень полезную лемму, хорошо согласующуюся с нашей интуицией касательно понятия "независимость" .

Задача 3.14. Пусть события $A, B \in \mathcal{F}$ независимы. Докажите, что произвольные события $C \in \{A, A^c\}$ и $D \in \{B, B^c\}$ также независимы.

Определение 3.12 обобщается на случай нескольких событий следующим образом:

Определение 3.15. События $A_1, \dots, A_n$ называются *независимыми в совокупности*, если для всех $2 \leq k \leq n$ и $1 \leq i_1 < \cdots < i_k \leq n$,

$$\mathbb{P}(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) = \mathbb{P}(A_{i_1})\mathbb{P}(A_{i_2}) \cdots \mathbb{P}(A_{i_k}).$$

При этом говорят, что события $A_1, \dots, A_n$ *попарно независимы*, если A_i и A_j независимы для любых $i \neq j$. Из независимости в совокупности очевидно следует попарная независимость. Обратное неверно:

Задача 3.16. Приведите пример вероятностного пространства $(\Omega, \mathcal{F}, \mathbb{P})$ и трех событий $A, B, C \in \mathcal{F}$, независимых попарно, но не в совокупности.

Подсказка: рассмотрите бросок двух различимых монеток с классическим определением вероятности. Придумайте подходящие события, каждое из которых состоит из двух элементарных исходов.

Когда пишут фразу типа "события $A_1, \dots, A_n$ независимы" , как правило имеют ввиду именно независимость в совокупности, хотя корректно задать уточняющий вопрос.

Понятие независимости в совокупности и попарной независимости естественным образом расширяется на случай счетного числа событий. Точнее, говорят, что события $A_1, A_2, \dots$ независимы в совокупности, если для любого n события $A_1, \dots, A_n$ независимы в совокупности в смысле определения 3.15. Заметим, что в этим случае непрерывность меры влечет

$$\mathbb{P}(A_{i_1} \cap A_{i_2} \cap \dots) = \prod_{j=1}^{\infty} \mathbb{P}(A_{i_j}).$$

Пример 3.17. В примере 2.3 о подбрасывании двух различимых кубиков с классическим определением вероятности зададимся вопросом о независимости события $A = \{$сумма результатов бросков кратна 3$\}$ и $B = \{$сумма результатов бросков четна$\}$. Вероятность события A была вычислена в примере 2.9 и составляет $\mathbb{P}(A) = 1/3$. Очевидно, $\mathbb{P}(B) = 1/2$. С другой стороны,

$$\mathbb{P}(A \cap B) = \mathbb{P}\{(1,5),(2,4),(5,1),(4,2),(3,3),(6,6)\} = \frac{1}{6}.$$

Так как $1/3 * 1/2 = 1/6$, события A и B независимы.

Пример 3.18. Являются ли независимыми различные броски в примере 2.1, 2.8 о подбрасывании симметричной монеты $n \geq 1$ раз (эквивалентно, о подбрасывании n различимых симметричных монет)? Интуиция говорит "да" , но заданный вопрос нужно как-то трансформировать в математический. Следующая интерпретация кажется разумной: являются ли независимыми в совокупности произвольные события $A_1, \dots, A_n$ вида

$$A_i = \{\omega = (\omega_1 \dots, \omega_n) \in \Omega : \omega_i \in B_i\}, \quad \text{где} \quad B_i \subset \{\text{"О"} , \text{"Р"} \}?$$

Ответ на этот вопрос оказывается положительным, но мы убедимся в этом только в частном случае, проверив независимость событий $A_1 = \{\omega_1 = \text{"О"} \}$ и $A_2 = \{\omega_2 = \text{"Р"} \}$, состоящих в том, что при первом броске выпал орел, а при втором – решка. Действительно,

$$\mathbb{P}(A_1 \cap A_2) = \mathbb{P}\{\omega_1 = \text{"О"} ,\ \omega_2 = \text{"Р"} ,\ (\omega_i)_{i \geq 3} \text{ произвольны}\} = \frac{2^{n-2}}{2^n} = \frac{1}{4},$$

и рассуждая аналогично находим, что $\mathbb{P}(A_1) = \mathbb{P}(A_2) = 1/2$.

Полная же проверка заявленной независимости в совокупности кажется громоздкой и одна мысль о ней удручает. Но оказывается, ее и не нужно проверять, потому что, как мы покажем в следующем параграфе, такая независимость гарантирована самой структурой вероятностного пространства, представляющегося в виде прямого произведения.

3.4 Вероятностная модель для независимых испытаний

Случай конечного числа экспериментов

Пусть у нас имеется n экспериментов, то есть n вероятностных пространств $(\Omega_i, \mathcal{F}_i, \mathbb{P}_i)$. Тогда мы можем отвечать на вопросы типа "с какой вероятностью в третьем эксперименте произойдет событие $A_3 \in \mathcal{F}_3$?" Ответ дается вероятностью $\mathbb{P}_3(A_3)$. А как нам ответить на вопрос "с какой вероятностью в первом эксперименте произойдет событие A_1, во втором — A_2, и так далее, а в последнем — A_n, если мы считаем, что эксперименты проводятся независимо?" Для этого необходимо построить новое вероятностное пространство $(\Omega, \mathcal{F}, \mathbb{P})$, на котором можно было бы "поселить" все n экспериментов независимым образом. То есть так, чтобы события B_i вида "в i-ом эксперименте произошло событие $A_i \in \mathcal{F}_i$" должны быть независимы в совокупности.

Предлагается это делать следующим образом. Возьмем в качестве $(\Omega, \mathcal{F}, \mathbb{P})$ прямое произведение вероятностных пространств $(\Omega_i, \mathcal{F}_i, \mathbb{P}_i)$, см. аппендикс B.3. Тогда упомянутые выше события B_i принимают вид $B_i = \Omega_1 \times \cdots \times \Omega_{i-1} \times A_i \times \Omega_{i+1} \times \cdots \times \Omega_n$, или, записывая более подробно,

$$B_i = \{\omega = (\omega_1, \ldots, \omega_n) \in \Omega : \omega_i \in A_i\}, \qquad A_i \in \mathcal{F}_i.$$

Следующая лемма утверждает, что предложенная конструкция дает именно то, что мы и хотели.

Лемма 3.19. *Для произвольных $A_i \in \mathcal{F}_i$, $1 \leq i \leq n$, события $B_1, \ldots, B_n$ независимы в совокупности и $\mathbb{P}(B_i) = \mathbb{P}_i(A_i)$.*

Доказательство. Равенство $\mathbb{P}(B_i) = \mathbb{P}_i(A_i)$ немедленно следует из определения вероятности $\mathbb{P}$:

$$\mathbb{P}(B_i) = \mathbb{P}_1(\Omega_1) \cdots \mathbb{P}_{i-1}(\Omega_{i-1}) \mathbb{P}_i(A_i) \mathbb{P}_{i+1}(\Omega_{i+1}) \cdots \mathbb{P}_n(\Omega_n) = \mathbb{P}_i(A_i).$$

Независимость проверяется аналогично:

$$\mathbb{P}(B_{i_1} \cdots B_{i_k}) = \mathbb{P}(\omega : \omega_{i_1} \in A_{i_1}, \ldots, \omega_{i_k} \in A_{i_k}) = \mathbb{P}_{i_1}(A_{i_1}) \cdots \mathbb{P}_{i_n}(A_{i_n}).$$

□

Пример 3.20. Вероятностное пространство, описывающее исход однократного подбрасывания несимметричной монетки с вероятностью выпадения орла равной $0 \leq p \leq 1$, разумно выбрать в виде $(\hat{\Omega}, \hat{\mathcal{F}}, \hat{\mathbb{P}})$, где $\hat{\Omega} = \{0, 1\}$, $\hat{\mathcal{F}} = 2^{\hat{\Omega}}$ и $\hat{\mathbb{P}}\{1\} = p$, $\hat{\mathbb{P}}\{0\} = 1 - p$. Тогда конструкция, предложенная выше, позволяет нам, без всяких дополнительных проверок и вычислений, предъявить вероятностное пространство $(\Omega, \mathcal{F}, \mathbb{P})$, описывающее $n \geq 1$ независимых подбрасываний такой монетки. Оно задается прямым произведением n копий вероятностного пространства $(\hat{\Omega}, \hat{\mathcal{F}}, \hat{\mathbb{P}})$, так что мы получаем

$$\Omega = \{\omega = (\omega_1, \ldots, \omega_n) : \omega_i \in \hat{\Omega}\}, \quad \mathcal{F} = \sigma(\hat{\mathcal{F}} \times \cdots \times \hat{\mathcal{F}}) = 2^{\Omega},$$

и

$$\mathbb{P}(\omega) = \hat{\mathbb{P}}(\omega_1) \cdots \hat{\mathbb{P}}(\omega_n) = p^{\sum_{i=1}^n \omega_i} (1-p)^{n - \sum_{i=1}^n \omega_i}.$$

Случай бесконечного числа экспериментов

Ключевым разделом теории вероятностей, которого мы скоро коснемся, являются так называемые предельные теоремы; например, закон больших чисел и Центральная предельная теорема. В частности, они освещают вопрос "можно ли что-нибудь сказать о статистике результатов большого числа независимых экспериментов, зная лишь некоторые характеристики распределений, описывающих эти эксперименты?" Математическая формулировка подобных вопросов требует взятия предела "число экспериментов n стремится к бесконечности" . Для этого нужно, чтобы вероятностное пространство $(\Omega, \mathcal{F}, \mathbb{P})$, на котором живут наши n экспериментов, не зависело от n. Это требует построения вероятностного пространства, на которое влезает счетное число независимых экспериментов.

В прошлом разделе мы построили подобное вероятностное пространство для произвольного фиксированного конечного числа экспериментов n, но, к сожалению, напрямую предложенная конструкция на случай счетного числа экспериментов не обобщается.

Действительно, можно увидеть, что счетное прямое произведение $\mathcal{F}_1 \times \mathcal{F}_2 \times \dots$ уже не является полукольцом множеств, так что теорема Каратеодори неприменима (см. аппендикс B.3).

Эта проблема имеет элегантное и естественное решение, которое позволяет строить вероятностное пространство, "поддерживающее" не только лишь счетное, но совершенно любое число экспериментов. Итак, пусть имеется набор вероятностных пространств $(\Omega_t, \mathcal{F}_t, \mathbb{P}_t)_{t\in\mathcal{T}}$, описывающих эксперименты, параметризованные произвольным множеством $\mathcal{T}$. Например, в случае $\mathcal{T} = [0, \infty)$, (непрерывный) параметр t можно интерпретировать как время, и думать, что в каждый момент времени совершается некоторый случайный эксперимент. [14]

Множество элементарных исходов Ω выберем, как и прежде, в виде

$$\Omega = \bigotimes_{t\in\mathcal{T}} \Omega_t.$$

Подмножества $C \subset \Omega$ вида

$$C = \{\omega = (\omega_t)_{t\in\mathcal{T}} : \omega_{t_1} \in A_1, \dots, \omega_{t_k} \in A_k\}, \quad \text{где} \quad k \in \mathbb{N}, \quad t_j \in \mathcal{T}, \quad A_j \in \mathcal{F}_{t_j} \quad \forall j \le k, \tag{3.10}$$

будем называть *цилиндрическими*. Менее формально, цилиндрические множества — это такие, которые несут информацию об исходе лишь *конечного* числа экспериментов, при этом не важно каким был исход остальных.

Обозначим символом $\mathcal{C}$ набор всех цилиндрических множеств. Если $|\mathcal{T}| < \infty$, то $\mathcal{C}$ совпадает с прямым произведением сигма-алгебр $\otimes_{t\in\mathcal{T}}\mathcal{F}_t$, однако в общем случае это неверно. Преимущество набора множеств $\mathcal{C}$ перед прямым произведением $\otimes_{t\in\mathcal{T}}\mathcal{F}_t$ видно из следующей задачи.

Задача 3.21. Докажите, что набор всех цилиндрических множеств образует полукольцо.

Определим функцию множеств $\mathbb{P}$ на полукольце $\mathcal{C}$ по формуле

$$\mathbb{P}(C) = \prod_{j=1}^{k} \mathbb{P}_{t_j}(A_j),$$

если $C \in \mathcal{C}$ имеет вид (3.10). Нетрудно видеть, что $\mathbb{P}$ сигма-аддитивна на $\mathcal{C}$. То есть, если непересекающиеся множества $C_1, C_2, \dots \in \mathcal{C}$ удовлетворяют $C := \cup_{i=1}^{\infty} C_i \in \mathcal{C}$, то $\mathbb{P}(C) = \sum_{i=1}^{\infty} \mathbb{P}(C_i)$.

Определение 3.22. Наименьшая σ-алгебра $\sigma(\mathcal{C})$, содержащая всевозможные цилиндрические подмножества, называется *цилиндрической σ-алгеброй.*

Обозначим $\mathcal{F} := \sigma(\mathcal{C})$. По теореме Каратеодори B.13, $\mathbb{P}$ однозначно продолжается до меры на измеримом пространстве $(\Omega, \mathcal{F})$ (которую мы по-прежнему будем обозначать $\mathbb{P}$).

Тройка $(\Omega, \mathcal{F}, \mathbb{P})$ и дает желаемое вероятностное пространство. Действительно, аналогично предыдущему параграфу положим для $s \in \mathcal{T}$

$$B_s = \{\omega = (\omega_t)_{t\in\mathcal{T}} \in \Omega : \omega_s \in A_s\}, \qquad A_s \in \mathcal{F}_s.$$

Лемма 3.23. *Для произвольных $n \ge 2$ и $A_{s_i} \in \mathcal{F}_{s_i}$, $1 \le i \le n$, события $B_{s_1}, \dots, B_{s_n}$ независимы в совокупности и $\mathbb{P}(B_{s_i}) = \mathbb{P}_{s_i}(A_{s_i})$.*

[14] Развитие этой идеи ведет к понятию "случайный процесс" .

Доказательство повторяет доказательство леммы 3.19.

Хотя, формально говоря, мы и поселили на одно вероятностное пространство произвольное число экспериментов, в конструкцию заложена некоторая априорная счетность – см. определение цилиндрической сигма-алгебры. Последняя лемма вообще рассматривает лишь конечные наборы множеств, но это связано с тем, что определение независимости в совокупности давалось лишь для конечного числа событий, и можно написать аналог леммы и для счетного числа событий. Действительно, рассмотрим событие вида

$$C = \{\omega = (\omega_t)_{t\in\mathcal{T}} \in \Omega : \omega_{s_1} \in A_{s_1}, \omega_{s_2} \in A_{s_2}, \dots\},$$

для счетного набора событий $A_{s_i} \in \mathcal{F}_i$. Ясно, что оно лежит в цилиндрической сигма-алгебре $\mathcal{F}$, и, используя непрерывность меры, мы находим $\mathbb{P}(C) = \prod_{i=1}^{\infty} \mathbb{P}_{s_i}(A_{s_i})$. Однако, событие

$$C' = \{\omega = (\omega_t)_{t\in\mathcal{T}} \in \Omega : \omega_s \in A_s \quad \text{для всех} \quad s \in \mathcal{T}'\}$$

для множества $\mathcal{T}'$ мощностью континуум, уже не лежит в цилиндрической сигма-алгебре $\mathcal{F}$, так что мы не можем вычислить его вероятность.

3.5 Леммы Бореля-Кантелли

В этом параграфе мы докажем два утверждения, называемые леммами Бореля-Кантелли, отвечающими на вопрос типа "с какой вероятностью в последовательности экспериментов некоторое событие произойдет бесконечное число раз?"

Пусть, как всегда, $(\Omega, \mathcal{F}, \mathbb{P})$ — вероятностное пространство, а $A_i \in \mathcal{F}$, $i \geq 1$, — последовательность событий на нем. Рассмотрим событие

$$A_\infty = \{\omega \in \Omega : \omega \in A_i \quad \text{для бесконечного числа событий } A_i\} = \bigcap_{n=1}^{\infty} \bigcup_{i=n}^{\infty} A_i.$$

Лемма 3.24 (лемма Бореля-Кантелли)**.** *i) Если $\sum_{i=1}^{\infty} \mathbb{P}(A_i) < \infty$, то $\mathbb{P}(A_\infty) = 0$.*
ii) Если $\sum_{i=1}^{\infty} \mathbb{P}(A_i) = \infty$ и события A_i независимы в совокупности, то $\mathbb{P}(A_\infty) = 1$.

Доказательство. i) Используя вложенность событий $B_n := \bigcup_{i=n}^{\infty} A_i$ и непрерывность вероятности, находим

$$\mathbb{P}(A_\infty) = \lim_{n\to\infty} \mathbb{P}(B_n) \leq \lim_{n\to\infty} \sum_{i=n}^{\infty} \mathbb{P}(A_i) = 0,$$

так как ряд $\sum_{i=1}^{\infty} \mathbb{P}(A_i)$ сходится.

ii) Заметим, что $A_\infty^c = \bigcup_{n=1}^{\infty} B_n^c$, где $B_n^c = \bigcap_{i=n}^{\infty} A_i^c$. Тогда

$$\mathbb{P}(A_\infty^c) \leq \sum_{n=1}^{\infty} \mathbb{P}(B_n^c).$$

Используя независимость событий A_i^c, находим

$$\mathbb{P}(B_n^c) = \prod_{i=n}^{\infty} \mathbb{P}(A_i^c) = \prod_{i=n}^{\infty} (1 - \mathbb{P}(A_i)) = 0,$$

так как $\sum_{i=1}^{\infty} \mathbb{P}(A_i) = \infty$.

□

Используя лемму Бореля-Кантелли, нетрудно показать, что если вы набираете случайный текст, то с вероятностью 1 вы напечатаете Войну и Мир бесконечное число раз.

4 Случайные величины

Эксперименты, для которых применяются вероятностные описания, связаны как правило с числовыми измерениями, и основным объектом для вероятностного описания измерений служит *случайная величина*. Ее можно определить двумя эквивалентными способами:

1. Числовая функция ξ на вероятностном пространстве Ω

2. Структура вероятностного пространства на множестве Ω, состоящем из чисел.

Вторая формулировка вытекает из первой: достаточно рассматривать множество чисел вида $\xi(\omega)$, $\omega \in \Omega$ и сопоставить каждому числу вероятность его прообраза в Ω. Первая же формулировка получается из второй при помощи тождественного отображения $\Omega \to \Omega$. При этом следует помнить, что указанные определения используются в дискретном случае без ограничений, но в более общей ситуации требуют уточнения: например, первая формулировка включает требование измеримости функции. Теперь обсудим все это более подбробно.

4.1 Случайные величины и их распределения

Пусть $(\Omega, \mathcal{F}, \mathbb{P})$ — вероятностное пространство.

Определение 4.1. Измеримая функция $\xi : (\Omega, \mathcal{F}) \mapsto (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ называется *случайной величиной.* [15]

Аналогично можно определить случайные величины со значениями в абстрактном измеримом пространстве $(X, \mathcal{G})$ как измеримые функции $\xi : (\Omega, \mathcal{F}) \mapsto (X, \mathcal{G})$. Очевидные примеры: $(X, \mathcal{G}) = (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ и $(X, \mathcal{G}) = (\mathbb{C}^d, \mathcal{B}(\mathbb{C}^d))$. Более сложные примеры возникают в теории случайных процессов, где X часто является каким-либо функциональным пространством (непрерывных функций, соболевским, etc). Чтобы подчеркнуть, что ξ принимает значения в множестве X, иногда говорят, что ξ — X-значная случайная величина.

Однако, в наших лекциях случайные величины почти всегда будут принимать значения либо в $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, либо в $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. В этом последнем случае $\xi = (\xi_1, \dots, \xi_d)$, и мы будем стараться называть ξ *случайным вектором*, либо d-мерной случайной величиной. Ее компоненты ξ_j при этом являются случайными величинами. Но часто мы будем заговариваться и говорить про ξ просто "случайная величина" . Размерность обычно будет понятна из контекста.

Следующее определение основано на понятии образа меры при отображении, см. аппендикс B.4.

Определение 4.2. Вероятностная мера $\mathbb{P}_\xi := \xi_*(\mathbb{P})$ на измеримом пространстве $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ называется *распределением случайной величины* ξ.

Другими словами, $\mathbb{P}_\xi$ — такая вероятностная мера на $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, что

$$\mathbb{P}_\xi(A) = \mathbb{P}\big\{\omega \in \Omega : \xi(\omega) \in A\big\}, \quad \forall A \in \mathcal{B}(\mathbb{R}).$$

Обычно запись $\mathbb{P}\big\{\omega \in \Omega : \xi(\omega) \in A\big\}$ сокращают до $\mathbb{P}\big\{\xi \in A\big\}$ или $\mathbb{P}\big(\xi \in A\big)$. Мы тоже будем следовать этой славной традиции.

Аналогичным образом определяют распределение случайного вектора $\xi = (\xi_1, \dots, \xi_d)$ как меру $\mathbb{P}_\xi := \xi_*(\mathbb{P})$ на $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. В этом случае меру $\mathbb{P}_\xi$ также называют *совместным распределением* случайных величин $\xi_1, \dots, \xi_d$. Аналогичное определение работает и для случайных величин со значениями в абстрактном измеримом пространстве.

[15] В определении случайной величины используется борелевская сигма-алгебра, а не более широкая сигма-алгебра множеств, измеримых по Лебегу, чтобы иметь возможность брать композиции случайных величин, не теряя измеримости.

Пример 4.3. Рассмотрим эксперимент, состоящий из $n \geq 1$ независимых подбрасываний монетки с вероятностью выпадения орла равной p и вероятностную модель для этого эксперимента, предложенную в примере 3.20. Пусть ξ —случайная величина, описывающая число орлов, выпавших за n бросков,

$$\xi(\omega) := \sum_{i=1}^{d} \omega_i.$$

Тогда ее распределение $\mathbb{P}_\xi$ задано соотношением

$$\mathbb{P}_\xi\{k\} = \mathbb{P}\{\xi = k\} = C_n^k p^k (1-p)^{n-k}, \qquad k = 1, 2, \ldots, n.$$

Такое распределение называется *биномиальным* с параметрами n, p.
Пишут $\xi \sim Binomial(n, p)$.

Важное замечание. Когда мы обсуждаем случайные величины, обычно нам интересно отвечать на вопросы типа "с какой вероятностью ξ принимает значения в множестве A?" , или, другими словами, "чему равно $\mathbb{P}_\xi(A)$?" . Поэтому, как правило, единственное, что нам нужно знать о случайной величине, — это ее распределение $\mathbb{P}_\xi$. Конкретная же структура вероятностного пространства $(\Omega, \mathcal{F}, \mathbb{P})$ и явный вид функции $\xi : \Omega \mapsto \mathbb{R}$ бывают важны довольно редко, и обычно при решении задач они не уточняются.

Каждая случайная величина ξ определяет новое вероятностное пространство

$$(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P}_\xi).$$

Зачем же тогда рассматривать двухслойную конструкцию, состоящую из исходного вероятностного пространства $(\Omega, \mathcal{F}, \mathbb{P})$ и пространства $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P}_\xi)$, если можно сразу рассмотреть последнее и получить интересующую нас информацию о вероятностях вида $\mathbb{P}(\xi \in A) = \mathbb{P}_\xi(A)$? Это упростило бы теорию, так как можно было бы не вводить понятие случайной величины. Если нас интересует лишь одна случайная величина ξ, то так сделать действительно можно, но на практике часто оказывается важным иметь возможность ввести на *том же* вероятностном пространстве новую случайную величину ζ с другим интересующим нас распределением (тогда, например, можно будет поточечно вычислять сумму $\xi(\omega) + \zeta(\omega)$). В примере 4.3 это может быть, скажем, результат третьего броска. Вероятностное пространство $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P}_\xi)$ может оказаться слишком бедным для того, чтобы случайная величина ζ была на нем определена. Поэтому предложенная двухслойная конструкция удобна. В ней мы всегда заранее и по умолчанию предполагаем, что исходное вероятностное пространство $(\Omega, \mathcal{F}, \mathbb{P})$ достаточно богато, чтобы на нем были определены все рассматриваемые нами случайные величины и используемые конструкции.

Пример 4.4. Пусть $M \in \mathcal{B}(\mathbb{R}^d)$ — ограниченное борелевское множество в $\mathbb{R}^d$. Рассмотрим случайный вектор ξ со значениями в $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ с распределением

$$\mathbb{P}(\xi \in A) = \mathbb{P}_\xi(A) = \frac{Leb(A \cap M)}{Leb(M)}, \quad \forall A \in \mathcal{B}(\mathbb{R}^d),$$

где Leb обозначает меру Лебега. Такое распределение называют *равномерным* распределением на M, а случайный вектор ξ – *равномерно распределенным* на M. Пишут $\xi \sim Uniform(M)$.

Отметим, что, в соответствии с замечанием выше, мы указали лишь распределение случайной величины, но не выбор вероятностного пространства. Его можно выбирать по-разному, и разным выборам будут соответствовать разные случайные величины. Простейший выбор такой:

$$\Omega = \mathbb{R}^d, \quad \mathcal{F} = \mathcal{B}(\mathbb{R}^d), \quad \mathbb{P}(A) = \frac{Leb(A \cap M)}{Leb(M)} \quad \text{и} \quad \xi(\omega) = \omega.$$

Это в точности упомянутое выше вероятностное пространство вида $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \mathbb{P}_\xi)$.

4.2 Типы случайных величин и векторов

Выделяют три типа случайных величин и векторов, классификация происходит из классификации их распределений, см. аппендикс B.5. А именно, случайную величину (или вектор) ξ называют:

• *Дискретной*, если ее распределение $\mathbb{P}_\xi$ является дискретной мерой. То есть, существует не более, чем счетное множество $X \subset \mathbb{R}^d$, такое что $\mathbb{P}(\xi \in X) = 1$. Это близко к тому, чтобы сказать, что случайная величина принимает не более, чем счетное число значений, но не совсем то. Точки $a \in \mathbb{R}^d$, такие что $\mathbb{P}_\xi(a) = \mathbb{P}\{\xi = a\} > 0$, называют *атомами меры* $\mathbb{P}_\xi$.

• *Сингулярной*, если ее распределение является сингулярной непрерывной мерой. То есть, $\mathbb{P}(\xi = x) = 0$ для всех $x \in \mathbb{R}^d$, но существует множество $X \in \mathcal{B}(\mathbb{R}^d)$ нулевой меры Лебега, такое что $\mathbb{P}(\xi \in X) = 1$. Такие случайные величины довольно экзотичны.

• *абсолютно непрерывной*, если ее распределение абсолютно непрерывно. То есть, случайная величина ξ (а точнее, ее распределение $\mathbb{P}_\xi$) имеет плотность p_ξ — неотрицательную интегрируемую функцию, такую что

$$\mathbb{P}(\xi \in A) = \int_A p_\xi(x)\, dx, \quad \forall A \in \mathcal{B}(\mathbb{R}^d).$$

В частности, $\int_{\mathbb{R}^d} \rho_\xi(x)\, dx = 1$.

Замечание 4.5. Обратите внимание на небольшое несоответствие в терминологии между мерами и случайными величинами: меру называют сингулярной, если она либо дискретна, либо непрерывна сингулярна. А случайную величину называют сингулярной, если ее распределение только лишь непрерывно сингулярно.

Сингулярные и абсолютно непрерывные случайные величины ξ вместе называют *непрерывными*. Их характеризует равенство $\mathbb{P}(\xi = x) = 0$ для всех $x \in \mathbb{R}^n$.

Кроме указанных выше трех типов случайных величин, разумеется, бывают случайные величины смешанных типов. По теореме B.19 о разложении меры распределение произвольной случайной величины ξ однозначно представляется в виде

$$\mathbb{P}_\xi = c_d \nu_d + c_s \nu_s + c_a \nu_a,$$

где ν_d, ν_s и ν_a — дискретная, непрерывная сингулярная и абсолютно непрерывная вероятностные меры, а c_d, c_s, c_a — неотрицательные константы (появляющиеся из нормировки мер μ_d, μ_s, μ_a из теоремы B.19, чтобы получить вероятностные меры $\nu_{d,s,a}$: $\mu_a = c_a \nu_a$, и аналогично для $\nu_{d,s}$).

Замечание 4.6. Аналогично обсуждению ниже примера 4.3, можно построить вероятностное пространство и случайные величины ξ_d, ξ_s, ξ_a на нем, так, чтобы их распределения совпадали с мерами ν_d, ν_s, ν_a. Разумеется, как правило $\xi \neq c_d \xi_d + c_s \xi_s + c_a \xi_a$. Более того, чтобы вычислить распределение суммы случайных величин, нужно знать их совместное распределение.

В дальнейшем нам будет полезно следующее следствие теоремы Фубини.

Лемма 4.7. *Пусть* $\xi = (\xi_1, \dots, \xi_d)$ *— абсолютно непрерывный случайный вектор с плотностью* p_ξ*. Тогда его компоненты* ξ_j *тоже являются абсолютно непрерывными случайными величинами, а их плотность имеет вид*

$$p_{\xi_j}(x) = \int_{\mathbb{R}^{n-1}} p_\xi(x_1, \dots, x_{j-1}, x, x_{j+1}, \dots, x_n)\, dx_1 \cdots dx_{j-1} dx_{j+1} \cdots dx_n, \qquad x \in \mathbb{R}. \quad (4.1)$$

Доказательство. Для удобства обозначений проведем доказательство для $j = 1$. Для любого $A \in \mathcal{B}(\mathbb{R}^n)$,

$$\mathbb{P}(\xi_1 \in A) = \int_{A \times \mathbb{R}^{n-1}} p_\xi(x_1, \dots x_n)\, dx_1 \cdots dx_n = \int_A p_{\xi_1}(x_1)\, dx_1,$$

где в последнем равенстве мы использовали теорему Фубини, а p_{ξ_1} определено в (4.1). $\square$

Задача 4.8. Покажите, что обратное утверждение не верно: из абсолютной непрерывности компонент случайного вектора не следует абсолютная непрерывность самого вектора.

Пример 4.9. Биномиальная случайная величина $\xi \sim Binomial(n, p)$ дискретна. Ее распределение имеет вид $\mathbb{P}_\xi = \sum_{k=0}^n \mathbb{P}(\xi = k)\delta_k = C_n^k p^k q^{n-k} \delta_k$, где $q = 1 - p$, а δ_k обозначает дельта-меру в точке k (см. (B.3)). Последнее равенство удобнее записывать в виде таблички:

ξ	0	$\dots$	k	$\dots$	n
$\mathbb{P}_\xi$	$\binom{n}{0} q^n$	$\dots$	$\binom{n}{k} p^k q^{n-k}$	$\dots$	$\binom{n}{n} p^n$

Пример 4.10. Распределение дискретного случайного вектора (α, β) также удобно записывать в виде таблички. Например,

$\beta \setminus \alpha$	-1	0	$+1$
-1	$1/8$	$1/12$	$1/8$
0	$1/12$	$1/6$	$1/12$
$+1$	$1/8$	$1/12$	$1/8$

В частности, здесь написано, что $\mathbb{P}(\alpha = 0, \beta = 1) = 1/12$.

Пример 4.11. Равномерно распределенная на $M \in \mathcal{B}(\mathbb{R}^d)$ случайная величина $\xi \sim Uniform(M)$ абсолютно непрерывна. Ее плотность имеет вид

$$p_\xi(x) \equiv \frac{1}{|M|} \mathbb{I}_M(x),$$

где $\mathbb{I}_M$ обозначает индикаторную функцию множества M, а $|M|$ — его меру Лебега. Обычно последнее равенство и приводится в качестве определения равномерного распределения. В частности, если $\xi \sim Uniform([0, 1])$, то $p_\xi(x) = \mathbb{I}_{[0,1]}(x)$.

Пример сингулярной случайной величины мы дадим ниже, после того как определим, что такое функции распределения.

4.3 Функции распределения

Обычно самый практичный способ работать с распределением случайной величины — использовать табличку, если случайная величина дискретна, и использовать плотность, если она абсолютно непрерывна. Эти способы не универсальны, так как, помимо экзотических сингулярных случайных величин, бывают еще и случайные величины смешанного типа. Это одна из многих причин, из-за которой удобно ввести следующий универсальный инструмент, однако же хорошо работающий только в размерности $d = 1$.

Итак, почти до конца этого параграфа мы рассматриваем случайную величину ξ со значениями лишь в $\mathbb{R}$.

Определение 4.12. Функция $F_\xi : \mathbb{R} \mapsto [0,1]$, определенная равенством

$$F_\xi(x) := \mathbb{P}(\xi \le x),$$

называется *функцией распределения случайной величины* ξ.

Эквивалентно, $F_\xi(x) = \mathbb{P}_\xi\big((-\infty, x]\big)$.

Лемма 4.13. *Функция распределения F_ξ однозначно определяет распределение $\mathbb{P}_\xi$. То есть, из равенства $F_\xi = F_\eta$ следует $\mathbb{P}_\xi = \mathbb{P}_\eta$.*

Доказательство. Полуинтервалы вида $(-\infty, x]$ образуют полукольцо множеств. Поэтому утверждение леммы следует из теоремы Каратеодори. □

Перечислим некоторые свойства функции распределения.

Лемма 4.14. *Функция распределения F_ξ случайной величины ξ удовлетворяет следующим свойствам:*
1) F_ξ не убывает, то есть $F_\xi(x) \le F_\xi(y)$ при $x \le y$.
2) $\lim\limits_{x\to-\infty} F_\xi(x) = 0$ и $\lim\limits_{x\to\infty} F_\xi(x) = 1$.
3) F_ξ непрерывна справа, то есть $\lim\limits_{y\to x+} F_\xi(y) = F_\xi(x)$.

Доказательство. Первое свойство следует из того, что событие $\{\omega : \xi(\omega) \le x\}$ содержится в событии $\{\omega : \xi(\omega) \le y\}$ при $x \le y$. Второе и третье свойства следуют из непрерывности меры. Действительно, пересечение событий $\{\omega : \xi(\omega) < x\}$ по всем $x < 0$ пусто, а значит

$$\lim_{x\to-\infty} F_\xi(x) = \lim_{x\to-\infty} \mathbb{P}(\xi \le x) = 0,$$

согласно лемме B.9(4). Так как $F_\xi(x) = 1 - \mathbb{P}(\xi > x)$, предел $\lim\limits_{x\to\infty} F_\xi(x)$ находится аналогично, что влечет пункт 2 леммы. Пункт 3 аналогичным образом следует из леммы B.9(3). □

Итак, к настоящему моменту мы увидели, что функция распределения F_ξ случайной величины ξ однозначно определяет ее распределение $\mathbb{P}_\xi$ и обладает свойствами 1-3, перечисленными в лемме. Оказывается, верно и обратное, в следующем смысле.

Всякую функцию $F : \mathbb{R} \mapsto [0,1]$, удовлетворяющую свойствам 1-3 леммы 4.14, будем называеть *функцией распределения*. Рассмотрим функцию множеств μ, определенную на всех полуинтервалах вида $(x, y]$, где $x \ge -\infty$, $y \in \mathbb{R}$, по правилу

$$\mu\big((x,y]\big) = F(y) - F(x), \qquad F(-\infty) := 0. \tag{4.2}$$

Напомним, что множество таких полуинтервалов образует полукольцо. Можно показать, что функция множеств μ — счетно аддитивна, см. [KS, §3.2]. Тогда, по теореме Каратеодори, она единственным образом продолжается до меры μ на минимальной сигма-алгебре, порожденной рассматриваемым полукольцом. А последняя и является борелевской сигма-алгеброй $\mathcal{B}(\mathbb{R})$. Для иллюстрации отметим, что полученная мера удовлетворяет соотношениям

$$\mu([a,b]) = F(b) - \lim_{x\to a-} F(x), \qquad \mu((a,b)) = \lim_{x\to b-} F(x) - F(a).$$

В частности, $\mu(\mathbb{R}) = 1$, так что полученная мера является вероятностной.

Рассмотрим теперь вероятностное пространство $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu)$ и случайную величину на нем $\xi(\omega) \equiv \omega \ \forall \omega \in \mathbb{R}$. Тогда, согласно (4.2),

$$F_\xi(x) = \mu(\xi \le x) = F(x).$$

Таким образом, мы построили случайную величину ξ с функцией распределения F_ξ, равной F. Заметим, что распределение $\mathbb{P}_\xi$ (мера μ выше) определялось однозначно, а вот случайная величина ξ, конечно, не однозначно (например, можно было выбрать какое-нибудь другое вероятностное пространство).

Функции распределения различных типов случайных величин

Рассмотрим случайную величину ξ. Как мы обсудили выше, между распределениями случайных величин $\mathbb{P}_\xi$ на $\mathbb{R}$ (являющимися вероятностными мерами) и функциями распределения F_ξ существует взаимно однозначное соответствие. Следовательно, можно выделить классы функций распределения, соответствующие дискретным, сингулярным непрерывным и абсолютно непрерывным распределениям. Сейчас мы этим и займемся, однако прежде докажем следующую общую лемму.

Лемма 4.15. *Множество точек разрыва функции распределения F_ξ случайной величины ξ не более чем счетно. Оно состоит из всех точек $x \in \mathbb{R}$, в которых $\mathbb{P}(\xi = x) > 0$.*

Доказательство. Первое утверждение — хорошо известное свойство монотонных функций. Для обоснования второго утверждения достаточно рассматривать точки разрыва слева, потому что функция распределения непрерывна справа, согласно лемме 4.14(3),. Тогда желаемое утверждение следует из соотношения

$$\mathbb{P}(\xi = x) = \lim_{\varepsilon \to 0+} \mathbb{P}(x - \varepsilon < \xi \le x) = F_\xi(x) - \lim_{\varepsilon \to 0+} F_\xi(x - \varepsilon),$$

где в первом равенстве мы использовали непрерывность вероятности. □

• *Дискретные случайные величины*

Лемма 4.16. *Случайная величина ξ дискретна в том и только том случае, когда ее функция распределения F_ξ имеет ступенчатый вид.*

Доказательство. Если случайная величина дискретна, то утверждение очевидно: функция распределения F_ξ имеет ступенчатый вид со скачками в точках $x \in \mathbb{R}$, для которых $\mathbb{P}(\xi = x) > 0$. Высота скачка в точности равна $\mathbb{P}(\xi = x)$.

Обратное следует из того, что функция распределения F_ξ, как и всякая монотонная функция, не может иметь более, чем счетное число скачков. Учитывая ступенчатость F_ξ, отсюда легко видеть, что ξ дискретна. □

• *Абсолютно непрерывные случайные величины*

Напомним, что функция $F(x)$ вида

$$F(x) = \int_{-\infty}^{x} p(y)\, dy, \tag{4.3}$$

где функция p интегрируема, называется *абсолютно непрерывной*. Функция p при этом называется *производной* функции F, что обозначается обычным образом $p = F'$. Такая терминология оправдана: если функция p непрерывна, то, по теореме о дифференцируемости интеграла по верхнему пределу, действительно имеем $p = F'$ в классическом смысле. В общем случае, когда p лишь интегрируема, из курса анализа известно, что функция F дифференцируема почти во всех (по мере Лебега) точках $x \in \mathbb{R}$, и в этих точках верно соотношение $p(x) = F'(x)$.

Лемма 4.17. *Случайная величина ξ абсолютно непрерывна в том и только том случае, когда ее функция распределения F_ξ абсолютно непрерывна. При этом плотность распределения p_ξ является производной функции распределения,*

$$F'_\xi(x) = p_\xi(x). \tag{4.4}$$

Доказательство. Если случайная величина ξ абсолютно непрерывна, то она имеет плотность p_ξ. А значит

$$F_\xi(x) = \mathbb{P}(\xi \le x) = \int_{-\infty}^{x} p_\xi(y)\,dy, \tag{4.5}$$

то есть функция распределения F_ξ абсолютно непрерывна.

Обратно, допустим, что случайная величина ξ имеет абсолютно непрерывную функцию распределения F_ξ, то есть, существует интегрируемая функция p, для которой верно (4.5). Покажем, что в этом случае p является плотностью распределения ξ, так что ξ — абсолютно непрерывна. Действительно, достаточно проверить, что

$$\mathbb{P}(\xi \in A) = \int_A p(y)\,dy \qquad \forall A \in \mathcal{B}(\mathbb{R}).$$

Ввиду однозначности продолжения меры, утверждаемой теоремой Каратеодори, достаточно проверить последнее равенство для множеств вида $A = (a, b]$. А для них имеем

$$\mathbb{P}((a,b]) = \mathbb{P}(\xi \le b) - \mathbb{P}(\xi \le a) = F_\xi(b) - F_\xi(a) = \int_a^b p(y)\,dy,$$

что и требовалось. □

Пример 4.18. Пусть $\xi \sim Uniform([a,b])$. Тогда

$$F_\xi(x) = \begin{cases} 0, & \text{если } x \le a, \\ \frac{x-a}{b-a}, & \text{если } a < x \le b, \\ 1, & \text{если } x > b. \end{cases}$$

- *Сингулярные случайные величины.*

Согласно лемме 4.15, функция распределения сингулярной случайной величины непрерывна. Однако, она не может быть представлена в виде интеграла от плотности (4.3), а множество ее точек роста имеют нулевую меру Лебега. Мы не будем останавливаться на этом экзотическом случае подробно, а вместо этого рассмотрим пример такой функции распределения: это знаменитая канторова лестница.

Как известно, она непрерывна, множество ее точек роста является канторовым множеством и имеет нулевую меру Лебега. Эта функция почти всюду дифференцируема, но в точках дифференцируемости ее производная равна нулю. При этом в нуле она равна нулю, а в единице — единице, так что для случайной величины ξ с такой функцией распределения имеем $\mathbb{P}(\xi \in [0,1]) = 1$.

Многомерные функции распределения

В многомерном случае функции распределения определяются аналогично, однако с ними гораздо менее удобно работать. Но тем не менее они полезны: например, с их помощью бывает удобно проверять независимость случайных величин, речь об этом пойдет в следующем параграфе.

Пусть $\xi = (\xi_1, \dots, \xi_d)$ — случайный вектор.

Определение 4.19. Функция $F_\xi(x) : \mathbb{R}^d \mapsto [0,1]$, определенная равенством

$$F_\xi(x) = \mathbb{P}\{\xi_1 \le x_1, \dots, \xi_d \le x_d\},$$

где $x = (x_1, \dots, x_d)$, называется *функцией распределения* случайного вектора ξ.

Лемма 4.20. *Функция распределения F_ξ однозначно определяет распределение $\mathbb{P}_\xi$ в смысле леммы 4.13.*

Доказательство этой леммы аналогично доказательству ее одномерного аналога — леммы 4.13. Также имеет место аналог леммы 4.14, однако формулировки в многомерном случае оказываются сложнее, и мы их не приводим. См. [KS, теорема 1.47].

4.4 Независимость случайных величин

Теперь введем следующее ключевое определение.

Определение 4.21. Случайные величины ξ и η, определенные на одном и том же вероятностном пространстве $(\Omega, \mathcal{F}, \mathbb{P})$, называются *независимыми*, если

$$\mathbb{P}(\xi \in A, \eta \in B) = \mathbb{P}(\xi \in A)\mathbb{P}(\eta \in B) \quad \forall\, A, B \in \mathcal{B}(\mathbb{R}). \tag{4.6}$$

Так как равенство (4.6) можно переписать в виде

$$\mathbb{P}(\xi^{-1}(A) \cap \eta^{-1}(B)) = \mathbb{P}(\xi^{-1}(A))\mathbb{P}(\eta^{-1}(B)),$$

то определение независимости эквивалентно звучит так: для любых борелевских множеств $A, B \in \mathcal{B}(\mathbb{R})$ события $\xi^{-1}(A)$ и $\eta^{-1}(B)$ независимы.

Третий способ записать определение независимости — следующий:

$$\mathbb{P}_{(\xi,\eta)}(A \times B) = \mathbb{P}_\xi(A)\mathbb{P}_\eta(B) \quad \forall\, A, B \in \mathcal{B}(\mathbb{R}). \tag{4.7}$$

Определение 4.22. Случайные величины $\xi_1, \dots, \xi_n$ называются *независимыми в совокупности*, если

$$\mathbb{P}(\xi_1 \in A_1, \dots, \xi_n \in A_n) = \mathbb{P}(\xi_1 \in A_1) \cdots \mathbb{P}(\xi_n \in A_n) \quad \forall\, A_j \in \mathcal{B}(\mathbb{R}). \tag{4.8}$$

Заметим, что, в отличие от определения 3.15 независимости в совокупности событий, в определении 4.22 мы не рассматриваем всевозможные подпоследовательности $i_1, \dots, i_k$. Их рассмотрение не было бы ошибкой, но это излишне, так как это покрывается выбором в качестве некоторых A_j всего множества Ω.

Определения выше дословно переносятся на случай случайных векторов заменой $\mathbb{R}$ на $\mathbb{R}^d$, а также случайных величин со значениями в абстрактном измеримом пространстве $(X, \mathcal{G})$.

Предложение 4.23. *Независимость в совокупности случайных величин $\xi_1, \dots, \xi_n$ эквивалентна каждому из следующих тождеств:*
1) $\mathbb{P}_{(\xi_1,\dots,\xi_n)} = \mathbb{P}_{\xi_1} \times \dots \times \mathbb{P}_{\xi_n}$;
2) $F_{(\xi_1,\dots,\xi_n)}(x_1, \dots, x_n) = F_{\xi_1}(x_1) \cdots F_{\xi_n}(x_n)$;
3) в дополнительном предположении, что случайный вектор $\xi = (\xi_1, \dots, \xi_n)$ абсолютно непрерывен:

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

4) в дополнительном предположении, что случайные величины $\xi_1, \dots, \xi_n$ дискретны:

$$\mathbb{P}(\xi_1 = x_1, \dots, \xi_n = x_n) = \mathbb{P}(\xi_1 = x_1) \cdots \mathbb{P}(\xi_n = x_n),$$

для всех $x_j \in \operatorname{supp} \mathbb{P}_{\xi_j}$ (то есть таких x_j, что $\mathbb{P}(\xi_j = x_j) > 0$ — атомов мер $\mathbb{P}_{\xi_j}$).

Доказательство. 1) Из формулы 1) независимость следует по ее определению, аналогично (4.7). Обратно, пусть случайные величины независимы в совокупности. То есть,

$$\mathbb{P}_{(\xi_1,\ldots,\xi_n)}(A_1 \times \cdots \times A_n) = \mathbb{P}_{\xi_1}(A_1) \times \cdots \times \mathbb{P}_{\xi_n}(A_n) \quad \forall A_1, \ldots, A_n \in \mathcal{B}(\mathbb{R}).$$

Теперь требуемый результат следует из единственности продолжения меры с полукольца множеств $\mathcal{B}(\mathbb{R}) \times \cdots \times \mathcal{B}(\mathbb{R})$ на минимальную сигма-алгебру, им порожденную (которая совпадает с $\mathcal{B}(\mathbb{R}^n)$), обеспеченной теоремой Каратеодори.

2) Если случайные величины $\xi_1, \ldots, \xi_n$ независимы в совокупности, то формула 2) следует по определению. Обратно, пусть верна формула 2). Рассмотрим случайный вектор η, имеющий распределение $\mathbb{P}_\eta = \mathbb{P}_{\xi_1} \times \cdots \times \mathbb{P}_{\xi_n}$. Тогда $F_\eta(x) = F_{\xi_1}(x_1) \cdots F_{\xi_n}(x_n)$, а значит, по предположению, $F_\eta = F_{(\xi_1,\ldots,\xi_n)}$. Так как, согласно лемме 4.20, функция распределения однозначно определяет распределение, мы находим $\mathbb{P}_\eta = \mathbb{P}_{(\xi_1,\ldots,\xi_n)}$, то есть имеет место формула 1). Значит, по уже доказанному, случайные величины $\xi_1, \ldots, \xi_n$ независимы.

3) Напомним, что абсолютная непрерывность компонент ξ_j вектора ξ следует из леммы 4.7. Пусть случайные величины $\xi_1, \ldots, \xi_n$ независимы. Тогда для любых событий $A_1, \ldots, A_n \in \mathcal{B}(\mathbb{R})$ из формулы (4.8) следует

$$\int_{A_1 \times \cdots \times A_n} p_{(\xi_1,\ldots,\xi_n)}(x_1, \ldots, x_n)\, dx_1 \ldots dx_n = \int_{A_1} p_{\xi_1}(x_1)\, dx_1 \cdots \int_{A_n} p_{\xi_n}(x_n)\, dx_n.$$

Это влечет равенство 3). Обратно, пусть имеет место равенство 3). Тогда из него сразу следует равенство 2), которое по уже доказанному влечет желаемую независимость.

4) Независимость по определению влечет требуемое равенство. Докажем обратное: для любых $A_1, \ldots, A_n \in \mathcal{B}(\mathbb{R})$

$$\begin{aligned} \mathbb{P}\{\xi_1 \in A_1, \ldots, \xi_n \in A_n\} &= \sum_{x_1 \in A_1 \cap \operatorname{supp} \mathbb{P}_{\xi_1}} \cdots \sum_{x_n \in A_n \cap \operatorname{supp} \mathbb{P}_{\xi_n}} \mathbb{P}(\xi_1 = x_1, \ldots, \xi_n = x_n) \\ &= \sum_{x_1 \in A_1 \cap \operatorname{supp} \mathbb{P}_{\xi_1}} \mathbb{P}(\xi_1 = x_1) \cdots \sum_{x_n \in A_n \cap \operatorname{supp} \mathbb{P}_{\xi_n}} \mathbb{P}(\xi_n = x_n) \\ &= \mathbb{P}\{\xi_1 \in A_1\} \cdots \mathbb{P}\{\xi_n \in A_n\}. \end{aligned}$$

□

Следствие 4.24. *Если известны распределения $\mathbb{P}_{\xi_1}, \ldots, \mathbb{P}_{\xi_n}$ каждой из случайных величин $\xi_1, \ldots, \xi_n$, и известно, что случайные величины независимы, то известно и совместное распределение этих случайных величин, то есть распределение случайного вектора $\xi = (\xi_1, \ldots, \xi_n)$:*

$$\mathbb{P}_\xi = \mathbb{P}_{\xi_1} \times \cdots \times \mathbb{P}_{\xi_n}.$$

Задача 4.25. Пусть в примере 4.4 с равномерным распределением $M = [0,1]^d$. Докажите, что компоненты случайного вектора ξ независимы и имеют равномерное распределение на $[0,1]$. Верно ли это утверждение, если M — шар в $\mathbb{R}^d$?

Задача 4.26. Пусть вероятностное пространство имеет вид $([0,1], \mathcal{B}([0,1]), Leb)$. Рассмотрим пару функций $x(t) = 2t$, $y(t) = 1 - t^2$ как случайные величины. Зависимы ли они? Тот же вопрос для пары отображений $x(t) = \operatorname{sign}[\sin(2\pi t)]$ и $y(t) = \operatorname{sign}[\sin(4\pi t)]$.

Возвращаясь к связи интуитивного и формального понимания независимости, посмотрим, как обстоит дело с традиционно понимаемой функциональной зависимостью; обычно это утверждение $\xi_2 = g(\xi_1)$ о возможности выразить одну функцию через другую.

Задача 4.27. 1) Пусть случайные величины ξ_1, ξ_2 удовлетворяют соотношению $\xi_2 = g(\xi_1)$ для некоторой функции g. Верно ли, что они зависимы? Если нет, то при каких условиях?

2) Найдите пример двух зависимых случайных величин, таких что не существует функции g, для которой соотношение выше было бы верным.

Таким образом, зависимость в смысле теории вероятностей отвечает более широкому смыслу, нежели функциональная зависимость.

Сумма независимых случайных величин

Напомним, что для интегрируемых функций $f, g \in L_1(\mathbb{R})$ определена их свертка

$$(f * g)(x) := \int_{-\infty}^{\infty} f(x-y)g(y)\,dy;$$

она также удовлетворяет $f * g \in L_1(\mathbb{R})$.

Лемма 4.28. *Если случайные величины ξ и η независимы, абсолютно непрерывны и имеют плотности p_ξ и p_η, то случайная величина $\xi + \eta$ также абсолютно непрерывна и*

$$p_{\xi+\eta} = p_\xi * p_\eta.$$

Доказательство. Согласно предложению 4.23(3), случайный вектор (ξ, η) абсолютно непрерывен и имеет плотность $p_{(\xi,\eta)}(x, y) = p_\xi(x)p_\eta(y)$. Вычислим с ее помощью функцию распределения случайной величины $\xi + \eta$:

$$F_{\xi+\eta}(t) = \int_{x+y \le t} p_{(\xi,\eta)}(x, y)\,dxdy = \int_{-\infty}^{t}\int_{-\infty}^{\infty} p_{(\xi,\eta)}(s-y, y)\,dyds = \int_{-\infty}^{t} (p_\xi * p_\eta)(s)\,ds,$$

где мы использовали замену переменных $(s, y) := (x + y, y)$. □

Модель независимых случайных величин

Пусть имеется последовательность вероятностных пространств $(\Omega_i, \mathcal{F}_i, \mathbb{P}_i)$, $1 \le i \le n$, на i-ом пространстве определена случайная величина ξ_i. Как «поселить» их все на одно вероятностное пространство независимым образом? Для этого можно воспользоваться конструкцией, предложенной в параграфе 3.4. Действительно, достаточно выбрать вероятностное пространство $(\Omega, \mathcal{F}, \mathbb{P})$ в виде прямого произведения вероятностных пространств $(\Omega_i, \mathcal{F}_i, \mathbb{P}_i)$, и рассмотреть случайный вектор $\widetilde{\xi} = (\widetilde{\xi}_1, \ldots, \widetilde{\xi}_n)$ с компонентами $\widetilde{\xi}_i(\omega) := \xi_i(\omega_i)$. Нетрудно видеть, что $\mathbb{P}_{\widetilde{\xi}_i} = \mathbb{P}_{\xi_i}$ и компоненты вектора $\widetilde{\xi}$ независимы в совокупности. Убедимся в этом: в

$$\mathbb{P}\{\widetilde{\xi}_j \in A\} = \mathbb{P}\{\omega \in \Omega : \xi_j(\omega_j) \in A\} = \mathbb{P}_j\{\xi_j \in A\} = \mathbb{P}_{\xi_j}(A), \qquad \forall A \in \mathcal{B}(\mathbb{R}).$$

Далее, для всех $A_1, \ldots, A_n \in \mathcal{B}(\mathbb{R})$,

$$\mathbb{P}_{\widetilde{\xi}}(A_1 \times \cdots \times A_n) = \mathbb{P}\{\omega : \xi_1(\omega_1) \in A_1, \ldots, \xi_n(\omega_n) \in A_n\} = \mathbb{P}_{\xi_1}(A_1)\cdots\mathbb{P}_{\xi_n}(A_n) = \mathbb{P}_{\widetilde{\xi}_1}(A_1)\cdots\mathbb{P}_{\widetilde{\xi}_n}(A_n).$$

В случае бесконечного числа экспериментов нужно аналогичным образом воспользоваться соответствующей конструкцией вероятностного пространства из параграфа 3.4.

5 Характеристики случайных величин

5.1 Математическое ожидание

Пусть как всегда $(\Omega, \mathcal{F}, \mathbb{P})$ — вероятностное пространство, а $\xi : \Omega \mapsto \mathbb{R}$ — случайная величина.

Определение 5.1. *Математическим ожиданием* $\mathbb{E}\xi$ случайной величины ξ называется интеграл

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

Математическим ожиданием n-мерного случайного вектора $\eta = (\eta_1, \dots, \eta_n)$ называют вектор матожиданий его компонент $\mathbb{E}\eta := (\mathbb{E}\eta_1, \dots, \mathbb{E}\eta_n)$. См. параграф B.6, где напоминается конструкция интеграла Лебега и его основные свойства.

Как видно из определения, матожидание существует не для любой случайной величины, но только для интегрируемых по мере $\mathbb{P}$. Такие случайные величины образуют банахово пространство $L_1(\Omega, \mathbb{P})$ с нормой $\|\xi\|_{L_1} := \mathbb{E}|\xi|$, см. параграф 5.3.

Пример 5.2. Пусть Ω — не более, чем счетное множество. Тогда функция ξ автоматически является простой и, согласно опрделению интеграла Лебега от простых функций,

$$\mathbb{E}\xi = \sum_{\omega\in\Omega} \xi(\omega)\mathbb{P}\{\omega\}. \tag{5.1}$$

В частности, если $|\Omega| = n$ и вероятность определена классическим образом $\mathbb{P}(\omega) = 1/n$, матожидание случайной величины ξ совпадает со средним арифметическим ее значений

$$\mathbb{E}\xi = \sum_{\omega\in\Omega} \frac{\xi(\omega)}{n}.$$

Из определение выше создается иллюзорное впечатление, что матожидание $\mathbb{E}\xi$ зависит от структуры вероятностного пространства, а не только от распределения $\mathbb{P}_\xi$. Это не так: согласно теореме B.24 о замене переменной в интеграле Лебега, имеет место следующий результат.

Теорема 5.3. *Пусть $f : \mathbb{R}^n \mapsto \mathbb{R}$ — измеримая функция, а $\eta = (\eta_1, \dots, \eta_n)$ — случайный вектор. Тогда*

$$\mathbb{E}f(\eta) = \int_{\mathbb{R}^n} f(x)\, d\mathbb{P}_\eta(x), \quad \text{где} \quad x = (x_1, \dots, x_n). \tag{5.2}$$

Интеграл в правой части равенства существует тогда и только тогда, когда существует интеграл в левой части.

В частности,

$$\mathbb{E}\xi = \int_{\mathbb{R}} x\, d\mathbb{P}_\xi(x).$$

Замечание 5.4. Можно показать, что интеграл Лебега по мере $\mathbb{P}_\xi$ совпадает с интегралом Стилтьеса по dF_ξ, где F_ξ — функция распределения ξ. Поэтому в литературе вместо $d\mathbb{P}_\xi$ часто пишут dF_ξ.

Пример 5.5. Пусть ξ — дискретная случайная величина, а $X \subset \mathbb{R}$ — не более, чем счетное множество, такое что $\mathbb{P}\{\xi \in X\} = 1$. Тогда

$$\mathbb{E}\xi = \int_{\mathbb{R}} x\, d\mathbb{P}_\xi(x) = \sum_{x\in X} x\, \mathbb{P}_\xi\{x\}, \tag{5.3}$$

и, более обще, для функции $f : \mathbb{R} \mapsto \mathbb{R}$,

$$\mathbb{E}f(\xi) = \int_{\mathbb{R}} f(x)\, d\mathbb{P}_\xi(x) = \sum_{x\in X} f(x)\, \mathbb{P}_\xi\{x\}. \tag{5.4}$$

Формула (5.3) допускает наглядную интерпретацию: если думать про x как про координату точки, а про $\mathbb{P}_\xi\{x\}$ — как про массу, сидящую в этой точке, то $\mathbb{E}\xi$ дает центр масс системы точек X. Формула (5.1) допускает аналогичную интерпретацию, где ω — имена точек, $\xi(\omega)$ — их координаты, а $\mathbb{P}_\xi\{\omega\}$ — веса, ассоциированные с точками. Заметим, что при этом координаты различных точек могут совпадать. Если же собрать вместе все точки с одинаковыми координатами и сложить их массы, то формула (5.1) превращается в точности в формулу (5.3). Проверьте это!

Пример 5.6. Если случайная величина ξ абсолютно непрерывна, то $d\mathbb{P}_\xi(x) = p_\xi(x)\,dx$, так что

$$\mathbb{E}\xi = \int_{-\infty}^{\infty} xp_\xi(x)\,dx \quad \text{и} \quad \mathbb{E}f(\xi) = \int_{-\infty}^{\infty} f(x)p_\xi(x)\,dx. \tag{5.5}$$

Пример 5.7. • Если $\xi \sim Bernoulli(p)$, $0 \le p \le 1$, то есть $\mathbb{P}(\xi = 1) = p$ и $\mathbb{P}(\xi = 0) = 1 - p$, то

$$\mathbb{E}\xi = 0 \times (1 - p) + 1 \times p = p.$$

• Если $\xi \sim Exp(\lambda)$, $\lambda > 0$, то есть $p_\xi(x) = \lambda e^{-\lambda x}\mathbb{I}_{[0,\infty)}(x)$, то

$$\mathbb{E}\xi = \lambda \int_0^\infty xe^{-\lambda x}\,dx = -xe^{-\lambda x}\Big|_0^\infty + \int_0^\infty e^{-\lambda x}\,dx = \lambda^{-1}.$$

Обозначение. Говорят, что свойство $\mathcal{P}$ выполнено *почти наверное* (п.н.) или *почти всюду* (п.в.), если оно выполняется с вероятностью 1. Например, говоря, что случайная величина ξ неотрицательна почти всюду, $\xi \ge 0$ п.в., имеется ввиду, что $\mathbb{P}\{\xi \ge 0\} = 1$. [16]

Перечислим некоторые свойства математического ожидания.

Предложение 5.8. *Пусть $\xi, \eta \in L_1(\Omega, \mathbb{P})$. Тогда*

1. *Если $\xi = C$ п.в. для какой-нибудь константы C, то $\mathbb{E}\xi = C$.*
2. *Если $\xi \ge \eta$, то $\mathbb{E}\xi \ge \mathbb{E}\eta$. В частности, $\mathbb{E}|\xi| \ge |\mathbb{E}\xi|$, и если $\xi \ge 0$, то $\mathbb{E}\xi \ge 0$.*
3. *$\mathbb{E}(a\xi + b\eta) = a\mathbb{E}\xi + b\mathbb{E}\eta$ для произвольных констант a, b.*
4. *Если $\xi = \mathbb{I}_A$ п.н. для события $A \in \mathcal{F}$, то $\mathbb{E}\xi = \mathbb{P}(A)$.*
5. *Если ξ и η независимы, то $\mathbb{E}(\xi\eta) = \mathbb{E}\xi\mathbb{E}\eta$.*

Доказательство. Пункты 1-4 немедленно следуют из свойств интеграла Лебега. Докажем пункт 5. Используя в первом равенстве ниже теорему 5.3, а во втором — теорему 4.23(1) вместе с теоремой Фубини, находим

$$\mathbb{E}(\xi\eta) = \int_{\mathbb{R}^2} xy\, d\mathbb{P}_{(\xi,\eta)}(x, y) = \int_{-\infty}^{\infty} x\, d\mathbb{P}_\xi(x) \int_{-\infty}^{\infty} y\, dP_\eta(y) = \mathbb{E}\xi\mathbb{E}\eta.$$

Заметим что из существования интегралов справа следует существование интеграла слева, а значит и матожидания $\mathbb{E}(\xi\eta)$. □

[16] На английском: п.в. = a.e. (almost everywhere), п.н. = a.s. (almost surely).

Задача 5.9. Пусть известно, что $\mathbb{E}(\xi\eta) = \mathbb{E}\xi\mathbb{E}\eta$. Верно ли, что случайные величины ξ и η обязательно независимы?

Задача 5.10. Элементарное на первый взгляд доказательство предложения 5.8(5) опирается аш на три мощных результата: теоремы 5.3, 4.23 и теорему Фубини. Чтобы понять, что там на самом деле произошло, докажите это утверждение "руками" для случая, когда случайные величины ξ и η дискретны, опираясь на формулу (5.3).

5.2 Дисперсия

Пусть случайная величина ξ такова, что $\mathbb{E}\xi^2 < \infty$. Такие случайные величины называются *квадратично интегрируемыми.*

Определение 5.11. *Дисперсией* $\operatorname{Var}\xi$ случайной величины ξ называется число

$$\operatorname{Var}\xi := \mathbb{E}\big(\xi - \mathbb{E}\xi\big)^2.$$

Среднеквардатичным или *стандартным* отклонением случайной величины ξ называют число $\sqrt{\operatorname{Var}\xi}$. Стандартное отклонение используется для описания средней величины отклонения случайной величины от своего матожидания. Для этой цели можно было бы использовать величину $\mathbb{E}|\xi - \mathbb{E}\xi|$, но работать с модулем гораздо менее удобно, чем с квадратом.

Раскрывая квадрат, находим альтернативную формулу для дисперсии

$$\operatorname{Var}\xi = \mathbb{E}\big(\xi^2 - 2\xi\mathbb{E}\xi + (\mathbb{E}\xi)^2\big) = \mathbb{E}(\xi^2) - (\mathbb{E}\xi)^2, \tag{5.6}$$

где мы использовали, что $\mathbb{E}(2\xi\mathbb{E}\xi) = 2(\mathbb{E}\xi)^2$, поскольку $\mathbb{E}\xi$ — константа. В этой формуле скрыт "физический" смысл дисперсии, зато она удобна для вычислений.

Пример 5.12. 1. Если случайная величина ξ дискретна, а $X = \xi(\Omega)$ обозначает ее множество значений, то, согласно (5.4),

$$\operatorname{Var}\xi = \sum_{x\in X} \big(x - \mathbb{E}\xi\big)^2\, \mathbb{P}\{\xi = x\} = \sum_{x\in X} x^2\, \mathbb{P}\{\xi = x\} - (\mathbb{E}\xi)^2.$$

2. Если случайная величина ξ абсолютно непрерывна, то, согласно (5.5),

$$\operatorname{Var}\xi = \int_{-\infty}^{\infty} \big(x - \mathbb{E}\xi\big)^2 p_\xi(x)\,dx = \int_{-\infty}^{\infty} x^2 p_\xi(x)\,dx - (\mathbb{E}\xi)^2.$$

Заметим, что соотношения $\mathbb{E}(\xi^2) < \infty$ достаточно для существования дисперсии. Действительно, так как $|\xi| \le \xi^2 + 1$, оно влечет

$$|\mathbb{E}\xi| \le \mathbb{E}|\xi| \le \mathbb{E}(\xi^2 + 1) = \mathbb{E}(\xi^2) + 1 < \infty,$$

так что и $\operatorname{Var}\xi < \infty$, согласно формуле (5.6). [17]

Задача 5.13. Придумайте случайную величину, у которой матожидание конечно, а дисперсия — нет.

Обозначение: далее мы будем опускать скобки и писать $\mathbb{E}\xi^p$ вместо $\mathbb{E}(\xi^p)$. Напротив, в выражении $(\mathbb{E}\xi)^p$ мы всегда будем ставить скобки, так что путаницы не возникнет.

Перечислим некоторые свойства дисперсии.

[17] Пожалуй, здесь было бы проще воспользоваться неравенством Коши-Буняковского $|\mathbb{E}\xi| \le \sqrt{\mathbb{E}(\xi^2)}$, см. следующий параграф.

Предложение 5.14. *Пусть случайные величины ξ и η квадратично интегрируемы. Тогда*

1. $\operatorname{Var}\xi \geq 0$.
2. $\operatorname{Var}\xi = 0 \quad \Leftrightarrow \quad \xi = C$ *п.в. для некоторой константы C (разумеется, равной $\mathbb{E}\xi$).*
3. $\operatorname{Var}(C\xi) = C^2 \operatorname{Var}\xi$ *и* $\operatorname{Var}(\xi + C) = \operatorname{Var}\xi$ *для произвольной константы $C \in \mathbb{R}$.*
4. *Если ξ и η независимы, то* $\operatorname{Var}(\xi+\eta) = \operatorname{Var}\xi + \operatorname{Var}\eta$.

Доказательство. Пункты 1-3 сразу следуют из определения дисперсии и свойств интеграла Лебега. Для доказательства пункта 4 воспользуемся формулой (5.6) и раскроем скобки:

$$\operatorname{Var}(\xi+\eta) = \mathbb{E}\xi^2 + \mathbb{E}\eta^2 + 2\mathbb{E}(\xi\eta) - \big((\mathbb{E}\xi)^2 + (\mathbb{E}\eta)^2 + 2\mathbb{E}\xi\mathbb{E}\eta\big) = \operatorname{Var}\xi + \operatorname{Var}\eta + 2\big(\mathbb{E}(\xi\eta) - \mathbb{E}\xi\mathbb{E}\eta\big).$$

Остается заметить, что $\mathbb{E}(\xi\eta) = \mathbb{E}\xi\mathbb{E}\eta$ в силу независимости ξ и η. □

Задача 5.15. Пусть $\operatorname{Var}(\xi+\eta) = \operatorname{Var}\xi + \operatorname{Var}\eta$. Верно ли, что случайные величины ξ и η независимы?

Обозначения $\mathbb{E}$ и Var происходят от английских слов expectation и variation. В русскоязычной литературе матожидание и дисперсия часто обозначаются буквами $\mathbb{M}$ и $\mathbb{D}$.

5.3 Пространства L_p случайных величин и высшие моменты

Напомним, что (банахово) пространство $L_p(\Omega,\mathbb{P})$, $p \geq 1$, состоит из измеримых функций[18] $\xi : \Omega \mapsto \mathbb{R}$ с конечной нормой

$$\|\xi\|_{L_p} = \Big(\int_\Omega |\xi(\omega)|^p \, d\mathbb{P}(\omega)\Big)^{1/p}.$$

Как мы уже выяснили, на языке теории вероятностей измеримые функции называются случайными величинамим, а указанная норма удобно записывается с помощью матожидания:

$$\|\xi\|_{L_p} = \big(\mathbb{E}|\xi|^p\big)^{1/p}.$$

В частности, пространства интегрируемых и квадратично интегрируемых функций, в которых мы определяли матожидание и дисперсию, есть ни что иное как $L_1(\Omega,\mathbb{P})$ и $L_2(\Omega,\mathbb{P})$.

Пространство $L_2(\Omega,\mathbb{P})$ является гильтертовым и поэтому, как всегда, играет особую роль. Скалярное произведение в нем имеет вид

$$\langle \xi,\eta\rangle = \int_\Omega \xi(\omega)\eta(\omega)\, d\mathbb{P}(\omega) = \mathbb{E}(\xi\eta).$$

Напомним неравенство Коши-Буняковского: для $\xi,\eta \in L_2(\Omega,\mathbb{P})$

$$|\langle \xi,\eta\rangle| \leq \|\xi\|_{L_2}\|\eta\|_{L_2}.$$

На языке теории вероятностей оно переписывается в виде

$$|\mathbb{E}(\xi\eta)| \leq \big(\mathbb{E}\xi^2\mathbb{E}\eta^2\big)^{1/2}.$$

[18] Точнее говоря, классов эквивалентности, но вы это знаете и мы будем говорить просто "функций" .

В частности, выбирая $\eta = 1$, находим $|\mathbb{E}\xi| \le \sqrt{\mathbb{E}\xi^2}$.

Неравенство Гёльдера обобщает неравенство Коши-Буняковского: для $p, q > 1$ таких, что $\frac{1}{p} + \frac{1}{q} = 1$ и $\xi \in L_p$, $\eta \in L_q$, имеем $\xi\eta \in L_1$ и

$$\|\xi\eta\|_{L_1} \le \|\xi\|_{L_p}\|\eta\|_{L_q}, \quad \text{то есть} \quad \mathbb{E}|\xi\eta| \le (\mathbb{E}|\xi|^p)^{1/p}(\mathbb{E}|\eta|^q)^{1/q}.$$

Определение 5.16. Матожидание $\mathbb{E}|\xi|^p$, $p \in \mathbb{N}$, называется p-ым моментом случайной величины ξ.

С помощью формулы (5.6) дисперсия случайной величины выражается через первый и второй моменты. Предложение 5.14(2) нам гарантирует, что если дисперсия равна нулю, то случайная величина почти наверное совпадает со своим матожиданием. В общем случае, разумеется, знания двух первых моментов недостаточно, чтобы восстановить распределение случайной величины. Однако, чем больше моментов нам известно, тем большей информацией мы обладаем. Оказывается, при разумных предположениях на скорость роста моментов, распределение случайной величины *однозначно* определяется своими моментами. То есть, если у двух случайных величин все моменты совпадают, то и их распределения совпадают. Мы коснемся этого позже, когда будем говорить о характеристических функциях случайных величин. Также возникает обратный вопрос: каким условиям должна удовлетворять последовательность чисел $(m_j)_{j\ge1}$, чтобы она являлась моментами какой-нибудь случайной величины? Это знаменитая, красивая и совершенно нетривиальная задача, называемая Проблемой моментов. См. Википедию https://en.wikipedia.org/wiki/Moment_problem и книгу Н.И. Ахиезер, "Классическая проблема моментов" .

5.4 Ковариация и коэффициент корреляции

Определение 5.17. *Ковариацией* случайных величин $\xi, \eta \in L_2(\Omega, \mathbb{P})$ называется число

$$\operatorname{Cov}(\xi, \eta) := \mathbb{E}\big[(\xi - \mathbb{E}\xi)(\eta - \mathbb{E}\eta)\big].$$

Раскрывая скобки, находим альтернативную формулу, из которой видно, что квадратичной интегрируемости действительно достаточно для существования ковариации:

$$\operatorname{Cov}(\xi, \eta) = \mathbb{E}(\xi\eta) - \mathbb{E}\xi\mathbb{E}\eta.$$

Ковариация является «мерой совместной изменчивости»: если она положительна и велика, то случайные величины значительно отклоняются от своих средних, и при том отклонения в основном происходят в одну и ту же сторону. Если же ковариация отрицательна, то отклонения в основном происходят в противоположные стороны.

Предложение 5.18. *Пусть $\xi, \eta, \zeta \in L_2(\Omega, \mathbb{P})$. Ковариация удовлетворяет следующим свойствам:*

1. *Симметричность:* $\operatorname{Cov}(\xi, \eta) = \operatorname{Cov}(\eta, \xi)$*;*
2. *Билинейность:* $\operatorname{Cov}(a\xi + b\zeta, \eta) = a\operatorname{Cov}(\xi, \eta) + b\operatorname{Cov}(\zeta, \eta)$*;*
3. $\operatorname{Cov}(\xi, \xi) = \operatorname{Var}\xi$*;*
4. $|\operatorname{Cov}(\xi, \eta)| \le \sqrt{\operatorname{Var}\xi}\sqrt{\operatorname{Var}\eta}$.
5. *Если ξ и η независимы, то* $\operatorname{Cov}(\xi, \eta) = 0$.

Доказательство. Первые три свойства сразу следуют из определения ковариации, а четвертое — из неравенства Коши-Буняковского. Последнее свойство следует из предложения 5.8(5). □

Если $\mathrm{Cov}(\xi, \eta) = 0$, то говорят, что случайные величины ξ и η *некоррелированы.* Из независимости следует некоррелированность, **но не наоборот.** Некоррелированность говорит скорее о симметричности совместного распределения $\mathbb{P}_{(\xi,\eta)}$, чем о независимости.

Задача 5.19. Пусть вектор (ξ, η) имеет равномерное распределение в круге радиуса 1. Докажите, что $\mathrm{Cov}(\xi, \eta) = 0$, но ξ и η не являются независимыми.

Говорят, что случайная величина ξ *центрирована*, если $\mathbb{E}\xi = 0$. Множество центрированных квадратично интегрируемых случайных величин образует подпространство в L_2, а ковариация на нем задает скалярное произведение. Действительно, в этом случае $\mathrm{Cov}(\xi, \eta) = \mathbb{E}(\xi\eta)$.

Случайные величины с очень большой дисперсией и малой «совместной изменчивостью» могут иметь большую ковариацию. Поэтому удобно ее нормировать:

Определение 5.20. *Коэффициентом корреляции* (линейным коэффициентом корреляции или коэффициентом корреляции Пирсона) случайных величин $\xi, \eta \in L_2$ называется число

$$\rho_{\xi,\eta} = \frac{\mathrm{Cov}(\xi, \eta)}{\sqrt{\mathrm{Var}\,\xi}\sqrt{\mathrm{Var}\,\eta}}.$$

Лемма 5.21. *Верно неравенство $|\rho_{\xi,\eta}| \leq 1$, при этом $|\rho_{\xi,\eta}| = 1$ тогда и только тогда, когда ξ и η линейно зависимы, то есть $\xi = a\eta + b$ п.н. для некоторых констант a, b. Знаки a и $\rho_{\xi,\eta}$ совпадают.*

Доказательство. Неравенство следует из предложения 5.18(4). Остальное следует из неравенства Коши-Буняковского, которое утверждает, что $|\langle f, g\rangle| = \|f\|\|g\|$ в том и только том случае, когда функции f и g линейно зависимы, то есть $f = ag$. Теперь нужно подставить $f = \xi - \mathbb{E}\xi$, $g = \eta - \mathbb{E}\eta$, что дает $\xi = a\eta + (\mathbb{E}\xi - a\mathbb{E}\eta)$. Утверждение насчет знака очевидно. □

5.5 Производящая функция случайной величины

Для принимающей *только неотрицательные целые значения дискретной случайной величины* η возможно определить **производящую функциию** $Q_\eta(z)$ рядом

$$Q_\eta(z) = \mathbb{P}(\eta = 0) + \mathbb{P}(\eta = 1)z + \ldots + \mathbb{P}(\eta = k)z^k + \ldots \tag{5.7}$$

Из анализа известно, что этот степенной ряд заведомо сходится при $|z| \leqslant 1$ поскольку сумма всех вероятностей $\mathbb{P}(\eta = k)$ равна 1. Можно также переписать этот ряд в виде

$$\mathbb{P}(\eta = 0) + \mathbb{P}(\eta = 1)z + \ldots + \mathbb{P}(\eta = k)z^k + \ldots = \mathbb{E}\left(z^\xi\right),$$

где мы принимаем $0^0 := \lim\limits_{x\to 0} x^0 = 1$. Выбор традиционного обозначения z для переменной не случаен: в будущем, при рассмотрении характеристических функций, будет прояснена возможная роль комплексных чисел в этой конструкции (уже сейчас видно, что функция $Q_\eta(z)$ аналитична в круге $|z| < 1$).

Основная идея использования производящей функции возникает при рассмотрении значения ее производной:

$$Q'_\eta(z) = \mathbb{P}(\eta = 1) + 2 \cdot \mathbb{P}(\eta = 2)z + \ldots + k \cdot \mathbb{P}(\eta = k)z^{k-1} + \ldots \tag{5.8}$$

$$Q'_\eta(1) = 0 \cdot \mathbb{P}(\eta = 0) + 1 \cdot \mathbb{P}(\eta = 1) + 2 \cdot \mathbb{P}(\eta = 2) + \ldots + k \cdot \mathbb{P}(\eta = k) + \ldots = \mathbb{E}(\eta) \tag{5.9}$$

Почленным дифференциированием ряда получаются и другие полезные формулы. Например, дважды продифференциировав производящую функцию получим ряд $\sum k(k-1)\mathbb{P}(\eta = k)z^{k-2}$, после подстановки $z = 1$ имеем формулу $\mathbb{E}\left(\eta(\eta - 1)\right) = Q''_\eta(1) = \mathbb{E}(\eta^2) - \mathbb{E}(\eta)$. Если к этому выражению добавить $\mathbb{E}(\eta) - (\mathbb{E}(\eta))^2$, то возникнет формула, выражающая дисперсию через производящую функцию

$$\mathrm{Var}(\eta) = Q''_\eta(1) + Q'_\eta(1) - \left(Q'_\eta(1)\right)^2.$$

Предложение 5.22. *Производящая функция Q_η случайной величины η однозначно определяет ее распределение $\mathbb{P}_\eta$.*

Доказательство. Пусть производящая функция Q_ξ случайной величины ξ совпадает с Q_η:

$$\sum_{k=0}^{\infty} z^k \mathbb{P}(\xi = k) = \sum_{k=0}^{\infty} z^k \mathbb{P}(\eta = k),$$

при $|z| \le 1$, причем оба ряда абсолютно сходятся, так как $\sum_{k=0}^{\infty} \mathbb{P}(\xi = k) = \sum_{k=0}^{\infty} \mathbb{P}(\eta = k) = 1$. Отсюда следует совпадение коэффициентов этих рядов, то есть равенство $\mathbb{P}(\xi = k) = \mathbb{P}(\eta = k)$ для всех k. □

Предложение 5.23. *Пусть случайная величина η, как и выше, принимает только целые неотрицательные значения, все ее моменты конечны и $\sum_{k=0}^{\infty} \frac{\varepsilon^k \mathbb{E}|\eta|^k}{k!} < \infty$ для некоторого $\varepsilon > 0$. Тогда ее моменты однозначно задают ее распределение $\mathbb{P}_\eta$.*

Доказательство. Пусть $0 < z < 1$. Положим $t = \ln z$. Тогда

$$Q_\eta(z) = \mathbb{E}e^{t\eta} = \mathbb{E}\sum_{k=0}^{\infty} \frac{(t\eta)^k}{k!} = \sum_{k=0}^{\infty} \frac{t^k \mathbb{E}\eta^k}{k!},$$

по лемме Фату, если $|t| \le \varepsilon$. Таким образом, моменты однозначно определяют производящую функцию $Q(z)$ при $e^{-\varepsilon} \ge z < 1$. Так как производящая функция аналитична в круге $|z| < 1$, это однозначно задает ее во всем круге. Теперь результат следует из предложения 5.22.

Сумма двух независимых случайных величин

Пусть ξ и η неотрицательные *независимые* целочисленные случайные величины с распределениями вероятностей $\mathbb{P}(\xi = j) = a_j$ и $\mathbb{P}(\eta = k) = b_k$. Событие $(\xi = j, \eta = k)$ в силу независимости имеет вероятность $a_j b_k$. Сумма $\xi + \eta$ есть новая случайная величина, и событие $\xi + \eta = m$ есть объединение несовместных событий вида $(\xi = j, \eta = m-j)$ и поэтому соответствующая вероятность суммы событий равна сумме вероятностей $a_0 b_m + a_1 b_{m-1} + \ldots + a_m b_0$. При перемножении соответствующих рядов, определяющих производящие функции для ξ и η, получается то же самое выражение. Поэтому для *независимых* целочисленных случайных величин ξ и η с неотрицательными значениями $Q_{\xi+\eta}(z) = Q_\xi(z) Q_\eta(z)$

Производящие функциии удобны для вычисления моментов дискретных случайных величин специального вида и *для вычислений с суммами независимых случайных величин.* Действительно, для дискретных с.в. специального вида закон распределения суммы нескольких независимых случайных величин получается так: надо перемножить соответствующие производящие функции и это произведение даст производящую функцию суммы. Разложение ее в ряд Маклорена определит соответствующие вероятности в дискретном распределении.

Суммы случайного числа независимых случайных величин

Пусть на Ω задана i.i.d последовательность $\{\xi_n\}$ (то есть последовательность независимых и одинаково распределенных) принимающих только неотрицательные целые значения дискретных случайных величин, у всех, тем самым, одинаковая производящая функция $Q_\xi(z)$. Пусть кроме того на Ω задана еще независимая от $\{\xi_n\}$ и тоже принимающая только неотрицательные целые значения дискретная случайная величина η, ее производящую функцию обозначим $Q_\eta(z)$. Тогда можно рассмотреть с.в. β, которая на любом исходе $\omega \in \Omega$ принимает значение равное сумме

$$\beta(\omega) = \xi_1(\omega) + \ldots + \xi_{\eta(\omega)}(\omega)$$

число слагаемых определено значением $\eta(\omega)$. Нетрудно видеть, что по формуле полной вероятности

$$\mathbb{P}(\beta = n) = \sum_{k=1}^{\infty} \mathbb{P}(\eta = k)\mathbb{P}(\xi_1 + \ldots + \xi_k = n)$$

Чтобы найти вероятность $\mathbb{P}(\xi_1 + \ldots + \xi_k = n)$ достаточно разложить $(Q_\xi(z))^k$ в ряд и взять коэффициент при z^n. Таким образом, производящая функция $Q_\beta(z)$ отвечает выражению

$$Q_\beta(z) = \sum_{k=0}^{\infty} \mathbb{P}(\eta = k)(Q_\xi(z))^k = Q_\eta\left(Q_\xi(z)\right)$$

то есть композиции производящих функций. Пример: если величины ξ_k i.i.d индикаторы с вероятностями p и $q = 1 - p$, тогда $Q_\xi(z) = q + pz$ и, значит, $Q_\beta(z) = Q_\eta(q + pz)$.

5.6 Медиана и квантили

Пусть ξ — случайная величина. Ее *α-квантилем*, $0 < \alpha < 1$, называется число q_α, такое что

$$\mathbb{P}(\xi \le q_\alpha) \ge \alpha \quad \text{и} \quad \mathbb{P}(\xi \ge q_\alpha) \ge 1 - \alpha.$$

Квантиль всегда существует, но, вообще говоря, определен не однозначно.

Однако в случае, когда обратная к функции F_ξ^{-1} в точке α существует и однозначна, предыдущие неравенства эквиваленты равенству

$$\mathbb{P}(\xi \le q_\alpha) = F_\xi(q_\alpha) = \alpha,$$

то есть

$$q_\alpha = F_\xi^{-1}(\alpha). \tag{5.10}$$

Медианой m_ξ случайной величины ξ называют квантиль $q_{1/2}$. Медиана — альтернативный математическому ожиданию способ понимания того, что такое « среднее» случайной величины.

Пример 5.24. 1) Пусть $\xi \sim Exp(1)$. Тогда

$$F_\xi(x) = \int_0^x e^{-t}\,dx = 1 - e^{-x}.$$

Таким образом, функция $F_\xi : (0,\infty) \mapsto (0,1)$ биективна, а значит можно воспользоваться формулой (5.10). Находим

$$1 - e^{-q_\alpha} = \alpha,$$

так что $q_\alpha = -\ln(1-\alpha)$.

2) Пусть $\mathbb{P}(\xi = 1) = 9/10$, а $\mathbb{P}(\xi = 1000) = 1/10$. Тогда $m_\xi = 1$. Если с.в. ξ отражает распределеие зарплат сотрудников какой-нибудь организации, то такое понимание средней зарплаты может быть гораздо разумнее, чем матожидание $\mathbb{E}\xi = 9/10 + 100$.

3) Пусть $\xi \sim Bernoulli(1/2)$, т.е. $\mathbb{P}(\xi = 0) = 1/2 = \mathbb{P}(\xi = 1)$. Тогда медианой является любое число из отрезка $[0,1]$.

5.7 Неравенства для моментов

В параграфе 5.3 мы уже напомнили некоторые неравенства, известные из курса анализа, и дали их вероятностную интерпретацию. Обсудим некоторые из них более подробно и приведем другие.

Неравенство Коши-Буняковского

Пусть случайные величины ξ и η таковы, что $\mathbb{E}\xi^2 < \infty$ и $\mathbb{E}\eta^2 < \infty$. Тогда

$$\Big(\mathbb{E}(\xi\eta)\Big)^2 \leqslant \mathbb{E}\xi^2\mathbb{E}\eta^2$$

Доказательство. Если $\mathbb{E}\xi^2 = 0$ или $\mathbb{E}\eta^2 = 0$, то $\xi = 0$ почти наверно или $\eta = 0$ почти наверное соответственно, так что утверждение верно. Пусть теперь оба вторых момента отличны от нуля. Перейдем к с.в. $\tilde{\xi} = \frac{\xi}{\sqrt{\mathbb{E}\xi^2}}$ и $\tilde{\eta} = \frac{\eta}{\sqrt{\mathbb{E}\eta^2}}$ имеем в силу обычного неравенства $2|xy| \leqslant x^2 + y^2$ для $x, y \in \mathbb{R}$

$$\begin{aligned} 2\mathbb{E}|\tilde{\xi}\tilde{\eta}| &\leqslant \mathbb{E}\tilde{\xi}^2 + \mathbb{E}\tilde{\eta}^2 = 2 \\ \mathbb{E}|\tilde{\xi}\tilde{\eta}| &\leqslant 1 \end{aligned}$$

□

Неравенство Йенсена

Пусть $g = g(x)$ — выпуклая вниз измеримая по Борелю числовая функция и $\mathbb{E}|\xi| < \infty$, тогда $g\left(\mathbb{E}\xi\right) \leqslant \mathbb{E}g(\xi)$.

Доказательство. Из-за выпуклости $\forall\ x_0 \in \mathbb{R}\ \exists \lambda(x_0)$ такое, что

$$\forall\ x_0 \in \mathbb{R}\ g(x) \geqslant g(x_0) + (x - x_0)\lambda(x_0)$$

Выбирая теперь $x_0 = \mathbb{E}\xi\ x = \xi$ имеем

$$g(\xi) \geqslant g(\mathbb{E}\xi) + (\xi - \mathbb{E}\xi)\lambda(\mathbb{E}\xi)$$

и, применяя математическое ожидание к последнему неравенству, получаем требуемое: $g\left(\mathbb{E}\xi\right) \leqslant \mathbb{E}\left[g(\xi)\right]$ □

Неравенство Ляпунова

Для $0 < s < t$ выполнено
$$(\mathbb{E}|\xi|^s)^{1/s} \leqslant \left(\mathbb{E}|\xi|^t\right)^{1/t}$$

Доказательство. для $r = t/s > 1$ и выпуклой функции $g(x) = |x|^r$ для с.в. $\eta = |\xi|^s$ неравенство Йенсена дает требуемое в виде $[\mathbb{E}(|\xi|^s)]^{t/s} \leqslant \mathbb{E}(|\xi|^t)$.

Заметим, что в терминах анализа это означает, что пространство L_t непрерывно вложено в пространство L_s при $0 < s < t$.

6 Важнейшие дискретные распределения

В естествознании имеется базовый набор примеров дискретных вероятностных пространств и связанных с ними случайных величин. Значение этих общепринятых примеров чрезвычайно велико, без их знания невозможно никакое дальнейшее углубление в теорию, поскольку более сложные конструкции обычно принято сводить к базовым. Очень подробно свойства этих базовых моделей разобраны в книге В.Феллер "Введение в теорию вероятностей и ее приложения", том 1, поиск на указатель терминов.

В этом параграфе всегда $0 \le p \le 1$ и $q = 1 - p$.

6.1 Распределние Бернулли (индикатор)

Случайная величина принимает всего два значения:

ξ	0	1
$\mathbb{P}_\xi$	q	$p.$

Говорят, что ξ имеет распределение Бернулли, пишут $\xi \sim Bernoulli(p)$. Еще иногда говорят, что ξ— индикаторная случайная величина

Модель

Распределение возникает при рассмотрении опыта с двумя исходами (однократное бросание неровной монеты), которые мы назовем герб и решка или успех и неудача.

Производящая функция

Очевидно, имеет вид $Q(z) = (q + pz)$.

Основные характеристики

Непосредственным вычислением устанавливаем, что $\mathbb{E}\eta = p$ и $\operatorname{Var}\eta = pq$.

6.2 Биномиальное распределение

Следующее распределение случайной величины ξ называется *биномиальным распределением* с параметрами $n \geq 1$ и p:

ξ	0	$\dots$	k	$\dots$	n
$\mathbb{P}_\xi$	$\binom{n}{0}q^n$	$\dots$	$\binom{n}{k}p^k q^{n-k}$	$\dots$	$\binom{n}{n}p^n$

Пишут $\xi \sim Binomial(n, p)$ или $B(n, p)$.

Модель

Биномиальное распределение возникает при рассмотрении n-кратных независимых повторений опыта с двумя исходами. Таким образом, про ξ можно думать как про число орлов, выпавших при n независимых бросаниях монеты с вероятностью выпадения орла равной p. Точнее, нетрудно проверить, что случайная величина

$$S_n := \xi_1 + \cdots + \xi_n, \tag{6.1}$$

где $\xi_j \sim Bernoulli(p)$ и независимы, имеет биномиальное распределение $B(n,p)$.

Производящая функция

В силу (6.1), производящая функция биномиального распределения равна

$$Q_\xi(z) = (q + pz)^n. \tag{6.2}$$

Основные характеристики

Воспользуемся свойствами математического ожидания и дисперсии, и формулой (6.1), чтобы убедиться в том, что $\mathbb{E}(\xi) = np$ (аддитивность математического ожидания), $\mathrm{Var}(\xi) = npq$ (дисперсия суммы независимых с.в. равна сумме их дисперсий).

6.3 Геометрическое распределение

Это (не путать с геометрическими вероятностями!) распределение целочисленной неотрицательной случайной величины, принимающей бесконечно много значений:

ξ	0	1	$\ldots$	k	$\ldots$
$\mathbb{P}_\xi$	p	qp	$\ldots$	$q^k p$	$\ldots$

Пишут $\xi \sim Geom(p)$.

Модель

Геометрическое распределение возникает при рассмотрение независимых повторений эксперимента с двумя исходами до наступления первого успеха, где вероятность успеха в единичном эксперименте равна p. Если речь идет о подбрасывании монеты (успех – выпадения герба), то элементарные исходы кодируются последовательностями букв вида: Г, РГ, РРГ, РРРГ,. . . . Вероятности элементарных событий выбираются по правилу независимых повторений. На исходе РРРРРР. . . РГ (k решек подряд) случайная величина ξ принимает значение равное k (число неудач до первого успеха). Надо только проверить, что это действительно распределение, в самом деле:

$$p + qp + q^2 p + \ldots = p(1 + q + q^2 + \ldots) = \frac{p}{1-q} = 1$$

Производящая функция

$$Q(z) = p + qpz + q^2pz^2 + \ldots = p(1 + qz + (qz)^2 + \ldots) = \frac{p}{1-qz}$$

Основные характеристики

По формуле $\mathbb{E}(\xi) = Q'_\xi(1) = \frac{q}{p}$

- Найдите самостоятельно дисперсию геометрического распределения.

6.4 Распределение Паскаля

β	0	1	...	k	...
	p^r	$\binom{r}{1}qp^r$	...	$\binom{r+k-1}{k}q^kp^r$	...

, где $r \geq 1$. Пишут $\beta \sim BN(n,p)$ (Blaise Pascal).

Проверка свойств и заодно обобщенная формула Бинома Ньютона

Для распределения Паскаля часто используется также название «отрицательное биномиальное распределение», его происхождение объясняется так:

$$\binom{x}{k} = \frac{x(x-1)(x-2)\dots(x-k+1)}{k!} \Longrightarrow \qquad \forall r > 0 \qquad \binom{-r}{k} = (-1)^k\binom{r+k-1}{k}$$

Следовательно вероятности можно записать в ином, более удобном для запоминания, виде:

β	0	1	...	k	...
	p^r	$\binom{-r}{1}(-q)p^r$	...	$\binom{-r}{k}(-q)^kp^r$	...

Общеизвестна каноническая формула Бинома Ньютона, которая доказывается раскрытием скобок и использованием числа сочетаний для подсчета способов получить при приведении подобных членов данную степень. В частности:

$$(1-q)^r = \sum_{k=0}^{r}\binom{r}{k}(-q)^k$$

Рассмотрим теперь степенной ряд $1+q+q^2+\ldots = 1/(1-q)$ и продифференциируем его почленно r раз, в результате получится ряд

$$\frac{(-1)^r r!}{(1-q)^{r+1}} = r! + \frac{(r+1)!}{1!}q + \frac{(r+2)!}{2!}q^2 + \ldots = r!\sum_{k=0}^{\infty}\binom{r+k}{k}q^k$$

$$(-1)^{r-1}p^{-r} = \frac{(-1)^{r-1}}{(1-q)^r} = (-1)^{r-1}\sum_{k=0}^{\infty}\binom{r+k-1}{k}q^k = (-1)^{r-1}\sum_{k=0}^{\infty}(-1)^k\binom{-r}{k}(-1)^k(-q)^k$$

$$(1-q)^{-r} = \sum_{k=0}^{\infty}\binom{-r}{k}(-q)^k$$

Последнее равенство – это обобщение Бинома Ньютона для отрицательных степеней, а предпоследняя объясняет, почему сумма всех вероятностей в распределении Паскаля действительно равна 1.

Модель

Распределение Паскаля возникает при рассмотрение независимых повторений эксперимента с двумя исходами до наступления r-го успеха. Если речь идет о подбрасывании монеты (успех – выпадения герба), то элементарные исходы кодируются заканчивающимися буквой Г последовательностями вида: РРРГРРРГРГ...РГ, общее число букв Г в последовательности равно r. Если число решек в последовательности равно k, то случайная величина β принимает на таком элементарном исходе значение k (число неудач до r-го успеха). Вероятности элементарных событий выбираются по правилу независимых повторений. Нетрудно заметить, что всего таких последовательностей фиксированной длины $r+k$ существует в точности $\binom{r+k-1}{r-1} = \binom{r+k-1}{k}$.

Заметим, что неудачи располагались между последовательными успехами – пусть α_1 – число неудач до первого успеха, α_2 – число неудач от первого успеха до второго и так далее. Мы видим, что $\beta = \alpha_1 + \alpha_2 + \ldots + \alpha_r$. Теперь осталось заметить, что случайные величины α_i взаимно независимы и каждая из них имеет геометрическое распределение.

Производящая функция

Согласно свойству производящей функции для суммы независимых случайных величин производящая функция распределения Паскаля совпадает с

$$\left(\frac{p}{1-qz}\right)^r$$

Основные характеристики

Представление $\beta = \sum \alpha_k$ сразу дает ответ для математического ожидания распределения Паскаля: $\mathbb{E}(\xi) = rq/p$. Дисперсия суммы независимых случайных величин равна сумме дисперсий слагаемых, выразите ее в явном виде самостоятельно.

практике.

6.5 Распределение Пуассона

Это дискретное неотрицательное целочисленное распределение, принимающее значения k с вероятностью $\frac{\lambda^k}{k!}e^{-\lambda}$.

γ	0	$\ldots$	k	$\ldots$
	$e^{-\lambda}$	$\ldots$	$\frac{\lambda^k}{k!}e^{-\lambda}$	$\ldots$

Вспоминая разложение экспоненты в ряд нетрудно видеть, что производящая функция пуассоновского распределения с параметром λ равна $\exp(\lambda(z-1))$ (проверить!). Заметим, что из этого вытекает, что сумма двух *независимых* пуассоновских величин с параметрами λ и μ будет опять пуассоновской случайной величиной уже с параметром $\lambda+\mu$. Зная явный вид производящей функции несложно проверить, что математическое ожидание пуассоновского распределения равно его дисперсии и равно λ.

Предельная теорему Пуассона

Распределение Пуассона происходит из следующей предельной теоремы.

Теорема 6.1. *Зафиксируем $\lambda > 0$. Пусть $\xi_n \sim Binomial(n, p_n)$, где $p_n = \lambda/n + o(1/n)$. Тогда*

$$\mathbb{P}(\xi_n = m) \to \mathbb{P}(\xi = m) \quad \text{при} \quad n \to \infty \quad \forall m = 0, 1, \ldots,$$

где $\xi \sim Poisson(\lambda)$.

Имеем

$$\mathbb{P}(\xi = m) = \binom{n}{m} p_n^m (1-p_n)^{n-m} = \frac{n(n-1)\ldots(n-m+1)}{m!}\left(\frac{\lambda}{n} + o(1/n)\right)^m \left(1 - \frac{\lambda}{n} - o(1/n)\right)^{n-m}.$$

Теперь заметим, что $(\lambda/n + o(1/n))^m = (\lambda/n)^m + o(1/n^m)$, и далее

$$\lim_{n\to\infty} n\ldots(n-m+1)(\lambda/n)^m = \lambda^m, \quad \text{а также} \lim_{n\to\infty}\left(1 - \frac{\lambda}{n} - o(1/n)\right)^{n-m} = e^{-\lambda}$$

(второй замечательный предел). □

Это вычисление взято из книги В.Тутубалин "Теория вероятностей и случайных процессов. Основы математического аппарата и прикладные аспекты", Глава 1, раздел 4.3, в котором стоит посмотреть любопытные примеры интерпретаций формулы Пуассона на практике.

Возникает вопрос о скорости сходимости в теореме Пуассона (и в каком смысле ее понимать). Приведем без доказательства следующий результат, принадлежащий Ю.В. Прохорову, см. [Shi-1, параграф 1.6.4].

Теорема 6.2. *В условиях теоремы Пуассона 6.1,*

$$\sum_{m=0}^{\infty} \left|\mathbb{P}(\xi_n = m) - \mathbb{P}(\xi = m)\right| \leq \frac{2\lambda}{n} \min(2, \lambda).$$

Практическая интерпретация распределения Пуассона

Типичное применение – оценка вероятности возникновения нескольких маловероятных событий. Например, известно, что вероятность дефекта у типовой детали равна $p \ll 1$, тогда вероятность обнаружить в партии из m деталей менее k дефектных оценивается величиной $\mathbb{P}(\xi < k) = p_0 + p_1 + \ldots + p_{k-1}$, где $\xi \sim Poisson(mp)$, так что $p_i = \frac{(mp)^j}{j!} e^{-mp}$. При этом используется соображение, что математическое ожидание числа дефектных деталей в партии из m деталей равно mp – соответствующему параметру распределения Пуассона, совпадающему как указано выше с его математическим ожиданием.

Другой пример: нам сообщают, что вероятность неисправности самолета равна 10^{-6} за один час полета, то в принципе ясно что такое число могло возникнуть как частное $p = n_T/T$ от деления общего числа зафиксированных неисправностей n_T на большое количество летных часов T. Тогда, поскольку доставшийся нам параметр $\mu = 10^{-6}$ весьма мал и относится к схеме «повторений опыта каждый час» разумно считать, что для наблюдений за время $t \ll T$ часов имеет место распределение Пуассона с параметром $\lambda = \mu t$, так что можно оценивать вероятности возникновения, скажем, более трех неисправностей за время t (что может служить основой для чрезвычайной ситуации). Важным приложением также является учет вероятности того, что за время $t \ll T$ часов не возникнет ни одной неисправности, это, естественно, вычисляется по формуле $e^{-\mu t}$, что обычно понимается как *вероятность того, что время ожидания неисправности превосходит* t. Такая модель для появления редких событий во времени имеет наименование *пуассоновского потока событий с параметром* μ.

7 Важнейшие непрерывные распределения

Здесь предлагается минимально-необходимый список абсолютно непрерывных случайных величин (задаваемых с помощью их плотностей).

7.1 Равномерные распределение

1. **Равномерное** на отрезке $[a, b]$ распределение

$$F_\xi(x) = \begin{cases} 0 & x \leqslant a \\ \frac{x-a}{b-a} & a < x \leqslant b \\ 1 & x > b \end{cases}$$

его также можно определить через плотность:

$$\boxed{p_\xi(x) = \begin{cases} 0 & x < a \\ \frac{1}{b-a} & a \leqslant x \leqslant b \\ 0 & x > b \end{cases}}$$

Это вероятностная модель для встроенного в язык C и `openssl` компьютерного генератора `rand` псевдослучайных чисел. Пишут $\xi \sim Uniform([a, b])$.

2. Равномерное распределение в общем многомерном случае рассматривалось в примерах 4.4 и 4.11.

7.2 Гауссовское или нормальное распределение

Гауссова или **нормально распределенная случайная величина** принимает любые вещественные значения, закон распределения задается формулой плотности, участвуют два параметра m и $\sigma > 0$:

$$\boxed{p_\xi(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}}$$

Пишут $\xi \sim \mathcal{N}(m, \sigma^2)$. Можно показать, что

$$\mathbb{E}\xi = m, \qquad \operatorname{Var}\xi = \sigma^2.$$

Случай $\sigma = 0$ удоно включить в класс нормальных распредлений, в этом ситуации принимают $\xi = m$ п.н.

Распределение $\mathcal{N}(0, 1)$ называют *стандартным* нормальным распределением.

Неформальный смысл вытекает из Центральной предельной теоремы, которая появится в наших лекциях позже. Она утверждает, что гауссова случайная величина – предел *нормализованных* сумм очень большого числа независимых одинаково распределенных случайных величин, независимо от распределения этих случайных величин.

7.3 Показательное или экспоненциальное распределение

с параметром $\lambda > 0$ имеет плотность

$$\boxed{p_\xi(x) = \begin{cases} 0 & x < 0 \\ \lambda e^{-\lambda x} & x \geqslant 0 \end{cases}}$$

Пишут $\xi \sim Exp(\lambda)$. Такое распределение так же называется *экспоненциальным.*

Неформальный смысл: Случайная величина измеряет длину интервала времени между наступлениями «успехов» при «непрерывных» повторениях опыта, когда «успех» маловероятен настолько, что среднее число «успехов» за единичное время равно λ.

7.4 Логнормальное распределение

Экспонента гауссовой $\mathcal{N}(m, \sigma^2)$ случайной величины, называется логнормальной случайной величиной. Элементарное упражнение дает явную формулу для ее плотности:

$$\boxed{f(x) = \begin{cases} 0 & x \leqslant 0 \\ \frac{1}{x\sigma\sqrt{2\pi}} \exp\left[-\frac{(\ln x - m)^2}{2\sigma^2}\right] & x > 0 \end{cases}}$$

7.5 Семейства распределений

1. **Гамма распределения** Как известно из курса математического анализа гамма-функция от аргумента $t > 0$ определена следующей формулой

$$\Gamma(t) = \int_0^\infty u^{t-1} e^{-u} du.$$

В комплексном анализе рассказывается как она аналитически продолжается на другие комплексные значения аргумента. Известно, что гамма-функция интерполирует значения факториала, а именно при натуральном k

$$\Gamma(k+1) = k!$$

Из определения гамма функции следует, что при любых $b > 0$, $r > 0$ интеграл функции положительного аргумента

$$\boxed{p_{[b,r]}(x) = \frac{1}{\Gamma(r)} b^r x^{r-1} e^{-bx}, \quad x > 0,}$$

равен 1, а потому эта функция задает плотность распределения, называющуюся *гамма-плотностью с параметром r и масштабным параметром b.* Пишут $\xi \sim \Gamma(r, b)$.

При $r = 1$ соответствующая гамма-плотность совпадает с плотностью показательного распределения с параметром b. Упражнение по математическому анализу на вычисление несобственных интегралов показывает, что математическое ожидание для гамма-распределения с параметром r и масштабным параметром b — это r/b, а дисперсия, соответственно, r/b^2.

Предложение 7.1. *Семейство гамма-плотностей замкнуто относительно операции свертки:*

$$f_{[b,r_1]} * f_{[b,r_2]} = f_{[b,r_1+r_2]}$$

Доказательство. Прямое вычисление свертки дает необходимое выражение □
□

Согласно лемме 4.28 это означает, что если $\xi_j \sim Gamma(r_j, b)$, $j = 1, 2$, и ξ_1, ξ_2— независимы, то $\xi_1 + \xi_2 \sim Gamma(r_1 + r_2, b)$.

2. **Распределения** χ^2 - суммы квадратов независимых стандартных нормальных. Рассмотрим распределение квадрата нормальной случайной величины $\mathcal{N}(0, 1)$. Прямым вычислением проверяется,что плотность распределения квадрата нормальной случайной величины $\mathcal{N}(0, 1)$ при любом x совпадает с точностью до постоянного коэффициента с $f_{[\frac{1}{2},\frac{1}{2}]}(x)$, а значит, этот коэффициент равен единице и мы вдобавок получаем, что $\Gamma(\frac{1}{2}) = \sqrt{\pi}$. Воспользовавшись Предложением 7.1 мы сразу получаем, что сумма n квадратов независимых нормальных случайных величин распределена с плотностью $f_{[\frac{1}{2},\frac{n}{2}]}(x)$ — это и есть формула плотности распределения χ^2 с n степенями свободы.

3. **Распределения Коши.** Для двух независимых стандартных нормальных (то есть распределенных как $\mathcal{N}(0, 1)$) случайных величин α, β отношение α/β распределено

по закону Коши. Общее же определение плотности Коши (с масштабным параметром $t > 0$) таково:

$$\boxed{p_\xi(x) = \frac{t}{\pi}\frac{1}{t^2 + x^2}}$$

Пишут $\xi \sim Cauchy(t)$.

Важно, что свертка плотности Коши с параметром u и плотности Коши с параметром v явно вычисляется несложным интегрированием, и в ответе получается плотность Коши с параметром $u+v$. Значит, сумма независимых с.в. с распределениями Коши с параметрами u и v также имеет распределение Коши с параметром $u+v$. Применительно к статистическим формулам это означает, что усреднение i.i.d. в последовательности распределенных по закону Коши с.в. также распределено по закону Коши!

Из формулы для плонтости видно, что распределение Коши не имеет моментов: уже интеграл $\int_{\mathbb{R}} x\, p_\xi(x)\, dx$ не определен в смысле Лебега.

8 Вокруг закона больших чисел и центральной предельной теоремы

8.1 Неравенства Маркова и Чебышёва

Лемма 8.1 (Неравенство Маркова). *Пусть ξ - неотрицательная случайная величина, у которой есть математическое ожидание. Тогда*

$$\mathbb{P}\{\xi \geqslant R\} \leqslant \frac{\mathbb{E}\xi}{R}, \qquad \forall R > 0.$$

Доказательство. $\mathbb{E}(\xi) \geqslant \mathbb{E}\left(\xi \cdot \mathbb{I}_{\{\xi \geqslant R\}}\right) \geqslant R\,\mathbb{E}\left(\mathbb{I}_{\{\xi \geqslant R\}}\right) = R\,\mathbb{P}\{\xi \geqslant R\}$. □

Напомним, что существование второго момента $\mathbb{E}\xi^2$ влечет существование матожидания $\mathbb{E}\xi$.

Следствие 8.2 (Неравенство Чебышёва). *Пусть случайная величина ξ такова, что $\mathbb{E}\xi^2 < \infty$. Тогда*

$$\mathbb{P}\{|\xi - \mathbb{E}\xi| \geqslant R\} \leqslant \frac{\operatorname{Var}\xi}{R^2} \qquad \forall R > 0.$$

Доказательство.

$$\mathbb{P}\{|\xi - \mathbb{E}\xi| \geqslant R\} = \mathbb{P}\left\{(\xi - \mathbb{E}\xi)^2 \geqslant R^2\right\} \leq \frac{\mathbb{E}(\xi - \mathbb{E}\xi)^2}{R^2} = \frac{\operatorname{Var}\xi}{R^2},$$

согласно неравенству Маркова. □

Пример 8.3 («Правило трех сигм»)**.** Общепринято обозначение $\sigma^2 := \operatorname{Var}\xi$, причем $\sigma > 0$. Если в неравенстве Чебышёва положить $R = 3\sigma$, то получается универсальная оценка $\frac{8}{9}$ для вероятности случайной величине оказаться не далее 3σ от своего математического ожидания. Для конкретных распределений эта универсальная оценка может быть существенно улучшена. Например, для нормального распределения 8/9 заменяется на $1 - 0.0028$.

Следствие 8.4 (Неравенство Чернова). *Пусть случайная величина ξ такова, что $\mathbb{E}e^{t\xi} < \infty$ для любого $t \geq 0$. Тогда*

$$\mathbb{P}\{\xi \geqslant \varepsilon\} \leqslant \inf_{t \geqslant 0} e^{-t\varepsilon}\mathbb{E}\left(e^{t\xi}\right)$$

Доказательство. Для $t > 0$ по монотонности экспоненты $\mathbb{P}(\xi \geqslant \varepsilon) = \mathbb{P}\left(e^{t\xi} \geqslant e^{t\varepsilon}\right)$, далее используем неравенство Маркова. □

8.2 Закон больших чисел в форме Чебышёва

Пусть последовательность случайных величин (ξ_n) и случайная величина ξ определены на одном и том же вероятностном пространстве.

Определение 8.5. Говорят, что последовательность (ξ_n) сходится к ξ *по вероятности*, если для всякого $\varepsilon > 0$,

$$\mathbb{P}\{|\xi_n - \xi| \geq \varepsilon\} \to 0 \quad \text{при} \quad n \to \infty.$$

Пишут $\xi_n \xrightarrow{\mathbb{P}} \xi$ при $n \to \infty$.

Это ни что иное как известная из теории меры сходимость по мере.

Теорема 8.6 (Закон больших чисел)**.** *(1) Пусть случайные величины $\xi_1, \xi_2, \ldots$ независимы и их дисперсии равномерно ограничены, то есть существует такая постоянная C, что $\operatorname{Var} \xi_j < C$ для всех j. Тогда случайная величина $S_n := \sum_{j=1}^{n} \xi_j$ удовлетворяет*

$$\frac{S_n - \mathbb{E}S_n}{n} \xrightarrow{\mathbb{P}} 0 \quad \textit{при} \quad n \to \infty.$$

(2) Пусть случайные величины $\xi_1, \xi_2, \ldots$ — iid [19] *и $\mathbb{E}\xi_j^2 < \infty$. Тогда*

$$\frac{S_n}{n} \xrightarrow{\mathbb{P}} \mathbb{E}\xi_1 \quad \textit{при} \quad n \to \infty.$$

Отметим, что условия второго пункта усиливают условия первого. На самом деле, сходимость в законе больших чисел можно усилить (имеет место сходимость *почти наверное*), а условия ослабить (достаточно существования только первого момента). Мы займемся этими вопросами позже.

Доказательство. (1) Согласно неравенству Чебышёва,

$$\mathbb{P}\left(\left|\frac{S_n}{n} - \mathbb{E}\frac{S_n}{n}\right| \geq \varepsilon\right) \leq \frac{\operatorname{Var}(S_n/n)}{\varepsilon^2} = \frac{\operatorname{Var} S_n}{\varepsilon^2 n^2}. \tag{8.1}$$

Так как в силу независимости $\operatorname{Var} S_n = \sum_{j=1}^{n} \operatorname{Var} \xi_j \leq Cn$, правая часть выражения (8.1) стремится к нулю при $n \to \infty$.

(2) Так как ξ_j одинаково распределены, $\mathbb{E}S_n = n\mathbb{E}\xi_1$. Теперь утверждение следует из пункта (1). □

Пример 8.7. Пусть независимые случайные величины $\xi_1, \xi_2, \ldots$ имеют распределение $Bernoulli(p)$ (орел/решка, вероятность орла $= p$). Закон больших чисел утверждает, что $S_n/n \to \mathbb{E}\xi_1 = p$ по вероятности.

Так как S_n — ни что иное как число выпавших орлов, так что S_n/n — частота выпадения орлов за n бросков, мы получаем, что **частота успеха** (=орлов) **стремится к вероятности успеха**.

Итак, мы впервые нашли аккуратную связь между интуитивно близкими понятиями — частотой и вероятностью. Продемонстрируем эту связь в общем виде.

[19] Independent identically distributed, т.е. независимы и одинаково распределены.

Пример 8.8. Пусть $\xi_1, \xi_2, \dots$ — iid случайные величины, $A \in \mathcal{B}(\mathbb{R})$, а

$$\nu_n^A(\omega) := |\{1 \le k \le n : \xi_k(\omega) \in A\}|$$

— случайная величина, описывающая число тех испытаний среди первых n штук, результат которых лежит в множестве A. Заметим, что

$$\nu_n^A = \sum_{k=1}^{n} \mathbb{I}_A(\xi_k),$$

где $\mathbb{I}_A$ — индикатор множества A. Так как случайные величины $\mathbb{I}_A(\xi_k)$ тоже iid, закон больших чисел влечет сходимость

$$\frac{\nu_n^A}{n} \xrightarrow{\mathbb{P}} \mathbb{E}\,\mathbb{I}_A(\xi_1) = \mathbb{P}(\xi_1 \in A) \quad \text{при} \quad n \to \infty.$$

В левой части последнего выражения написано ни что иное как частота попадания в множество A.

О неверных интерпретациях ЗБЧ

Часто ссылаются на закон больших чисел и приводят утверждения, которые определенно из него не следуют. Если Петр и Павел по очереди бросают правильную монету 10 000 раз, то почему-то принято считать, что Петр будет в выигрыше приблизительно половину времени. Но это совсем не так: детальное изучение данной ситуации устанавливает, что такое равновесие *менее всего* вероятно. Вероятность того, что Петр будет в выигрыше не более 20 раз, несравненно больше вероятности того, что число игр, после которых он будет впереди, заключено между 4990 и 5010. Не существует никакой тенденции к выравниванию периодов лидерства. Закон больших чисел устанавливает только то, что для большого числа *различных* игр с бросанием монеты доля тех из них, в которых в данный момент в выигрыше находится герб, близка к $\frac{1}{2}$ и ничего не говорится о колебаниях лидерства в отдельной игре.

См. лекцию, где обсуждался закон арксинуса (сил писать подробно нет, а на лекции мы детально обсуждали этот феномен, см. также параграф 9.1); также далее появится пример про дискретное случайное блуждание.

8.3 Некоторые применения закона больших чисел

Метод Монте-Карло

В этом разделе мы обсудим важное приложение закона больших чисел — метод Монте-Карло. Поставим задачу численного нахождения определенного интеграла от непрерывной функции $f : \mathbb{R}^d \mapsto \mathbb{R}$

$$I := \int_U f(x)\,dx, \quad \text{где } U \text{ — ограниченная область в } \mathbb{R}^d.$$

Если d велико (скажем, $d = 10$), то классические методы работают очень плохо (см. лекцию). Необходимость же вычисления таких интегралов возникает спрошь и рядом, например, в физике. На помощь приходит теория вероятностей.

Пусть $\xi_1, \xi_2, \dots$ — последовательность независимых случайных величин, имеющих равномерное распределение в U. Тогда, согласно закону больших чисел,

$$\frac{1}{n}\sum_{i=1}^{n} f(\xi_i) \xrightarrow{\mathbb{P}} \mathbb{E}f(\xi_1) = \frac{1}{|U|}\int_U f(x)\,dx \quad \text{при} \quad n \to \infty,$$

так как $\xi_1 \sim Uniform(U)$. Как видно из доказательства ЗБЧ, скорость сходимости зависит только от дисперсии $\operatorname{Var}\xi_1$. Но

$$\operatorname{Var}\xi_1 = \mathbb{E}f(\xi_1)^2 - (\mathbb{E}f(\xi_1))^2 \le 2\max_{x\in U}|f(x)|^2.$$

Таким образом, **скорость сходимости в методе Монте-Карло не зависит от размерности d.**

Предложенный метод Монте-Карло начинает плохо работать, когда функция f принимает очень большие значения на подмножестве в U, имеющем малую меру. В этом случае, если n не слишком велико, то вероятность, что хотя бы один из $\xi_1,\dots,\xi_n$ попадет в это подмножество, мала, а значит среднее $\frac{1}{n}\sum_{i=1}^n f(\xi_i)$ при таких n может сильно отклониться от интеграла I. Эту проблему решает более хитрая версия метода Монте-Карло, использующая цепи Маркова. Вместо последовательности независимых равномерных случайных величин $\xi_1,\xi_2,\dots$, строится последовательность, образующая марковскую цепь, которая имеет специальное стационарное состояние.

Вероятностное доказательство теоремы Вейерштрасса

Пусть $f:[0,1]\mapsto\mathbb{R}$ — непрерывная функция. Рассмотрим ассоциированные с ней *полиномы Бернштейна*

$$B_n(x) := \sum_{k=0}^n f\Big(\frac{k}{n}\Big)C_n^k x^k(1-x)^{n-k}, \qquad 0\le x\le 1.$$

Теорема 8.9 (теорема Вейерштрасса)**.** *Многочлены $B_n(x)$ сходятся к функции $f(x)$ при $n\to\infty$ равномерно на $[0,1]$.*

Нижеследующее доказательство было предложено С.Н. Бернштейном.

Идея доказательства. Пусть S_n^x — случайная величина, имеющая биномиальное распределение [20] с параметрами (n,x), а $\eta_n^x := S_n^x/n$. Имеем

$$x = \mathbb{E}\eta_n^x \quad \text{и} \quad B_n(x) = \mathbb{E}f(\eta_n^x).$$

Согласно закону больших чисел, $\eta_n^x \mapsto x$ при $n\to\infty$ (это лишь идея доказательства, поэтому мы не уточняем в каком именно смысле имеет место сходимость). Так как f — непрерывна, находим $f(\eta_n)\to f(x)$. Вынося предел из-под интеграла, получаем $\mathbb{E}f(\eta_n)\to\mathbb{E}f(x)=f(x)$.

Доказательство. Записывая $f(x)$ в виде $\mathbb{E}f(x)$, занося модуль под матожидание и затем используя, что для всякого $\delta>0$ случайные величины $\mathbb{I}_{\{|\eta_n^x-x|\le\delta\}}$ и $\mathbb{I}_{\{|\eta_n^x-x|>\delta\}}$ в сумме дают единицу, находим

$$|B_n(x)-f(x)| \le \mathbb{E}|f(\eta_n^x)-f(x)| = \mathbb{E}\Big(|f(\eta_n^x)-f(x)|\mathbb{I}_{\{|\eta_n^x-x|\le\delta\}}\Big) + \mathbb{E}\Big(|f(\eta_n^x)-f(x)|\mathbb{I}_{\{|\eta_n^x-x|>\delta\}}\Big)$$
$$=: I_1 + I_2.$$

Непрерывная функция на отрезке равномерно непрерывна на нем, а значит для всякого $\varepsilon>0$ можно подобрать δ так, чтобы $|f(x)-f(y)|\le\varepsilon$ при $|x-y|\le\delta$. Тогда

$$I_1 \le \varepsilon\,\mathbb{E}\mathbb{I}_{\{|\eta_n^x-x|\le\delta\}} \le \varepsilon.$$

[20] Напоминание: другими словами, распределение S_n^x совпадает с распределением числа выпавших орлов при n независимых бросках монетки с вероятностью орла равной x.

Так как непрерывная функция ограничена на отрезке, для всяких $x, y \in [0,1]$ имеем $|f(y) - f(x)| \le 2\|f\|_\infty$, где $\|f\|_\infty = \max_{x\in[0,1]} |f(x)|$. Следовательно,

$$I_2 \le 2\|f\|_\infty \mathbb{E}\mathbb{I}_{\{|\eta_n^x - x| > \delta\}} = 2\|f\|_\infty \mathbb{P}\big(|\eta_n^x - x| > \delta\big). \tag{8.2}$$

Так как [21] $\operatorname{Var}\eta_n^x = nx(1-x)$, неравенство Чебышёва, примененное к правой части (8.2), влечет оценку

$$I_2 \le 2\|f\|_\infty \frac{nx(1-x)}{n^2\delta^2} \le \frac{2\|f\|_\infty}{n\delta^2} \le \varepsilon.$$

при достаточно большом n. Таким образом, $I_1 + I_2 < 2\varepsilon$. □

Неравенство Хёфдинга

Лемма 8.10. *Для с.в. ξ такой, что $a \leqslant \xi \leqslant b$ и $\mathbb{E}(\xi) = 0$ и любого $t > 0$ выполнено $\mathbb{E}\left(e^{t\xi}\right) \leqslant \exp\left[\frac{1}{8}t^2(b-a)^2\right]$.*

Доказательство. Выпуклость экспоненты дает неравенство $e^{t\xi} \leqslant \frac{b-\xi}{b-a}e^{ta} + \frac{\xi-a}{b-a}e^{tb}$. Применяя к обеим частям неравенства оператор математического ожидания с учетом $\mathbb{E}(\xi) = 0$ имеем $\mathbb{E}\left(e^{t\xi}\right) \leqslant \frac{b}{b-a}e^{ta} - \frac{a}{b-a}e^{tb} = e^{h(t(b-a))}$, причем функцию h можно выразить явно

$$h(x) = \frac{a}{b-a}x + \ln\left[1 + \frac{a}{b-a}(1-e^x)\right]$$

и потому в нуле $h'(0) = h(0) = 0$ и вдобавок $\forall x > 0 \;\; h''(x) \leqslant \frac{1}{4}$.

Разложение Тейлора с остаточным членом поэтому дает требуемое

$$h(x) = h(0) + h'(0)\cdot x + \frac{h''(\hat{x})x^2}{2} \leqslant \frac{x^2}{8}$$

что при подстановке $x = t(b-a)$ и дает $\mathbb{E}\left(e^{t\xi}\right) \leqslant \exp\left[\frac{1}{8}t^2(b-a)^2\right]$ □

Предложение 8.11 (Неравенство Хёфдинга). *(1) Пусть случайные величины $\xi_1, \xi_2, \dots$ независимы и $a_i \leqslant \xi_i \leqslant b_i$ для некоторых постоянных a_i, b_i. Тогда для произвольных $\varepsilon > 0$ и $t > 0$ их усреднение $\bar{\xi}_n := \sum_{i=1}^n \xi_i/n$ удовлетворяет*

$$\mathbb{P}\left\{|\bar{\xi}_n - \mathbb{E}\bar{\xi}_n| \geqslant \varepsilon\right\} \leqslant 2e^{-nt\varepsilon}\prod_{i=1}^n \exp\left[\frac{t^2(b_i-a_i)^2}{8}\right].$$

(2) Если вдобавок все случайные величины ξ_i одинаково распределены, так что, в частности, $a_i = a$, $b_i = b$ $\forall i$, то

$$\mathbb{P}\left\{|\bar{\xi}_n - \mathbb{E}\xi_1| \geqslant \varepsilon\right\} \leqslant 2\exp\left[-\frac{2n\varepsilon^2}{(b-a)^2}\right].$$

с.в. минимум в правой части достигается при $t = \frac{4\varepsilon}{(b-a)^2}$ и потому

Доказательство. Обозначим $\eta_i = \xi_i - \mathbb{E}\xi_i$ и $\bar{\eta}_n = \sum_{i=1}^n \eta_i/n$. Запишем неравенство Чернова (следствие 8.4) для $t > 0$, учитывая свойство независимости:

$$\mathbb{P}\left\{\overline{\eta} \geqslant \varepsilon\right\} = \mathbb{P}\left\{\sum_{i=1}^n \eta_i \geqslant n\varepsilon\right\} \leqslant e^{-nt\varepsilon}\mathbb{E}\left[\exp\left(t\sum_{i=1}^n \eta_i\right)\right] = e^{-nt\varepsilon}\prod_i \mathbb{E}\left(e^{t\eta_i}\right)$$

По лемме 8.10 $\mathbb{E}\left(e^{t\eta_i}\right) \leqslant \exp\left[\frac{1}{8}t^2(b_i-a_i)^2\right]$, теперь, используя $\mathbb{P}\left\{|\overline{\eta}_n| \geqslant \varepsilon\right\} = \mathbb{P}\left\{\overline{\eta}_n \geqslant \varepsilon\right\} + \mathbb{P}\left\{-\overline{\eta}_n \geqslant \varepsilon\right\}$, получаем требуемое. □ □

[21] Тут хотелось бы сразу воспользоваться законом больших чисел, который гарантирует, что правая часть (8.2) стремится к нулю при $n \to \infty$, но так мы не получим равномерности сходимости по x. Поэтому придется его честно передоказать.

Пример 8.12 (концентрация для многомерного куба). Посмотрим, что дает неравенство Хёфдинга в применении к n независимым равномерным на $[-\frac{1}{2}, \frac{1}{2}]$ случайным величинам, которые можно считать координатными функциями для равномерного распределения в n-мерном кубе с ребром единица:

$$\mathbb{P}\Big\{\Big|\sum_{k=1}^{n} \xi_k \cdot \frac{1}{\sqrt{n}}\Big| \geqslant r\sqrt{n}\Big\} = P\Big\{\Big|\sum_{k=1}^{n} \frac{\xi_k}{n}\Big| \geqslant r\Big\} \leqslant 2e^{-2nr^2} \tag{8.3}$$

Параметр r здесь имеет смысл выбирать в промежутке $[0, 1/2]$, иначе вероятность занулится.

Таким образом, если n велико, ортогональная проекция многомерной случайной величины на главную диагональ (с единичным направляющим вектором $\Big(\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{n}}, \ldots \frac{1}{\sqrt{n}}\Big)$) куба задаёт на этой главной диагонали вероятностную меру, с очень быстро убывающей плотностью при относительном удалении r от начала координат (напомним, что длина этой диагонали $\sqrt{n}$), а, значит, вся эта равномерная в объеме куба мера в геометрическом смысле сосредоточена вблизи гиперплоскости, ортогональной главной диагонали куба и проходящей через начало координат. Так, если, взять точку из равномерного в кубе распределения, то вероятность обнаружить ее не вблизи этой плоскости крайне мала.

Разумеется,все эти рассуждения (в силу симметрии) можно соотнести с другими диагоналями (с направляющими векторами вида $\Big(\pm\frac{1}{\sqrt{n}}, \pm\frac{1}{\sqrt{n}}, \ldots \frac{1}{\sqrt{n}}\Big)$).

8.4 Центральная предельная теорема

В этом параграфе обсуждается ключевой результат курса — центральная предельная теорема (ЦПТ). Пусть случайные величины $\xi_1, \xi_2, \ldots$ независимы, одинаково распределены и имеют конечную дисперсию (эквивалентно, $\mathbb{E}\xi_j^2 < \infty$). Обозначим

$$S_n := \sum_{i=1}^{n} \xi_i, \qquad m := \mathbb{E}\xi_1, \qquad \sigma^2 := \operatorname{Var}\xi_1, \qquad \sigma > 0.$$

Зададимся вопросом, как устроена сходимость в законе больших чисел

$$\frac{S_n - \mathbb{E}S_n}{n} = \frac{S_n}{n} - m \xrightarrow{\mathbb{P}} 0 \quad \text{при} \quad n \to \infty?$$

Дисперсия величины $(S_n - \mathbb{E}S_n)/n$ равна $n\sigma^2/n^2 = \sigma^2/n$, [22] и в этом смысле $(S_n - \mathbb{E}S_n)/n$ сходится к нулю со скоростью $\sim 1/\sqrt{n}$. Нормируем эту случайную величину так, чтобы она оставалась порядка единицы при всех n. Точнее, разделим ее на корень из свей дисперсии, то есть рассмотрим с.в.

$$\eta_n := \frac{S_n - \mathbb{E}S_n}{\sqrt{\operatorname{Var}S_n}} = \frac{S_n - nm}{\sigma\sqrt{n}}, \qquad \operatorname{Var}\eta_n = 1.$$

Теорема 8.13 (центральная предельная теорема)**.** *Пусть случайные величины $\xi_1, \xi_2, \ldots$ независимы, одинаково распределены и $\mathbb{E}\xi_j^2 < \infty$. Тогда*

$$\eta_n \xrightarrow{d} \eta \quad \textit{при} \quad n \to \infty, \quad \textit{где} \quad \eta \sim \mathcal{N}(0, 1).$$

[22]Напомним, что сдвиг не меняет дисперсию, то есть $\operatorname{Var}(S_n - \mathbb{E}S_n) = \operatorname{Var}S_n = n\sigma^2$.

Значок $\xrightarrow{d}$ обозначает *сходимость по распределению*, а $\mathcal{N}(0,1)$ — стандартное нормальное распределение. Аккуратное определение такой сходимости будет дано позже, в разделе о сходимостях. В **данном** же контексте она означает следующее:

$$\mathbb{P}\big\{\eta_n \in (a,b)\big\} \to \mathbb{P}\big\{\eta \in (a,b)\big\} \quad \text{при} \quad n \to \infty \quad \forall a \le b. \tag{8.4}$$

В частности, поточечно сходятся функции распределения

$$F_{\eta_n}(x) \to F_\eta(x) \quad \forall x \in \mathbb{R}.$$

Напомним, что плотность с.в. $\eta \sim \mathcal{N}(0,1)$ известна, так что

$$\mathbb{P}(\eta \in A) = \frac{1}{\sqrt{2\pi}} \int_A e^{-x^2/2}\, dx.$$

Итак, центральная предельная теорема утверждает, что, **каким бы ни было** распределение случайных величин ξ_j, распределение отклонения среднего S_n/n от матожидания m всегда управляется нормальным распределением:

$$\frac{S_n}{n} - m \sim \frac{\sigma}{\sqrt{n}}\eta \sim \mathcal{N}(0, \sigma^2/n).$$

Доказательство ЦПТ в общем виде будет проведено позже, после обсуждения аппарата характеристических функций. С его помощью оно получается коротко и элегантно. Сейчас же, чтобы продемонстрировать "что происходит" , мы разберем гораздо более громоздкое доказательство, работающее только в частном случае, зато получающееся в результате прямых вычислений.

Теорема Муавра-Лапласа

Пусть ξ_j имеют распределение $\mathbb{P}(\xi_i = -1) = \mathbb{P}(\xi_i = 1) = \frac{1}{2}$. В этой ситуации ЦПТ называют *теоремой Муавра-Лапласа*, которую мы и докажем в этом параграфе. Достаточно провести доказательство для случая $A = [a,b]$ и четных n. Заметим, что $m = 0$ и $\sigma^2 = 1$.

Для вероятностей значений S_{2n} с помощью комбинаторных соображений нетрудно найти точную формулу (см. параграф 9.1):

$$\begin{aligned}\mathbb{P}(S_{2n} = 2k) &= \binom{2n}{n+k} 2^{-2n} = \frac{(2n)!}{(n+k)!(n-k)!} 2^{-2n} \\ &\sim \frac{(2n)^{2n}}{(n+k)^{n+k}(n-k)^{n-k}} \cdot \frac{\sqrt{2\pi}\sqrt{2n}}{\sqrt{2\pi(n+k)}\sqrt{2\pi(n-k)}} 2^{-2n},\end{aligned}$$

где мы использовали формулу Стирлинга $n! \sim n^n e^{-n}\sqrt{2\pi n}$. Упрощая,

$$\begin{aligned}&\frac{(2n)^{2n}}{(n+k)^{n+k}(n-k)^{n-k}} \cdot \frac{\sqrt{2\pi}\sqrt{2n}}{\sqrt{2\pi(n+k)}\sqrt{2\pi(n-k)}} 2^{-2n} = \\ &= \left(1+\frac{k}{n}\right)^{-n-k}\left(1-\frac{k}{n}\right)^{-n+k} \cdot \frac{1}{\sqrt{\pi n}} \left(1+\frac{k}{n}\right)^{-1/2}\left(1-\frac{k}{n}\right)^{-1/2} \\ &= \left(1-\frac{k^2}{n^2}\right)^{-n}\left(1+\frac{k}{n}\right)^{-k}\left(1-\frac{k}{n}\right)^{k} \cdot \frac{1}{\sqrt{\pi n}} \left(1+\frac{k}{n}\right)^{-1/2}\left(1-\frac{k}{n}\right)^{-1/2}\end{aligned}$$

Нам нужны два элементарных утверждения из математического анализа:

1. Если $c_j \to 0$, $a_j \to \infty$ и $a \cdot c_j \to \lambda = const$, то $(1+c_j)^{a_j} \to e^{\lambda}$. Действительно, при $x \to 0$ $\ln(1+x)/x \to 1$, поэтому $a_j \ln(1+c_j) \to \lambda$

2. Если $\max_{1 \leqslant j \leqslant n} |c_{j,n}| \to 0$ $\sum_{j=1}^{n} c_{j,n} \to \lambda$ и $\sup_n \sum_{j=1}^{n} |c_{j,n}| < \infty$, то $\prod_{j=1}^{n}(c_{j,n}+1) \to e^{\lambda}$ (это обобщение предыдущего)

Положим $2k/\sqrt{2n} = x$, при ограниченом x и $n \to \infty$ выполнено $k/n \to 0$ и $k = x\sqrt{n/2} \to \infty$. Известные пределы позволяют аппроксимировать значение вероятности:

$$\lim \left(1 - \frac{k^2}{n^2}\right)^{-n} = e^{x^2/2}, \quad \lim \left(1 + \frac{k}{n}\right)^{-k} = \lim \left(1 - \frac{k}{n}\right)^{k} = e^{-x^2/2} \quad \mathbb{P}(S_{2n} = 2k) \sim \frac{1}{\sqrt{\pi n}} e^{-x^2/2}$$

Это последнее утверждение называется **локальной предельной теоремой Муавра-Лапласа**. Следующий шаг — аппроксимировать вероятность

$$\mathbb{P}\left(a\sqrt{2n} \leqslant S_{2n} \leqslant b\sqrt{2n}\right) = \sum_{m \in [a\sqrt{2n}, b\sqrt{2n}] \cap \{2\mathbb{Z}\}} \mathbb{P}(S_{2n} = m)$$

Положим $m = x\sqrt{2n}$ и используем сумму аппроксимаций из локальной предельной теоремы. Поскольку расстояние между точками суммирования $x \in [a,b] \cap \{2\mathbb{Z}/\sqrt{2n}\}$ равно $\sqrt{\frac{2}{n}}$, то результат представим интегральной суммой и потому близок к соответствующему интегралу по отрезку:

$$\mathbb{P}\left(a\sqrt{2n} \leqslant S_{2n} \leqslant b\sqrt{2n}\right) \sim \sum_{x \in [a,b] \cap \{2\mathbb{Z}/\sqrt{2n}\}} \frac{1}{\sqrt{\pi}} e^{-x^2/2} \sqrt{\frac{2}{n}} \cdot \frac{1}{\sqrt{2}} \sim \int_a^b \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx.$$

Скорость сходимости в центральной предельной теореме

Возникает естественный вопрос о скорости сходимости в центральной предельной теореме (и о способе ее измерения). Для этого можно воспользоваться следующим результатом, который мы приведем без доказательства.

Теорема 8.14 (Неравенство Берри - Эссеена)**.** *В условиях центральной предельной теоремы 8.13, дополнительно предполагая, что* $\mathbb{E}|\xi_j|^3 < \infty$*, имеет место оценка*

$$\sup_{-\infty < x < \infty} \left|F_{\eta_n}(x) - F_{\eta}(x)\right| \leq C \frac{\mathbb{E}\left|\xi_1 - \mathbb{E}\xi_1\right|^3}{\sqrt{n}\sigma^3},$$

где F_{η_n} *и* F_{η} *обозначают функции распределения случайной величины* η_n *и нормально распределенной с.в.* η *соответственно.*

Доказательство этого результата можно найти в [Shi-1, параграф 3.11].

Без дополнительных предположений о природе суммируемых случайных величин порядок приведенной оценки не может быть улучшен. Отметим, что оптимальное значение абсолютной константы C до сих пор неизвестно, и заключается в промежутке $(2\pi)^{-1/2} \leq C < 0.7655$.

9 Связи вероятностной теории с естествознанием

Вопрос о том, какие вероятностные модели следует использовать для описаний природных процессов черезвычайно важен и далеко не прост. Например, в классической статистической физике, созданной Максвеллом и Больцманом, исследуют, например, молекулы газа, которые полашают различимыми. Это приводит к тому, что статистические формулы Максвелла-Больцмана для вероятностей распределения частиц с заданными энергиями выводились из модели, использующей число способов для разложения n различных шаров по k различимым ящиками. Если вспомнить школьные объяснения про модель идеального газа из N движущихся молекул-шариков массы m, то идеальному газу с фиксированной энергией E отвечает соотношение

$$\sum_{k=1}^{N} \frac{mv_k^2}{2} = \frac{m}{2} \sum_{k=1}^{N} v_k^2 = E$$

тем самым, речь идет о вероятностных свойствах точек на поверхности многомерной сферы S^{N-1} радиуса $\sqrt{E}$. Удивительные факты про эти свойства можно найти в Листке 1.

В атомной и субатомной физике микромира все оказалось сложнее: например, фотоны и атомные ядра в измерениях требуют иной модели. Она разработана Эйнштейном и Бозе, в ее основе лежит выражение для числа разложения n неразличимых частиц-шаров по k различимым областям-ящиками, говоря короче, в статистике Бозе-Эйнштейна частицы считаются неразличимыми друг от друга. Другие частицы микромира — например, электроны— оказались и неразличимы и вдобавок подчинены ограничительному правилу, что в одном ящике (тут, конечно же, надо еще объяснить что именно в теории элементарных частиц считается «ящиком», но наш курс не по физике и потому эта тема здесь не рассматривается) не может находиться более одной частицы – эти правила определяют статистику Ферми-Дирака.

Задачи проверки соответствия реальных экспериментов и той или иной модели составляют предмет Математической Статистики, а в Теории Вероятностей в основном рассматриваются уже заданные модели и для них выводятся правила исчисления вероятностей событий. Это замечание стоит помнить при первом изучении предмета: действительно, чтобы предложить модель для реальной ситуации требуется достаточно длинное и всесторонее исследование, поэтому популярные рассуждения типа «разъяснения парадокса Монти-Холла» (см. например, `https://ru.wikipedia.org/wiki/Парадокс_Монти_Холла`) по существу отношения к реальности имеют не больше, чем рассмотренный в самом начале пример с монетами и карманами.

9.1 Содержательный пример: модель случайного блуждания

Важная для приложений модель (одномерного, симметричного, дискретного) случайного блуждания заключается в следующем: рассматриваются конечные суммы в последовательности $\xi_1, \xi_2, \dots$ независимых и одинаково распределенных случайных величин, принимающих с равными вероятностями $1/2$ два значения $+1$ и -1, нас будут интересовать множества значений сумм $S_n = \xi_1 + \dots + \xi_n$. Последовательные целочисленные значения сумм S_k удобно представлять на графике движения частицы по прямой $0Y$ из точки ноль в дискретные моменты времени, то есть как ломаную линию, выходящую из начала координат и в каждый момент времени $t = 1, 2, \dots$ проходящую через точку с целочисленной ординатой y, таким через время t частица окажется в некоторой точке y оси $0Y$ $-t \leqslant y \leqslant t$. По построению все пути из t шагов равновероятны с вероятностью 2^{-t}, так что эта модель классическая. Такой путь-график отражает движение частицы, подвергающейся случайным воздействиям (перемещение за один шаг вниз или вверх), у этой казалось бы простой модели оказывается, тем не менее, много противоречащих интуиции свойств; в частности, свойства доли времени, которое наугад выбранная траектория проводит в положительной и соответственно в отрицательной части оси $0Y$.

Количество $N(t, y)$ путей ведущих из начала координат в точку y можно явно вычислить: действительно, разность $k - l$ количества шагов вверх и вниз равна y, а сумма

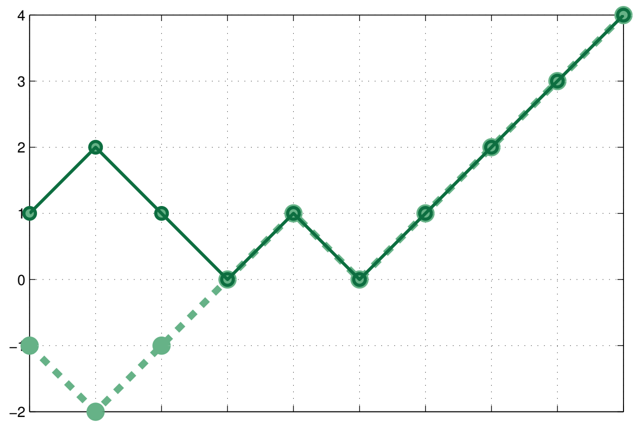


Рис. 2: Биективное соответствие путей, хотя бы раз принимающих значение 0, проходящих из точки с координатами $(t_1, 1)$ в точку $(t_2, 4)$ и всех вообще путей из точки $(t_1, -1)$ в точку $(t_2, 4)$

$k + l = t$.

- Проверьте, что формула для числа путей $N(t, y)$ из начала координат в точку y это $\binom{k+l}{k} = \binom{k+l}{l}$.

Посмотрим теперь на те пути, которые ведут в точку $y > 0$ и вдобавок в каждый момент времени $t = 1, 2, \dots$ принимают строго положительное значение, их будет уже меньше $N(t, y)$: во-первых, мы знаем, что на первом шагу путь идет в точку $+1$, во вторых, надо еще не учитывать те пути которые пересекали абциссу $y = 0$.

Лемма 9.1. *Путей из точки с координатами* $(1, 1)$ *в точку* (t, y) *хотя бы раз принимающих значение* 0 *столько же, сколько и всех путей из* $(1, -1)$ *в точку* (t, y).

Доказательство. Отразим симметрично как на Рис.2 часть пути до его *первого* попадания на абциссу, это задает требуемую биекцию. □

Таким образом, искомых «строго положительных» путей будет

$$N(t-1, y-1) - N(t-1, y+1) = \binom{k+l-1}{k-1} - \binom{k+l-1}{k+1} = \frac{k-l}{k+l}\binom{k+l}{k} = \frac{y}{t}N(t, y)$$

Пусть теперь $t = 2n$ четное число, тогда, рассуждение с указанной биекцией показывает, что

- количество путей из начала координат в точку $2n, 0$ *со строго положительными внутренними* вершинами будет $\frac{1}{n}\binom{2n-2}{n-1}$. Подсказка: $\frac{1}{2n-1}\binom{2n-1}{n-1} = \frac{1}{2n-1}\binom{2n-1}{n} = \frac{1}{n}\binom{2n-2}{n-1}$.
- количество путей из начала координат в точку $2n, 0$ *с неотрицательными внутренними вершинами* будет $\frac{1}{n+1}\binom{2n}{n}$. Подсказка: удалите из рассмотренного в предыдущей задаче пути его первый и последний сегмент и перенесите начало координат в точку $(1, 1)$ – это даст пути длины $2n - 2$ с неотрицательными внутренними вершинами, которых по предыдущей задаче в точности $\frac{1}{n}\binom{2n-2}{n-1}$, осталось поменять обозначения.

Разобранные методы непосредственного подсчета вероятностей при одномерном симметричном случайном блуждании позволяют объяснить ряд на первый взгляд странных свойств. Вам предлагается с помощью компьютерного эксперимента методом Монте-Карло увидеть один из таких как бы контринтуитивных эффектов. Например, казалось бы естественным предположить, что при симметричном блуждании типичные доли времени, приходящиеся на пребывании в положительной и отрицательной части прямой будут примерно равными для большинства траекторий. Попробуем проверить это, сгенерировав на компьютере большое число N траекторий длины $M \sim 100$, $N \gg M$ и указав для каждой число M_+ тех моментов времени, когда эта траектория была положительна. Интуиция симметрии подсказывает, что обычная траектория скорее всего примерно половину времени проводит в положительной полупрямой, проверим, так ли это, разбив возможный интервал $(0, M)$ на сегменты в числе, скажем $\left[\frac{\sqrt{N}}{3}\right]$ и для каждого сегмента указав количество сгенерированных траекторий со значением M_+ в этом сегменте. Эта процедура называется построением гистограммы для последовательных независимых испытаний, ей обычно сопоставляют столбчатую диаграмму. Упомянутое интуитивное соображение предсказывает, что средний столбик гистограммы должен бы оказаться выше прочих, но проделайте этот эксперимент сами и результат вас удивит: из него получается, что наугад выбранная траектория гораздо больше времени проводит в какой-то одной из половин, а траекторий с «интуитивно очевидным» поведением относительно мало.

Мы не будем здесь приводить аналитического объяснения странной формы гистограммы, формальный результат содержится в так называемлм Законе арксинуса[23].

Большие уклонения при случайном блуждании

Еще пример с одномерным дискретным случайным блужданием: рассматриваем сумму последовательных независимых шагов вправо-влево, то есть сумму независимых случайных величин, принимающих значения $\{+1, -1\}$ с вероятностями, соответственно $(p, 1-p)$. Разумеется, вероятности через n шагов оказаться на данном удалении от среднего *могут быть вычислены* через вероятности отклонений бернуллиевской с.в. $S_n = \sum_{i=1}^{n} \eta_i$, здесь η_k — независимые индикаторы, так что $\mathbb{E}(e^{t\eta}) = 1 - p + pe^t = g(t)$. В силу независимости слагаемых $\mathbb{E}(e^{tS_n}) = (g(t))^n$. Отсюда

$$\mathbb{P}\{S_n \geqslant n\varepsilon\} = \mathbb{P}\left\{\frac{S_n}{n} \geqslant \varepsilon\right\} \leqslant \inf_{t>0} \mathbb{E}\left[e^{t\left(\frac{S_n}{n}-\varepsilon\right)}\right] = \inf_{t>0} e^{-n\left[\frac{t}{n}\varepsilon - \ln g\left(\frac{t}{n}\right)\right]} =$$
$$\inf_{u>0} e^{-n[\varepsilon u - \ln g(u)]} = e^{-n\sup\limits_{u>0}[\varepsilon u - \ln g(u)]} = e^{-nr(\varepsilon)}$$

В статистической физике часто предпочитают иную (хоть и не вполне строгую) форму записи оценки вероятности уклонения: $\mathbb{P}\{S_n \simeq nx\} \sim \exp(-ns(x))$, при этом $s(x)$ называют **функцией Крамера**.

Энтропия как мера неопределенности

Для дискретной случайной величины с конечным набором вероятностей $p_1, p_2, \ldots p_m$ энтропией H называется

$$H = -\sum_{i=1}^{m} p_i \ln p_i$$

[23] Доказательство общего закона арксинуса можно посмотреть в учебниках Ширяева или Феллера

Ясно, что энтропия неотрицательна и равна нулю если только все вероятности, кроме одной, нулевые, то есть когда с.в. вырождается в константу (неслучайную величину). Наоборот, функция $f(x) = -x \ln x$ выпукла вверх и потому

$$\frac{f(x_1) + \ldots + f(x_m)}{m} \leqslant f\left(\frac{x_1 + \ldots + x_m}{m}\right)$$

$$H = -\sum_{i=1}^{m} p_i \ln p_i \leqslant -m \cdot \frac{p_1 + \cdots + p_m}{m} \cdot \ln\left(\frac{p_1 + \cdots + p_m}{m}\right) = \ln m$$

В частности, энтропия максимальна, когда все вероятности p_i равны друг-другу и вэтом смысле достигается максимальная неопределенность значений случайной величины.

Посмотрим теперь на модель траекторий дискретного случайного блуждания, необязательно симметричного. В вероятностном пространстве Ω траектории задают распределения вероятностей, которые (как показано выше) определяются формулой $P(\omega) = p^{\sum a_i} q^{n - \sum a_i}$. Траекторию назовем типичной, если $\sum a_i \sim np$ и $n - \sum a_i \sim nq$, $q = 1 - p$, и потому $P(\omega) \sim exp\,[-nH]$, где H – энтропия для распределения с вероятностями $p.q$. В силу ЗБЧ типичных траекторий большинство, а значит их число можно приблизить величиной e^{nH}.

Несложно вывести и обобщение данного результата, для случая повторений опыта с несколькими исходами вероятностей $p_1, p_2, \ldots p_k$, типичные траектории должны характеризоваться тем, что частоты в них близки к вероятностям, а оценка веса такой типичной траектории и их общего числа имеют тот же вид, что и в разобранном простейшем случае случайного блуждания.

10 Другие подходы в теории вероятностей. Неколмогоровские теории

Этот раздел включен в наш курс как дополнительный: действительно, разбираясь с как-будто бы хорошо известными примерами из естествознания мы часто встречаем аргументацию без явного упоминания модели вероятностного пространства. В каких-то случаях несложно такую модель восстановить, но и существуют эксперименты, в которых однозначный выбор модели затруднен или вообще невозможен (!). Краткие описания этих ситуаций приводятся ниже.

10.1 Частотный подход фон Мизеса

Изложение частотной теории вероятности Р. фон Мизеса опубликовано лет за десять до теории А.Н.Колмогорова и основано на конечных повторениях $\{s_1, ..., s_m\}$ одинаково организованного, но сохраняющего случайность в поведении, эксперимента s. Объектом изучения этой теории являются бесконечные последовательности, для которых выполнен принцип фон Мизеса. Он состоит в некоторой довольно сложно формулируемой стабилизации в бесконечной последовательности относительных частот $\nu_N(a) = \frac{n_N(a)}{N}$ для характеристики a эксперимента s, здесь $n_N(a)$ обозначает число проявлений характеристики a за первые N повторений. В частности, принцип статистической стабилизации относительных частот говорит: *если частота $\nu_N(a)$ стремится к пределу p_a, когда N стремится к бесконечности, то этот предел в частотной теории вероятностей полагают вероятностью характеристики a.* Таким образом само понятие вероятности в теории фон Мизеса основано на конструкции повторения эксперимента. Нестрого выражаясь: утверждение о сходимости частоты к вероятности в теории фон Мизеса было взято за аксиому, а прочие свойства вероятности пытались вывести (весьма громоздким образом) из этого утверждения. В теории А.Н.Колмогорова аксиомы формулируются иначе, а утверждение о связи частот и вероятностей при повторениях эксперимента превращается в теорему, известную под названием Закона Больших Чисел, (а также Усиленного Закона Больших Чисел) . Мы не будем впредь касаться теории фон Мизеса, хотя в научном мире у нее и сохраняется (узкий) круг стойких приверженцев.

Идея статистической проверки вероятностных моделей

Правильность значений $(p_1, p_2, \dots p_N)$, $0 \leqslant p_i \leqslant 1$ вероятностей элементарных событий назначенных для конкретной модели Ω может быть до известной степени проверена с помощью закона больших чисел, согласно которому в длинных сериях независимых экспериментов, происходящих при одинаковых условиях, частоты появления элементарных событий близки к их вероятностям. Важно при этом понимать, что такие методы могут лишь указать на странные отклонения наблюдаемых частот от предсказанных теорией: статистика может внести аргументы лишь для отклонения теоретической модели, но никак не может доказать ее правильность! Статистические оценки отклонений называются тестами, фактический список таких тестов очень велик. В этом смысле являются эвристическими все без исключения рекомендации *принятия* той или иной вероятностной модели для объяснения эксперимента.

10.2 Состояния и квантовая схема вероятностной модели.

Вероятности в конечном множестве элементарных исходов ω_k традиционной (не обязательно классической) модели задают простым списком $(p_1, p_2, \dots p_N)$. Если множество Ω из N элементарных исходов-событий зафиксировано, то все способы задания вероятностного пространства элементарных (и неэлементарных) событий можно описать при помощи следующей конструкции из линейной алгебры: пусть $0 \leqslant p_i = s_i^2 \leqslant 1$, рассмотрим в N-мерном евклидовом пространстве $L = L(\Omega)$ вектор $\mathbf{s}$ с координатами $(s_1, s_2, \dots s_N)$, физики склонны его называть *вектором состояния*, его длина равна единице потому что $\sum_i p_i = 1$. Всякому событию, составленному из элементарных исходов $\omega_{i_1}, \dots \omega_{i_m}$, сопоставим линейное подпространство $L(A) \subset L$ с ортонормальным базисом $\langle \omega_{i_1}, \dots \omega_{i_m} \rangle$, ортонормальный базис всего L будет при этом отождествляться с $\langle \omega_1, \dots \omega_N \rangle$

В евклидовом линейном векторном пространстве L определены ортогональные проекции векторов на подпространства. Заметим, что вероятность $\mathbb{P}(A)$, вычисленная по модели вероятностного пространства Ω совпадает с квадратом длины ортогональной проекции $|pr_A(\mathbf{s})|^2$ вектора состояния на линейное подпространство $L(A)$. Важно при этом, что так определенное семейство линейных подпространств однозначно отвечает семейству подмножеств в множестве всех элементов базиса и потому образует алгебру, которая неотличима от алгебры событий в Ω. Нетрудно видеть, что при таком понимании классическая модель отвечает вектору состояния, все координаты которого одинаковы и равны $1/\sqrt{N}$.

Это (на первый взгляд ненужное) усложнение и без того понятной традиционной конструкции помогает установить связь с понятием *квантовой вероятности*. Для исследования квантовых явлений понятия событий и случайных величин изложенная выше конструкция модифицируется: событиями называют теперь *произвольные линейные подпространства* $V \subset L(\Omega)$. Для физических применений существенно также, чтобы все эти линейные пространства рассматривалось над полем комплексных чисел $\mathbb{C}$ вместе с эрмитовым скалярным произведением, но пока можно не обращать внимания на эту деталь, поскольку здесь наша цель лишь в объяснении принципиальной разницы с традиционной теорией вероятностей. Объединение событий $U \cup V$ в квантовой теории вероятностей определяется как линейная оболочка объединения соответствующих подпространств, пересечение $U \cap V$ отвечает пересечению подпространств, противоположное событие $\overline{U}$ отвечает ортогональному дополнению $U^{\perp}$. Вероятность события U в квантовой теории вероятностей определяется через вектор состояния $\mathbf{s}$— вектор единичной длины в $L(\Omega)$ так же как и ранее: вероятность равна квадрату длины $|pr_U(\mathbf{s})|^2$ ортогональной проекции $pr_U(\mathbf{s}) \in U$, взятой «большом» линейном векторном пространстве $L(\Omega)$ вектора состояния $\mathbf{s}$ на линейное подпространство U. Физики, рассматривая комплексный случай, называют вектор $pr_U(\mathbf{s})$ *комплексной амплитудой события* U.

Ключевая разница между квантовой схемой и традиционной становится ясна из решения следующей (очень простой!) задачи:

Задача

Проверьте для событий квантовой теории вероятностей тождества

$$A \cap \left(\bigcup_i B_i\right) = \bigcup_i (A \cap B_i) \qquad \text{и} \qquad A \cup \bigcap_i B_i = \bigcap_i (A \cup B_i)$$

Специфика квантовых вероятностей

Таким образом в квантовой модели вероятностей события булеву алгебру относительно операций сложения, умножения и взятия противоположного события *уже не образуют*. В частности, для квантового случая невозможно указать полную группу событий, нет разбиения на «альтернативные гипотезы» и рассуждать в терминах условных вероятностей становится затруднительно.

Однако физический микромир оказался в удивительном согласии именно с квантовой схемой вычисления вероятностей, хотя некоторые свойства таких событий выглядят совершенно парадоксально, если неправильно интерпретировать их как элементы булевой алгебры. Кроме того, помимо обычной зависимости и независимости событий (понимаемых абсолютно в тех же равенствах как и в традиционной теории вероятностей) имеет место совершенно иная связь событий, называемая квантовой спутанностью.

Повторения эксперимента и квантовая спутанность

Если $L = L(\Omega)$ служит квантовой вероятностной моделью для одиночного эксперимента, то модели квантовых повторений сопоставляют линейное пространство $L^{\otimes m} = L \otimes L \otimes \ldots \otimes L$ – тензорное произведение пространств (а не пространство прямого произведения $L(\Omega) \times \ldots \times L(\Omega)$) как это предписано в традиционной теории вероятностей).

- Какую размерность имеет пространство событий трехкратного повторения, если $L(\Omega)$ двумерно?

Векторы состояния в пространстве $L^{\otimes m}$, отличные от вида $\mathbf{s} \otimes \mathbf{s} \otimes \ldots \otimes \mathbf{s}$, порождают для квантовой теории вероятности новые эффекты зависимостей между квантовыми событиями. Эти зависимости называются *квантовой спутанностью событий*, они, вообще говоря, никак не могут быть проинтерпретированы в терминах традиционной теории вероятностей. Популярные сегодня идеи квантовой защиты каналов информации и квантовой криптографии используют именно квантовую спутанность.

Квантовые и традиционные вероятности вокруг нас

В курсах физики правила исчисления вероятностей в квантовом случае приводятся как правило без мотивирующих объяснений. Поэтому следует помнить, что при рассмотрении физических квантовых эффектов употребляется вроде бы знакомая по традиционному подходу вероятностная терминология, но уже в совершенно новом значении; это порождает много известных из популярной литературы «парадоксов» микромира. Интересующиеся теорией исчисления квантовых вероятностей могут обратиться к первому разделу в целом достаточно трудной книги А.Холево "Вероятностные и статистические аспекты квантовой теории".

Наш вводный курс связан лишь с традиционной Теорией Вероятностей, но при этом надо четко представлять границы ее связей с реальностью: аксиоматический колмогоровский подход вовсе не покрывает всю сложность анализируемых явлений и каждый раз следует быть особенно аккуратным в выборе математической модели. Интересующимся этими вопросами можно предложить прочитать историю знаменитых неравенств Белла, которые оказываются справедливыми при классическом исчислении теории вероятностей, но явно нарушаются в квантовых экспериментах.

10.3 Байесовский подход

Байесовский подход в теории вероятностей вряд ли допускает строгое математическое обоснование: он по своей сути включает субъективное понятие вероятности. В общем и целом, байесовские методы не гарантируют проверяемую истинность и поэтому практическая статистика гораздо больше использует классические методы, хотя и байесовские идеи, безусловно, присутствуют как рецепт. В частности алгоритмы искусственного интеллекта и машинного обучения интенсивнее прочих используют байесовские методы. Начнем с изложения примера

Пример. Закон следования Лапласа

Пусть имеются $N+1$ урн, каждая из которых содержит N шаров; урна с номером k содержит k красных и $N-k$ белых шаров $k=0,1,2\dots N$. Из наугад выбранной урны n раз наудачу извлекают шары, причем вынутый шар каждый раз возвращают обратно. Предположим, что все n раз извлеченных шаров оказались красными (событие A). Найдем условную вероятность того, что следующее испытание тоже даст красный шар (событие B). Условная вероятность извлечь подряд n красных шаров из урны с номером k равна $(k/N)^n$. Отсюда,

$$\begin{aligned}\mathbb{P}(A) =& \frac{1^n+2^n+\dots N^n}{N^n(N+1)}\\ \mathbb{P}(AB)=\mathbb{P}(B) =& \frac{1^{n+1}+2^{n+1}+\dots N^{n+1}}{N^{n+1}(N+1)}\end{aligned}$$

Поскольку событие AB означает, что $n+1$ испытаний дали красные шары.

Искомая вероятность равна $\mathbb{P}(B|A)=\mathbb{P}(B)/\mathbb{P}(A)$ Суммы можно рассматривать как интегральные суммы по Риману, и поэтому при большом N

$$\frac{1}{N}\sum_{k=1}^{N}\left(\frac{k}{N}\right)^n \simeq \int_0^1 x^n dx = \frac{1}{n+1}$$

Следовательно, при большом N приближенно $\mathbb{P}(B|A)\simeq\frac{n+1}{n+2}$

Этой формуле можно дать следующее истолкование: если все возможные количества красных и белых шаров в урне равновероятны и если n испытаний дали красные шары, то вероятность появления красного шара при следующем испытании равна $\frac{n+1}{n+2}$. Это — так называемый закон следования Лапласа.

До развития современной теории понятие равновероятности часто отождествляли с «отсутствием предварительных сведений». Сам Лаплас для иллюстрации пользы своего закона вычислял вероятность того, что на следующий день взойдет Солнце, в предположении, что оно всходило ежедневно 5000 лет (или $n = 1826213$ дней подряд). Передают, что Лаплас готов был ставить 1826214 против 1 за то, что Солнце не изменит своего поведения; в наше время следовало бы увеличить ставку, поскольку регулярное движение Солнца наблюдалось в течение еще одного столетия. Чтобы отдать должное Лапласу и понять его намерения, потребовалось бы историческое исследование. Последователи Лапласа, однако, использовали аналогичные доводы в повседневной работе и рекомендовали физикам и инженерам применять такие методы в случаях, в которых формулы не имеют никакого действительного смысла. Даже если бы на минуту согласились допустить, что наша вселенная выбрана наугад из «множества вселенных», в котором все мыслимые возможности равновероятны, предполагаемый восход Солнца 5 февраля 3123 года до н. э. ничуть не более достоверен, чем то, что Солнце взойдет завтра; у нас совершенно одинаковые основания верить в оба эти события.

В чем состоит байесовский подход

Принято верить, что интерпретация условной вероятности в прикладных вопросах соответствует вероятностям в модели Ω, построенной для исходов с учетом знания условия. Но, вообще говоря, при разных условиях это могут быть разные модели, можно ли объединить их все в некоторой глобальной модели, зная условные вероятности? В практических приложениях иногда достаточно нетрудно охарактеризовать значения условных вероятностей, поэтому вероятность конкретного события бывает удобно находить через формулу полной вероятности. Благодаря второй формуле получил распространение эвристический *байесовский подход* поиска якобы наилучшей модели для описания эксперимента с известным результатом: выбранная группа разных моделей эксперимента трактуется как полная группа событий на неизвестном пока вероятностном пространстве и по формуле Байеса выбирают ту модель H_k, для которой условная вероятность $\mathbb{P}(H_k|A)$ наибольшая. Отметим, что при таком подходе вероятность считается *заведомо существующей безо всякой модели* и вопрос сводится лишь к тому как ее отыскать. В XX веке отношение к байесовским методам (кроме как в Англии) было весьма и весьма скептическим, поскольку в научной среде формула Байеса была дискредитирована метафизическими приложениями вроде описанного в законе следования Лапласа.

Но в наши дни этот подход реанимирован и широко используется при рассмотрении BigData, что внесло и вносит определенную путаницу в интерпретацию результатов, но необходимо признать, что множество реальных эффектов было выявлено именно в рамках байесовских методов. Так что основной проблемой с байесовскими методами следует считать отсутствие строгой теории, удовлетворяющей принятым в математике требованиям к логическим обоснованиям.

Практические соображения построения вероятностно-статистических моделей, которые мы обсуждали до сих пор, известны как частотные методы. Частотная точка зрения (разделяемая также фриквентистами в своем варианте теории вероятностей) основана на следующих постулатах:

1. Понятие вероятности связано с предельным поведением частот появления события при независимых повторениях и тем самым отражают свойства реального мира.

2. Параметры неизвестных распределений — это числа (или векторы), которые мы

предполагаем неизменными в идеальном эксперименте независимых повторений.

Напротив, подход, называемый байесовским, основан на следующих постулатах:

1. Вероятность описывает степень веры в истинность некоторого соотношения, а не предельную частоту. Допустимо делать вероятностные утверждения не только о данных в повторениях. Фразы типа «Вероятность того, что Альберт Эйнштейн выпил чашку чая 1 августа, 1948 г. »составляет 0,35» считаются осмысленными, так как отражают частную силу веры в истинность предложения. В частности, вероятностные утверждения допустимы для чисто-математических объектов, например для неизвестных констант.

2. Выводы байесовских рассуждений нацелены на то, чтобы точнее сформулировать законы распределения (случайной величины).

Попробуем разобрать конкретную байесовскую методику. Байесовский метод оценивания вероятностей (элементарных) исходов связан со следующей процедурой:

1. Выражаем распределение с.в. ξ через неизвестный параметр $\theta \in \mathbb{R}$. — это называется априорным распределением. Наши представления о возможных значениях параметра θ до того, как мы увидим какой-либо данные, в байесовском смысле равновероятны, это повод байесианцам называть θ случайной величиной.

2. Назначаем модель условных вероятностей при разных значениях параметра, то есть $p(\xi = a|\theta = b_j)$

3. По данным эксперимента над случайной величиной ξ имеем значения $x_1, x_2, \ldots x_n$, вычисляем в духе формулы Байеса апостериорное распределение параметра $p(\theta = b_j|\xi = x_1, \xi = x_2, \ldots \xi = x_n)$.

4. Выбираем значение параметра с максимальной апостериорной вероятностью и используем первый пункт. Всю указанную процедуру можно итерировать, байесианцы неясно почему предполагают, что от этого ответ о распределении ξ должен становиться точнее.

Чтобы увидеть, как выполняется третий шаг процедуры, сначала предположим, что θ дискретный параметр с конечным числом значений и что существует одно-единственное дискретное наблюдение x_1. Напоминаем, что параметр θ мы рассматриваем как случайную величину, а ее возможные значения обозначим t_k. Теперь, в этой дискретной версии видим формулу Байеса

$$\mathbb{P}(\theta = t_k|\xi = x_1) = \frac{\mathbb{P}(\theta = t_k, \xi = x_1)}{\mathbb{P}(\xi = x_1)} = \frac{\mathbb{P}(\xi = x_1|\theta = t_k)\mathbb{P}(\theta = t_k)}{\sum_i \mathbb{P}(\xi = x_1|\theta = t_i)\mathbb{P}(\theta = t_i)}$$

11 Метод характеристических функци

11.1 Определение и первые свойства

Пусть $\xi : \Omega \mapsto \mathbb{R}$ — случайная величина. Конструкция производящей функции была ограничена требованием целочисленности значений случайной величины ξ. Следующее определение вводит универсальный аналог производящей функции, подходящий как для целочисленных с.в., так и нет.

Определение 11.1. *Характеристической функцией* случайной величины ξ называется функция $\varphi_\xi : \mathbb{R} \mapsto \mathbb{C}$,

$$\varphi_\xi(t) = \mathbb{E}e^{it\xi}\,, \tag{11.1}$$

где i — мнимая единица.

Случайная величина $e^{it\xi}$ — комплекснозначная. Мы работали в основном с действительнозначными случайными величинами, но большинство теории переносится на комплексный случай дословно. Если это вас смущает, есть и другой путь: поскольку мы уже имели дело со случайными векторами, то в можно понимать комплекснозначную случайную величину $\zeta : \Omega \to \mathbb{C}$ как обычный двумерный случайный вектор с компонентами вещественной и мнимой части от ζ.

Чтобы лучше понять определение характеристической функции, разберем несколько частных случаев:

- Если ξ — целочисленная случайная величина, принимающая только неотрицательные значения, то

  $$\varphi_\xi(t) = e^{it0}\mathbb{P}(\xi = 0) + e^{it1}\mathbb{P}(\xi = 1) + \ldots + e^{itk}\mathbb{P}(\xi = k) + \ldots = Q_\xi(e^{it}),$$

  где Q_ξ обозначает производящую функцию ξ.

- В более общем случае, если ξ — произвольная целочисленная случайная величина, то характеристическая функция

  $$\varphi_\xi(t) = \sum_{k=-\infty}^{\infty} e^{ikt}\,\mathbb{P}(\xi = k) \tag{11.2}$$

  2π-периодична, а правая часть ее определения (11.2) представляет из себя ни что иное как ее **ряд Фурье**.

- Если ξ — абсолютно непрерывная случайная величина с плотностью p_ξ, то, записывая матожидание с помощью формулы (5.5), находим

  $$\varphi_\xi(t) = \int_{-\infty}^{\infty} e^{itx} p_\xi(x)\,dx. \tag{11.3}$$

  Мы видим, что в этом случае характеристическая функция есть ни что иное как **обратное преобразование Фурье** от плотности распределения (которая, напомним, интегрируема относительно Лебега, так что обратное преобразование Фурье хорошо определено).

  Разумеется, можно было бы определить характеристическую функцию не как обратное, а как прямое преобразование Фурье, и ее свойства от этого бы не изменились. Но в теории вероятностей и теории меры принято использовать формулу (11.3).

- В общем случае, применяя формулу (5.2), находим

  $$\varphi_\xi(t) = \int_{-\infty}^{\infty} e^{itx}\,d\mathbb{P}_\xi(x). \tag{11.4}$$

  Этот объект называется *преобразованием Фурье меры* $\mathbb{P}_\xi$.

Из указанной выше связи между преобразованием Фурье и характеристической функцией ясно, что последняя должна нести в себе очень много информации о распределении. В частности, из теоремы об обращении преобразования Фурье следует, что при дополнительных условиях регулярности, по характеристической функции абсолютно непрерывной случайной величины можно однозначно восстановить ее плотность. Если бы нас интересовали только целочисленные или абсолютно непрерывные случайные величины, для описания свойств их характеристических функций и связи последних с распределениями можно было бы просто сослаться на несколько (весьма нетривиальных) теорем из анализа об общих свойствах преобразования Фурье. Но уже для смешанных с.в. рассуждения усложняются, и потому в подробные курсы теории вероятностей традиционно включают множество общих теорем о характеристических функциях, которые в указанных частных случаях могут быть проинтерпретированы как утверждения из обычной теории преобразования Фурье.

В нашем коротком вводном курсе изложить эти теоремы со всеми доказательствами не представляется возможным, поэтому придется ограничиться довольно сжатым изложением с некоторыми пояснениями. Начнем с пояснений: почему вообще использование характеристических функций оказалось столь значимым для теории вероятностей. Три приведенные ниже пункта будут подробно обсуждаться в последующих параграфах.

Во-первых, как намекает нам связь с преобразованием Фурье, между распределениями и характеристическими функциями существует взаимно однозначное соответствие. Поэтому можно изучать свойства распределений, анализируя соответствующие характеристические функции — это так называемая теорема обращения, приведенная ниже, см. теорему 11.5.

Во-вторых, характеристические функции от суммы независимых случайных величин являются произведением характеристических функций слагаемых — это прямо следует из свойства математического ожидания произведения двух независимых случайных величин. Поэтому для последовательности независимых случайных величин $\{\xi_k\}$ и последовательности $\{S_n\}$ ее частичных сумм $S_n = \sum_1^n \xi_k$ характеристические функции имеют несложный вид: $\varphi_{S_n}(t) = \prod_1^n \varphi_{\xi_k}(t)$. См. теорему 11.8.

В третьих, важнейшим является то обстоятельство, что сходимость случайных величин по распределению, имеющая место, например, в центральной предельной теореме, *эквивалентна* поточечной сходимости $\varphi_n \to \varphi$ соответствующих характеристических функций. Это — основной инструмент для доказательства теорем о сходимости по распределению последовательностей случайных величин. См. теорему 12.12.

Пример 11.2. 1. Пусть $\xi \sim Bernoulli(p)$. Тогда, обозначая $q = 1 - p$, находим

$$\varphi_\xi(t) = \mathbb{E}e^{it\xi} = q + pe^{it}.$$

2. Пусть $\xi \sim Binomial(n,p)$. Подставляя $z = e^{it}$ в (6.2), находим $\varphi_\xi(t) = (q + pe^{it})^n$.

3. Пусть $\xi \sim \mathcal{N}(0,1)$. Имеем

$$\varphi_\xi(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-x^2/2+itx}\,dx = \frac{e^{-t^2/2}}{2\pi}\int_{-\infty}^{\infty} e^{-(x-it)^2/2}\,dx.$$

Так как интеграл от $e^{-z^2/2}$ по прямой $\Im z = it$ равен интегралу по действительной оси,

$$\varphi_\xi(t) = e^{-t^2/2}.$$

Мы видим, что характеристическая функция стандартной гауссовской случайной величины совпадает с ее плотностью. Это не удивительно: из курса анализа известно, что гауссовская экспонента является собственной функцией преобразования Фурье, что мы сейчас и передоказали. [24]

4. Пусть $\xi \sim \mathcal{N}(m, \sigma^2)$. Так как распределение ξ совпадает с распределением с.в. $\zeta = m + \sigma\eta$, где $\eta \sim \mathcal{N}(0, 1)$, имеем

$$\varphi_\xi(t) = \varphi_\zeta(t) = e^{itm}\mathbb{E}e^{it\sigma\eta} = e^{itm}\varphi_\eta(\sigma t) = e^{itm - \sigma^2 t^2/2}. \tag{11.5}$$

5. Характеристические функции остальных стандартных распределений можно легко найти в интернете, например, в Википедии.

В завершение первого знакомства с характеристическими функциями обсудим некоторых их свойства.

Лемма 11.3. *1. Для $a, b \in \mathbb{R}$, $\varphi_{a\xi+b}(t) = e^{ibt} \cdot \varphi_\xi(at)$.*

2. $|\varphi_\xi(t)| \leqslant \varphi_\xi(0) = 1$.

3. $\varphi_\xi(t) = \overline{\varphi_\xi(-t)}$.

4. $\varphi_\xi(t)$ равномерно непрерывна по аргументу $t \in \mathbb{R}$.

Доказательство. (1) Имеем

$$\varphi_{a\xi+b}(t) = \mathbb{E}\left(e^{it(a\xi+b)}\right) = e^{itb}\mathbb{E}\left(e^{ita\xi}\right) = e^{itb} \cdot \varphi_\xi(at).$$

(2,3) Следует непосредственно из определения.
(4) Имеем

$$|\varphi_\xi(t+h) - \varphi_\xi(t)| = \left|\mathbb{E}\big[e^{it\xi}(e^{ih\xi} - 1)\big]\right| \leqslant |\mathbb{E}(e^{ih\xi} - 1)| \underset{h\to 0}{\longrightarrow} 0,$$

по теореме Лебега о мажорируемой сходимости, см. теорему B.26 (**проверьте!** или посмотрите лекцию). ☐

Напомним, что для классического преобразования Фурье чем глаже функция, тем быстрее убывает ее преобразование Фурье, и наоборот. В следующей лемме мы доказываем аналогичную связь между гладкостью характеристической функции и "тяжестью хвостов распределения" (можно думать про скорость убывания плотности для абсолютно непрерывнвых случайных величин). Последняя выражается в терминах конечности моментов высоких порядков.

Лемма 11.4. *(1) Если $\mathbb{E}|\xi|^n < \infty$ для некоторого $n \in \mathbb{N}$, то функция φ_ξ является C^n-гладкой. Ее производные получаются дифференцированием в формуле (11.1) под знаком интеграла,*

$$\varphi^{(k)}(t) = \mathbb{E}\big[(i\xi)^k e^{it\xi}\big], \quad \text{и в частности} \quad \varphi^{(k)}(0) = i^k\mathbb{E}\xi^k. \tag{11.6}$$

Более того, в формуле Тейлора явно контролируется остаток:

$$\varphi_\xi(t) = \sum_{k=0}^{n} \frac{(it)^k}{k!}\mathbb{E}\xi^k + \varepsilon(t)\frac{(it)^n}{n!}, \tag{11.7}$$

[24] С этим связано появление гауссовского распределения в центральной предельной теореме.

где $|\varepsilon(t)| \leqslant 3\mathbb{E}|\xi|^n$ *и* $\lim_{t\to 0}\varepsilon(t) = 0$.

(2) Пусть все моменты $\mathbb{E}\xi^n$ *конечны и растут не слишком быстро — точнее, для некоторого* $t_0 > 0$ *выполнено* $t_0^n\mathbb{E}\xi^n/n! \to 0$ *при* $n \to \infty$. *Тогда*

$$\varphi_\xi(t) = \sum_{k=0}^{\infty} \frac{(it)^k}{k!}\mathbb{E}\xi^k.$$

(3) (Б/Д) Если производная $\varphi_\xi^{(2n)}(0)$ *определена для некоторого* n, *то* $\mathbb{E}|\xi|^{2n} < \infty$.

Подчеркнем, что формула (11.6) позволяет выразить моменты $\mathbb{E}\xi^k$ случайной величины ξ через производные характеристической функции в нуле.

Доказательство. (1) Фактически требуется доказать, что производную можно прокидывать внутрь матожидания (= интеграла), а в формуле Тейлора есть упомянутая оценка на $o(t^n)$. Удобно это проделать следующим образом. Допустим, что $\mathbb{E}\xi$ определено. Тогда

$$\frac{\varphi_\xi(t+h) - \varphi_\xi(t)}{h} = \mathbb{E}\Big[e^{it\xi}\frac{e^{ih\xi}-1}{h}\Big].$$

Теперь теорема Лебега о мажорируемой сходимости (теорема B.26) влечет, что

$$\frac{\varphi_\xi(t+h) - \varphi_\xi(t)}{h} \to \mathbb{E}\big(i\xi e^{it\xi}\big) \quad \text{при} \quad h \to 0.$$

Действительно, при каждом $\omega \in \Omega$,

$$e^{it\xi(\omega)}\frac{e^{ih\xi(\omega)}-1}{h} \to e^{it\xi(\omega)}i\xi(\omega).$$

Кроме того,

$$\Big|e^{it\xi}\frac{e^{ih\xi}-1}{h}\Big| = \Big|\frac{e^{ih\xi}-1}{h}\Big| \le |\xi|,$$

а $\mathbb{E}|\xi| < \infty$. Справедливость формулы (11.6) для $k > 1$ проверяется по индукции.

Утверждение (11.7) удобно доказывать следующим образом. Применяя Формулу Тейлора с остатком в форме Лагранжа к действительной и мнимой части [25] экспоненты $e^{iy} = \cos y + i\sin y$, а затем складывая результаты, для каждого $y \in \mathbb{R}$ находим [26]

$$e^{iy} = \sum_{i=0}^{n-1}\frac{(iy)^k}{k!} + \frac{(iy)^n}{n!}\big(\cos(\theta_1 y) + i\sin(\theta_2 y)\big) = \sum_{i=0}^{n}\frac{(iy)^k}{k!} + \frac{(iy)^n}{n!}\big(\cos(\theta_1 y) - 1 + i\sin(\theta_2 y)\big),$$

где $|\theta_{1,2}| \le 1$. Заменяя y на $t\xi$ и беря от обеих частей матожидание, получаем требуемую формулу с

$$\varepsilon(t) = \mathbb{E}\Big(\xi^n\big(\cos(\theta_1 t\xi) - 1 + i\sin(\theta_2 t\xi)\big)\Big).$$

Так как $\big|\big(\cos(\theta_1 t\xi) - 1 + i\sin(\theta_2 t\xi)\big| \le 3$, имеем $|\varepsilon(t)| \le 3\mathbb{E}|\xi|^n$. Кроме того, по теореме Лебега о мажорируемой сходимости, $\varepsilon(t) \to 0$ при $t \to 0$. Действительно, при всех $\omega \in \Omega$,

$$\cos(\theta_1 t\xi(\omega)) \to 1 \quad \text{и} \quad \sin(\theta_2 t\xi(\omega)) \to 0 \quad \text{при} \quad t \to 0.$$

[25] Применить формулу Тейлора сразу к экспоненте e^{iy} (как я ошибочно сделал на лекции) нельзя: остаток в форме Лагранжа пишется только для действительных функций.

[26] Проделайте это вычисление: нужно понять до какого порядка раскладывать косинус, а до какого синус (подсказка: до разных).

(3) Согласно пункту (1),

$$\Big|\varphi_\xi(t) - \sum_{k=0}^{n} \frac{(it)^k}{k!}\mathbb{E}\xi^k\Big| \le 3\frac{|t|^n \mathbb{E}|\xi|^n}{n!} \to 0 \quad \text{при} \quad |t| \le t_0 \quad \text{и} \quad n \to \infty.$$

(3) Доказательство этого утверждения леммы, как и всех остальных, а также некоторых других утверждений, можно найти в [Shi-1, теорема 2.12.1]. □

- Контрольный вопрос: верно ли, что любая функция $\varphi(t)$ является характеристической для некоторой случайной величины ξ, то есть $\varphi(t) = \varphi_\xi(t)$? Сперва обдумайте, а потом посмотрите в [KS] или загуглите теорему Бохнера-Хинчина.

11.2 Теорема обращения

Здесь мы доказываем теорему обращения, позволяющую по характеристической функции φ_ξ *произвольной* случайной величины ξ восстановить ее распределение, и извлекаем некоторые следствия из этого результата. Фактически теорема утверждает обратимость преобразования Фурье и относится к функциональному анализу.

Теорема 11.5 (Теорема обращения). *1. Для любого интервала (a, b) верно*

$$\mathbb{P}\{a < \xi < b\} + \frac{\mathbb{P}\{\xi = a\} + \mathbb{P}\{\xi = b\}}{2} = \lim_{R\to\infty} \frac{1}{2\pi} \int_{-R}^{+R} \frac{e^{-ita} - e^{-itb}}{it}\varphi_\xi(t)dt. \tag{11.8}$$

2. Если характеристическая функция φ_ξ интегрируема [27]*, то случайная величина ξ абсолютно непрерывна и ее плотность дается преобразованием Фурье*

$$p_\xi(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\varphi_\xi(t)\,dt. \tag{11.9}$$

Прежде, чем перейти к доказательству теоремы, нестрого объясним происхождение загадочной формулы (11.8). Для этого рассмотрим ситуацию, когда случайная величина ξ абсолютно непрерывна. Как мы обсудили в (11.3), в этом случае ее характеристическая функция есть ни что иное как обратное преобразование Фурье плотности p_ξ, а значит, при некоторых предположениях регулярности, верна формула (11.9) по теореме об обратном преобразовании Фурье. Тогда

$$\mathbb{P}(a < \xi < b) = \int_a^b p_\xi(x)\,dx = \frac{1}{2\pi}\int_{-\infty}^{\infty}\varphi_\xi(t)\int_a^b e^{-itx}\,dxdt, \tag{11.10}$$

где мы использовали теорему Фубини (теорема B.27). Так как

$$\int_a^b e^{-itx}\,dx = \frac{e^{-ita} - e^{-itb}}{it}, \tag{11.11}$$

мы получаем правую часть формулы (11.8). Остается заметить, что для абсолютно непрерывной с.в. ξ имеем $\mathbb{P}\{\xi = x\} = 0$ для любого $x \in \mathbb{R}$, так что левая часть (11.8) равна $\mathbb{P}(a < \xi < b)$.

[27] То есть, $\int_{-\infty}^{+\infty}|\varphi_\xi(t)|dt < \infty$.

Доказательство теоремы. (1) Доказательство, как и рассуждение выше, основано на применении теоремы Фубини. Обозначим через I_R интеграл в правой части равенства (11.8). По определению характеристической функции

$$I_R = \frac{1}{2\pi}\int_{-R}^{R} \frac{e^{-ita} - e^{-itb}}{it}\left[\int\limits_{-\infty}^{\infty} e^{itx} d\mathbb{P}_\xi\right] dt.$$

Применяя теорему Фубини (теорема B.27), находим

$$I_R = \frac{1}{2\pi}\int\limits_{-\infty}^{\infty}\left[\int\limits_{-R}^{R} \frac{e^{-ita} - e^{-itb}}{it} e^{itx} dt\right] d\mathbb{P}_\xi =: \frac{1}{2\pi}\int\limits_{-\infty}^{\infty} \Psi_R(x) d\mathbb{P}_\xi.$$

Теорему Фубини B.27 можно применить благодаря ее пункту (2):

$$\left|\frac{e^{-ita} - e^{-itb}}{it} e^{itx} dt\right| = \Big|\int\limits_{a}^{b} e^{-ity} dy\Big| \leqslant b - a \quad \Rightarrow$$

$$\int\limits_{-\infty}^{\infty}\left[\int\limits_{-R}^{R}\left|\frac{e^{-ita} - e^{-itb}}{it} e^{itx}\right| dt\right] d\mathbb{P}_\xi \leq \int\limits_{-\infty}^{\infty}\int\limits_{-R}^{R} (b-a)\, dt d\mathbb{P}_\xi = 2R(b-a) < \infty.$$

Далее,

$$\Psi_R(x) = \int_{-R}^{R} \frac{e^{it(x-a)} - e^{it(x-b)}}{it}\, dt = \int_{-R}^{R} \frac{\sin\big(t(x-a)\big) - \sin\big(t(x-b)\big)}{t}\, dt,$$

так как функции $\cos\big(t(x-a)\big)/t$ и $\cos\big(t(x-b)\big)/t$ нечетные. Используя нечетность синусов, а затем заменяя переменную интегрирования, находим

$$\Psi_R(x) = 2\int\limits_{0}^{R} \frac{\sin\big(t(x-a)\big) - \sin\big(t(x-b)\big)}{t}\, dt = 2\left[\int\limits_{0}^{R(x-a)} \frac{\sin u}{u} du - \int\limits_{0}^{R(x-b)} \frac{\sin u}{u} du\right]$$

Напомним, что $\int_0^\infty \frac{\sin x}{x}\, dx = \frac{\pi}{2}$. Значит,

$$\lim_{R\to\infty} \Psi_R(x) = 2\begin{cases} 0 & x < a \text{ или } x > b \\ \frac{\pi}{2} & x = a \text{ или } x = b \\ \pi & a < x < b \end{cases}.$$

Тогда, по теореме Лебега о мажорируемой сходимости находим

$$\lim_{R\to\infty} I_R = \frac{1}{2\pi}\int\limits_{-\infty}^{\infty} \lim_{R\to\infty} \Psi_R(x)\, d\mathbb{P}_\xi(x) = \mathbb{P}_\xi\big((a,b)\big) + \frac{\mathbb{P}_\xi(\{a\}) + \mathbb{P}_\xi(\{b\})}{2}.$$

(2) Пусть a и b таковы, что $\mathbb{P}(\xi = a) = \mathbb{P}(\xi = b) = 0$. Рассуждение похоже на цепочку равенств (11.10)-(11.11), пройденную в обратном направлении. А именно, используя в (11.8) тождество (11.11) и теорему Фубини, находим

$$\mathbb{P}(a < \xi < b) = \lim_{R\to\infty}\int_a^b \frac{1}{2\pi}\int_{-R}^{R} e^{-itx}\varphi_\xi(x)\, dtdx.$$

По условию теоремы $\int_{-\infty}^{\infty} |\varphi_\xi(x)|\,dx < \infty$, а значит применима теорема Лебега о мажорируемой сходимости. Находим

$$\mathbb{P}(a < \xi < b) = \int_a^b f(x)\,dx, \quad \text{где} \quad f(x) := \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\varphi_\xi(x)\,dx. \tag{11.12}$$

Докажем, что соотношение $\mathbb{P}(\xi = x) = 0$ выполнено для всех x. Для этого достаточно взять последовательность точек $a_n < x < b_n$, $a_n, b_n \to x$ при $n \to \infty$, таких, что $\mathbb{P}(\xi = a_n) = \mathbb{P}(\xi = b_n) = 0$, и перейти к пределу $n \to \infty$ в левой и правой частях равенства (11.12). [28] Теперь формула (11.12) доказана для всех $a < b$ и, устремляя $a \to -\infty$, мы находим, что функция распределения с.в. ξ имеет вид

$$F_\xi(b) = \int_{-\infty}^b f(x)\,dx \quad \forall b \in \mathbb{R},$$

то есть f — плотность ξ. □

Следствие 11.6. *Пусть $a < b$ — точки непрерывности функции распределения F_ξ. Тогда*

$$F_\xi(b) - F_\xi(a) = \lim_{R\to\infty} \frac{1}{2\pi}\int\limits_{-R}^{+R} \frac{e^{-ita} - e^{-itb}}{it}\varphi_\xi(t)dt. \tag{11.13}$$

Доказательство. Имеем

$$F_\xi(b) - F_\xi(a) = \mathbb{P}(a < \xi \le b) = \mathbb{P}(a < \xi < b) = \text{правая часть (11.13)},$$

согласно теореме 11.5, так как $\mathbb{P}(\xi = a) = \mathbb{P}(\xi = b) = 0$ по условию теоремы. □

Следствие 11.7. *Характеристическая функция однозначно задает распределение случайной величины. Точнее,*

$$\varphi_\xi = \varphi_\eta \quad \Leftrightarrow \quad \mathbb{P}_\xi = \mathbb{P}_\eta. \tag{11.14}$$

Доказательство. $\Leftarrow$ очевидно. Обратно, если $\varphi_\xi = \varphi_\eta$, то следствие 11.6 вместе с равенством $\lim\limits_{x\to-\infty} F_\xi(x) = \lim\limits_{x\to-\infty} F_\eta(x) = 0$ влекут $F_\xi = F_\eta$, а значит $\mathbb{P}_\xi = \mathbb{P}_\eta$. □

Ввиду последнего следствия и леммы 11.4 возникает естественный вопрос о взаимно однозначном соответствии распределения и его моментов. Однако без предположения о не слишком быстром росте моментов (например, как в пункте (2) леммы 11.4) такое утверждение будет неверно. Контрпример обеспечивает, к примеру, логнормальная с.в. β с плотностью $f_\beta(x) = \frac{1}{x\sqrt{2\pi}}\exp\left(-\ln^2 x/2\right)$, $x > 0$, подробности см. R.Durrett "Probability theory and Examples", §3.3.5.

11.3 Характеристическая функция случайного вектора

Характеристической функцией случайного вектора $\xi = (\xi_1, \dots, \xi_m)$ называется функция векторного аргумента

$$\varphi_\xi(t) = \mathbb{E}e^{it\cdot\xi} = \int\limits_{\mathbb{R}^m} e^{it\cdot x}\,d\mathbb{P}_\xi(x_1, \dots, x_m), \qquad t = (t_1, \dots t_m) \in \mathbb{R}^m,$$

[28] Такая последовательность существует, потому что множество точек y, где $\mathbb{P}(\xi = y) \neq 0$, не может быть более, чем счетным. Действительно, к примеру, это следует из того, что монотонная функция F_ξ не может иметь более, чем счетное число точек разрыва.

где $a \cdot b := a_1 b_1 + \dots + a_m b_m$ — скалярное произведение векторов $a, b \in \mathbb{R}^m$. **Все утверждения, доказанные в предыдущем параграфе, имеют аналоги в многомерном случае. В частности, верно (11.14).**

Следующая теорема дает очень удобный критерий независимости случайных величин в терминах характеристических функций.

Теорема 11.8 (Про независимость). *Для того чтобы компоненты случайного вектора $\xi = (\xi_1, \dots \xi_m) \in \mathbb{R}^m$ были независимы, необходимо и достаточно, чтобы его характеристическая функция была произведением характеристических функций компонент:*

$$\varphi_\xi(t) = \mathbb{E}\left[e^{i(t_1\xi_1 + \ldots + t_m\xi_m)}\right] = \prod_{k=1}^m \mathbb{E}e^{it_k\xi_k} = \prod_{k=1}^m \varphi_{\xi_k}(t_k).$$

Доказательство. Необходимость следует из того, что для независимых случайных величин матожидание произведения равно произведению матожиданий. Для доказательства достаточности сравним распределение случайного вектора ξ с распределением вектора $\eta = (\eta_1, \dots, \eta_m)$ такого, что $\mathbb{P}_{\xi_j} = \mathbb{P}_{\eta_j}$ для любого j, и $\eta_1, \dots, \eta_m$ независимы. Благодаря этой независимости и уже доказанной необходимости имеем

$$\varphi_\eta(t) = \prod_{k=1}^m \varphi_{\eta_k}(t_k) = \prod_{k=1}^m \varphi_{\xi_k}(t_k),$$

так как $\mathbb{P}_{\xi_j} = \mathbb{P}_{\eta_j}$, а значит $\varphi_{\eta_k} = \varphi_{\xi_k}$. По предположению, правая часть последнего произведения равна $\varphi_\xi(t)$, а значит $\varphi_\eta = \varphi_\xi$, так что распределения ξ и η совпадают. □

12 Сходимости случайных величин

12.1 Слабая сходимость мер

Пусть $(\mu_n)_{n\geq 1}$ и μ — вероятностные меры [29] на $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. [30] Обозначим через $C_b(\mathbb{R}^d)$ пространство непрерывных ограниченных функций $f : \mathbb{R}^d \mapsto \mathbb{R}$. Норма

$$\|f\|_\infty := \sup_{x\in\mathbb{R}} |f(x)|,$$

превращает $C_b(\mathbb{R}^d)$ в банахово пространство.

Определение 12.1. Последовательность мер μ_n *слабо* сходится к мере μ при $n \to \infty$, если для любой функции $f \in C_b(\mathbb{R}^d)$ выполнено

$$\int_{\mathbb{R}^d} f(x)\, d\mu_n \to \int_{\mathbb{R}^d} f(x)\, d\mu. \tag{12.1}$$

Обозначение:

$$\mu_n \rightharpoonup \mu \quad \text{при} \quad n \to \infty.$$

Пример 12.2. 1) Пусть x_n, x — точки в $\mathbb{R}^d$. Если $x_n \to x$ при $n \to \infty$, то $f(x_n) \to f(x)$ для любой непрерывной функции f, а значит $\delta_{x_n} \rightharpoonup \delta_x$.

[29] Все утверждения ниже имеют прямые аналоги и для не вероятностных конечных мер.

[30] Все, что происходит в этом параграфе, работает для мер на любом *польском* пространстве (то есть, полном метрическом сепарабельном). Это релевантно для теории случайных процессов, где случайные величины принимают значения в функциональных пространствах, например, в пространстве непрерывных функций. Мы же для простоты будем обсуждать только случай $\mathbb{R}^d$.

2) Пусть меры μ_n и μ абсолютно непрерывны и имеют плотности ρ_n и ρ, причем $\rho_n \to \rho$ в $L_1(\mathbb{R}^d, dx)$ (как всегда, dx обозначает меру Лебега), то есть $\int_{\mathbb{R}^d} |\rho(x) - \rho_n(x)|\, dx \to 0$ при $n \to \infty$. Тогда $\mu_n \rightharpoonup \mu$.

Действительно,

$$\Big| \int_{\mathbb{R}^d} f(x)\, d\mu_n - \int_{\mathbb{R}^d} f(x)\, d\mu \Big| \le \int_{\mathbb{R}^d} |f(x)| \big|\rho_n(x) - \rho(x)\big|\, dx \le C \int_{\mathbb{R}^d} \big|\rho_n(x) - \rho(x)\big|\, dx \to 0,$$

так как функция f ограничена константой.

Следующая теорема, называемая теоремой портманто, проясняет смысл слабой сходимости. Мы приводим ее без доказательства (его часть составляет одну из задач 3го листка).

Теорема 12.3 (портманто)**.** *Слабая сходимость мер $\mu_n \rightharpoonup \mu$ эквивалентна каждому из нижеследующих утверждений:*

1. *$\mu_n(A) \to \mu(A)$ для любого множества $A \in \mathcal{B}(\mathbb{R}^d)$ такого, что $\mu(\partial A) = 0$, где ∂A обозначает границу множества $A \in \mathcal{B}(\mathbb{R}^d)$.*
2. *$\limsup_{n\to\infty} \mu(A) \le \mu(A)$ для любого замкнутого множества $A \in \mathcal{B}(\mathbb{R}^d)$.*
3. *$\liminf_{n\to\infty} \mu(A) \ge \mu(A)$ для любого открытого множества $A \in \mathcal{B}(\mathbb{R}^d)$.*

Обсудим пункт (1) теоремы. Если в (12.1) выбрать $f = \mathbb{I}_A$, мы получим $\int f(x)\, d\mu_n = \mu_n(A)$ и, аналогично, $\int f(x)\, d\mu = \mu(A)$. Однако, индикаторная функция разрывна, поэтому вообще говоря слабая сходимость $\mu_n \rightharpoonup \mu$ не влечет сходимость $\mu_n(A) \to \mu(A)$. Пункт (1) теоремы утверждает, что последняя сходимость таки имеет место, если $\mu(\partial A) = 0$. На самом деле, это условие означает, что мы можем достаточно хорошо (в подходящем смысле) приблизить индикаторную функцию непрерывными. Обратное утверждение, что пункт (1) теоремы влечет слабую сходимость, следует из приближения непрерывной функции индикаторными.

Хорошее упражнение, помогающее понять происходящее, — рассмотреть последовательность мер $\delta_{1/n}$, которая слабо сходится к δ_0, согласно примеру 12.2(1). Проделаем это для пункта (1) теоремы, а для остальных оставим в качестве упражнения.

Пример 12.4. Пусть $A = \{0\}$. Так как $\mu(\partial A) = 1$, пункт (1) теоремы не утверждает, что $\mu_n(A) \to \mu(A)$. И действительно, этой сходимости нет: $\mu_n(A) = 0$, а $\mu(A) = 1$. Если же $A = (-\varepsilon, \varepsilon)$, $\varepsilon > 0$, то начиная с некоторого n, все $\mu_n(A) = 1$, так что сходимость $\mu_n(A) \to \mu(A)$ действительно имеет место. Из этого примера видно, что условие $\mu(\partial A) = 0$ — по существу. Если его убрать, взяв в качестве определения слабой сходимости просто сходимость $\mu_n(A) \to \mu(A)$ для любого $A \in \mathcal{B}(\mathbb{R}^d)$, то получится, что $\delta_{1/n}$ не сходится к δ_0, что очень неудобно с практической точки зрения.

В следующих двух определениях мы предполагаем, что у нас имеется набор вероятностных мер μ_α, параметризованных некоторым множеством индексов $\mathcal{U} \ni \alpha$ (произвольной мощности).

Определение 12.5. Набор мер $(\mu_\alpha)_{\alpha \in \mathcal{U}}$ называется *плотным*, [31] если для любого $\varepsilon > 0$ существует компакт K_ε, такой что $\mu_\alpha(K_\varepsilon) > 1 - \varepsilon$ для любого $\alpha \in \mathcal{U}$.

[31] На английском для этого понятия имеется специальное слово tight (а не dense). Но хорошего перевода на русский нет, и для двух понятий tight и dense используется один и тот же термин "плотный" (хотя, как мне было замечено на лекции, уместно было ты использовать слово "тесный").

Заметим, что если набор состоит из конечного числа мер, то он всегда плотен: в качестве K_ε можно взять шар достаточно большого радиуса. Если бы мы жили не на $\mathbb{R}^d$, а на произвольном польском пространстве, этот факт потребовал бы доказательства (но он верен), так как в бесконечномерии шар не компактен.

Определение 12.6. Набор мер $(\mu_\alpha)_{\alpha\in\mathcal{U}}$ называется *относительно компактным*, если из любой последовательности мер из этого набора можно выбрать слабо сходящуюся подпоследовательность.

Следующая теорема — мощнейший инструмент для доказательства слабой сходимости мер. Мы его будем использовать для доказательства ключевой теоремы 12.12.

Теорема 12.7 (Прохоров)**.** *Набор мер $(\mu_\alpha)_{\alpha\in\mathcal{U}}$ относительно компактен тогда и только тогда, когда он плотен.*

Мы не приводим доказательства этой теоремы, хотя оно не столь уж длинно. См. [KS, теорема 8.9].

Дополнительная информация о слабой сходимости мер

За детальным обсуждением с доказательствами материла этого параграфа см, например, [Dud, §11].

На самом деле, определение 12.1 эквивалентно аналогичному определению, где вместо всех непрерывных ограниченных функций рассматриваются только липшицевы ограниченные функции. Более того, слабая сходимость метризуема, то есть эквивалентна сходимости по так называемой дуально-липшицевой метрике:

$$\rho(\mu_n,\mu) := \sup_f \Big| \int_{\mathbb{R}^d} f(x)\,d\mu_n - \int_{\mathbb{R}^d} f(x)\,d\mu \Big|,$$

где супремум берется по всем липшицевым ограниченным функциям, таким, что

$$\|f\|_\infty + C_L(f) \le 1,$$

где $\|f\|_\infty = \sup_{x\in\mathbb{R}^d} |f(x)|$, а $C_L(f)$ — константа Липшица функции f. Это очень инструментальный факт. Во-первых, разность липшицевых функций удобно оценивать. Во-вторых, в терминах метрики можно говорить о скорости сходимости.

Есть и другие способы задать метрику на множестве вероятностных мер.

12.2 Сходимость случайных величин по распределению

Пусть имеется последовательность случайных величин или векторов $(\xi_n)_{n\ge1}$ и случайная величина ξ. Они *не обязаны* жить на одном и том же вероятностном пространстве, а могут быть определены на разны (впрочем, это не играет большой роли).

Определение 12.8. Последовательность ξ_n сходится к ξ *по распределению*, если их распределения слабо сходятся: $\mathbb{P}_{\xi_n} \rightharpoonup \mathbb{P}_\xi$ при $n\to\infty$. Обозначение:

$$\xi_n \xrightarrow{d} \xi \quad \text{при} \quad n\to\infty.$$

По определению слабой сходимости это означает, что

$$\int_{\mathbb{R}^n} f(x)\,\mathbb{P}_{\xi_n}(x) \to \int_{\mathbb{R}^n} f(x)\,\mathbb{P}_\xi(x),$$

для любой ограниченной непрерывной функции $f : \mathbb{R}^d \mapsto \mathbb{R}$. Но в соотношении выше в левой и правой частях написано ни что иное как матожидания $\mathbb{E}f(\xi_n)$ и $\mathbb{E}f(\xi)$. Таким образом, определение сходимости случайных величин по распределению переписывается следующим образом:

$$\mathbb{E}f(\xi_n) \to \mathbb{E}f(\xi) \quad \text{при} \quad n \to \infty, \tag{12.2}$$

для любой функции $f \in C_b(\mathbb{R}^d)$.

Теорема портманто 12.3 очевидным образом переговаривается в терминах сходимости по распределению случайных величин:

Теорема 12.9 (портманто)**.** *Сходимость по распределению $\xi_n \xrightarrow{d} \xi$ эквивалентна каждому из следующих свойств:*

1. $\mathbb{P}(\xi_n \in A) \to \mathbb{P}(\xi \in A)$ *для любого множества* $A \in \mathcal{B}(\mathbb{R}^d)$ *такого, что* $\mathbb{P}(\xi \in \partial A) = 0$.
2. $\limsup\limits_{n\to\infty} \mathbb{P}(\xi_n \in A) \le \mathbb{P}(\xi \in A)$ *для любого замкнутого множества* $A \in \mathcal{B}(\mathbb{R}^d)$.
3. $\liminf\limits_{n\to\infty} \mathbb{P}(\xi_n \in A) \ge \mathbb{P}(\xi \in A)$ *для любого открытого множества* $A \in \mathcal{B}(\mathbb{R}^d)$.

Пример 12.10. 1) Если $\xi_n = x_n$ п.н. и $\xi = x$ п.н. (x_n, x — константы), причем $x_n \to x$, то $\xi_n \xrightarrow{d} \xi$. Действительно, это следует из примера 12.2(1), так как $\mathbb{P}_{\xi_n} = \delta_{\xi_n}$ и $\mathbb{P}_\xi = \delta_x$.

2) Если случайные величины ξ_n и ξ абсолютно непрерывны и имеют плотности ρ_n и ρ, причем $\rho_n \to \rho$ в $L_1(\mathbb{R}^d, dx)$, то $\xi_n \xrightarrow{d} \xi$. Действительно, по определению, плотности ρ_n и ρ есть ни что иное как плотности мер $\mathbb{P}_{\xi_n}$ и $\mathbb{P}_\xi$, так что утверждение следует из примера 12.2(2).

3) Пусть ξ_n и ξ — целочисленные случайные величины, причем $\mathbb{P}(\xi_n = k) \to \mathbb{P}(\xi = k)$ при $n \to \infty$ для всех $k \in \mathbb{Z}$. Тогда $\xi_n \xrightarrow{d} \xi$. Проверку этого утверждения оставим в качестве упражнения. Ср. с задачей номер 2 из списка задач 9!

4) Пусть $\xi_n \sim \mathcal{N}(m_n, \sigma_n^2)$ и $m_n \to m$, $\sigma_n \to \sigma$. Тогда $\xi_n \xrightarrow{d} \xi$, где $\xi \sim \mathcal{N}(m, \sigma^2)$.

5) В центральной предельной теореме речь идет именно о сходимости по распределению, см. (8.4). Дело в том, что граница интервалов (a, b) состоит из двух точек a и b, и $\mathbb{P}(\eta = a) = \mathbb{P}(\eta = b) = 0$ для $\xi \in \mathcal{N}(0, 1)$, так как нормальное распределение абсолютно непрерывно. Дальше нужно применить первый пункт теоремы портманто.

Имеется следующая связь между сходимостью по распределению и сходимостью функций распределения, которая нередко берется за **определение** этой сходимости.

Теорема 12.11. *Сходимость случайных величин по распределению $\xi_n \xrightarrow{d} \xi$ эквивалентна поточечной сходимости их функций распределения $F_{\xi_n}(x) \to F_\xi(x)$ для всех x, где F_ξ непрерывно.*

Доказательство. Здесь мы докажем теорему только в одну сторону, а доказательство обратной импликации будет дано в параграфе 14.0.1, после обсуждения связи между сходимостями по распределению и почти наверное.

Допустим, что $\xi_n \xrightarrow{d} \xi$. Тогда для любой точки x, в которой функция F_ξ непрерывна

$$F_n(\xi) = \mathbb{P}\big\{\xi_n \in (-\infty, x]\big\} \to \mathbb{P}\big\{\xi \in (-\infty, x]\big\} = F_\xi(x),$$

где сходимость следует из пункта (1) теоремы портманто 12.9. Действительно, $\partial(\infty, x] = \{x\}$, а $\mathbb{P}\{\xi = x\} = F_\xi(x) - F_\xi(x-) = 0$. ☐

12.3 Связь сходимости по распределению с характеристическими функциями

Наконец то, ради чего весь сыр-бор.

Теорема 12.12. *Сходимость случайных векторов по распределению эквивалентна поточечной сходимости их характеристических функций:*

$$\xi_n \xrightarrow{d} \xi \quad \Leftrightarrow \quad \varphi_{\xi_n}(t) \to \varphi_\xi(t) \quad \text{для любого } t \in \mathbb{R}^d, \quad \text{при} \quad n \to \infty.$$

.

На самом деле, верно больше: если $\varphi_{\xi_n}(t) \to \varphi(t)$ поточечно для некоторой функции φ, непрерывной в нуле, то φ — характеристическая функция некоторой случайной величины ξ, и $\xi_n \xrightarrow{d} \xi$. См. [Shi-1, §3.3].

Доказательство. Мы проведем доказательство только для $d = 1$.

$\Rightarrow$ Требуемое утверждение сразу следует из определения сходимости по распределению 12.2. Действительно,

$$\varphi_{\xi_n}(t) = \mathbb{E}\cos(\xi_n t) + i\mathbb{E}\sin(\xi_n t) \to \mathbb{E}\cos(\xi t) + i\mathbb{E}\sin(\xi t) = \varphi_\xi(t).$$

$\Leftarrow$ Доказательство состоит из двух шагов. На первом мы докажем, что семейство распределений $(\mathbb{P}_{\xi_n})_{n\geq 1}$ плотно, а значит, благодаря теореме Прохорова, относительно компактно. На втором шаге мы покажем, что любая сходящаяся последовательность распределений из этого семейства обязана сходиться к распределению $\mathbb{P}_\xi$, в силу уже доказанной импликации $\Rightarrow$ теоремы.

Шаг 1. Доказательство плотности семейства мер $(\mathbb{P}_{\xi_n})_{n\geq 1}$ основано на следующей леммы, которую мы докажем после окончания доказательства теоремы.

Лемма 12.13. *Для любой случайной величины η и ее характеристической функции φ_η*

$$\mathbb{P}\Big\{|\eta| \leq \frac{2}{\tau}\Big\} \geq \Big|\frac{1}{\tau}\int_{-\tau}^{\tau} \varphi_\eta(t)\,dt\Big| - 1, \quad \text{для любого} \quad \tau > 0. \tag{12.3}$$

Разумеется, интересный случай в этой лемме, когда τ мало. В этой ситуации $\varphi_\eta(t) \approx \varphi_\eta(0) = 1$, так что правая часть неравенства (12.3) близка к единице. Теперь сделаем эти соображения количественными.

Зафиксируем произвольный $\varepsilon > 0$. Так как характеристическая функция непрерывна, существует $\tau > 0$, такой что $|\varphi_\xi(t) - 1| \leq \varepsilon/4$ при $|t| \leq \tau$. Тогда

$$\Big|\int_{-\tau}^{\tau} \varphi_\xi(t)\,dt\Big| = \Big|\int_{-\tau}^{\tau} (\varphi_\xi(t) - 1)\,dt + 2\tau\Big| \geq 2\tau - \int_{-\tau}^{\tau} \big|\varphi_\xi(t) - 1\big|\,dt \geq 2\tau(1 - \frac{\varepsilon}{4}).$$

Значит,

$$\Big|\frac{1}{\tau}\int_{-\tau}^{\tau} \varphi_\xi(t)\,dt\Big| \geq 2 - \frac{\varepsilon}{2}.$$

По предположению теоремы, $\varphi_{\xi_n}(t) \to \varphi_\xi(t)$ для любого t. Так как $|\varphi_{\xi_n}| \leq 1$, по теореме Лебега о мажорируемой сходимости,

$$\int_{-\tau}^{\tau} \varphi_{\xi_n}(t)\,dt \to \int_{-\tau}^{\tau} \varphi_\xi(t)\,dt \quad \text{при} \quad n \to \infty.$$

Значит, существует $N \geq 1$ такой, что при $n \geq N$,

$$\Big|\frac{1}{\tau}\int\limits_{-\tau}^{\tau} \varphi_{\xi_n}(t)\,dt\Big| \geq 2 - \varepsilon.$$

Тогда, в силу леммы 12.13,

$$\mathbb{P}\Big\{|\xi_n| \leq \frac{2}{\tau}\Big\} \geq 1 - \varepsilon.$$

Выберем теперь $t_N > 0$ такое, что $\mathbb{P}\{|\xi_n| \leq t_N\} \geq 1 - \varepsilon$ для всех $n < N$ (это можно сделать, так как таких n конечное число) и положим $R := \max(t_N, 2/\tau)$. Тогда

$$\mathbb{P}_{\xi_n}\big([-R, R]\big) = \mathbb{P}\big\{|\xi_n| \leq R\big\} \geq 1 - \varepsilon \quad \text{для любого} \quad n \geq 1,$$

то есть семейство мер $(\mathbb{P}_{\xi_n})_{n\geq 1}$ плотно (см. определение 12.5). Значит, по теореме Прохорова 12.7 оно относительно компактно.

Шаг 2. Проведем доказательство от противного. Пусть ξ_n не сходится по распределению к ξ. Тогда существует подпоследовательность ξ_{n_j} и непрерывная ограниченная функция $f : \mathbb{R} \mapsto \mathbb{R}$ такие, что

$$\Big|\mathbb{E}f(\xi_{n_j}) - \mathbb{E}f(\xi)\Big| \geq \delta \quad \text{для всех } j \geq 1 \text{ и некоторого } \delta > 0. \tag{12.4}$$

Но, по определению 12.6 относительной компактности, последовательность мер $\mathbb{P}_{\xi_{n_j}}$ содержит подпоследовательность $\mathbb{P}_{\xi_{n_{j_k}}}$, слабо сходящуюся при $k \to \infty$. Обозначим предельную меру μ и выберем случайную величину α с распределением μ (то есть $\mathbb{P}_\alpha = \mu$). Тогда $\xi_{n_{j_k}} \xrightarrow{d} \alpha$ при $k \to \infty$. По уже доказанной импликации $\Rightarrow$ теоремы, характеристическая функция $\varphi_{\xi_{n_{j_k}}}$ поточечно сходится к φ_α. Но по предположению теоремы, $\varphi_{\xi_n} \to \varphi_\xi$, так что $\varphi_\alpha = \varphi_\xi$, а значит распределения $\mathbb{P}_\alpha$ и $\mathbb{P}_\xi$ случайных величин α и ξ совпадают (по следствию 11.7). Таким образом, $\xi_{n_{j_k}} \xrightarrow{d} \xi$, что противоречит (12.4). □

Доказательство леммы 12.13. Записывая характеристическую функцию в виде (11.4) и применяя теорему Фубини, находим

$$I_\eta := \frac{1}{2\tau}\int_{-\tau}^{\tau} \varphi_\eta(t)\,dt = \frac{1}{2\tau}\int_{-\infty}^{\infty}\int_{-\tau}^{\tau} e^{itx}\,dt d\mathbb{P}_\eta(x) = \int_{-\infty}^{\infty} \frac{\sin(x\tau)}{x\tau}\,d\mathbb{P}_\eta(x).$$

Разбивая последний интеграл на два, получаем

$$|I_\eta| \leq \int_{|x|\leq\frac{2}{\tau}} \Big|\frac{\sin(x\tau)}{x\tau}\Big|\,d\mathbb{P}_\eta(x) + \int_{|x|>\frac{2}{\tau}} \Big|\frac{\sin(x\tau)}{x\tau}\Big|\,d\mathbb{P}_\eta(x) =: J_1 + J_2.$$

Так как $|\sin y/y| \leq 1$ при $|y| \leq 2$,

$$J_1 \leq \mathbb{P}_\eta\Big(\Big[-\frac{2}{\tau}, \frac{2}{\tau}\Big]\Big).$$

Ограничивая синус единицей, находим

$$J_2 \leq \frac{1}{2}\mathbb{P}_\eta\Big(\mathbb{R} \setminus \Big[-\frac{2}{\tau}, \frac{2}{\tau}\Big]\Big) = \frac{1}{2} - \frac{1}{2}\mathbb{P}_\eta\Big(\Big[-\frac{2}{\tau}, \frac{2}{\tau}\Big]\Big).$$

Таким образом,

$$|I_\eta| \leq \frac{1}{2} + \frac{1}{2}\mathbb{P}_\eta\Big(\Big[-\frac{2}{\tau}, \frac{2}{\tau}\Big]\Big), \quad \text{то есть} \quad \mathbb{P}_\eta\Big(\Big[-\frac{2}{\tau}, \frac{2}{\tau}\Big]\Big) \geq 2|I_\eta| - 1.$$

□

13 Закон больших чисел в форме Хинчина и центральная предельная теорема: доказательства

В этом разделе мы соберем урожай с предыдущих параграфов: с помощью метода характеристических функций мы усилим закон больших чисел и докажем центральную предельную теорему. Как всегда, для последовательности случайных величин $(\xi_n)_{n\geq 1}$ мы будем обозначать

$$S_n := \sum_{j=1}^{n} \xi_j.$$

Теорема 13.1 (ЗБЧ в форме Хинчина)**.** *Пусть $(\xi_n)_{n\geq 1}$ — последовательность iid интегрируемых (т.е. $\mathbb{E}|\xi_1| < \infty$) случайных величин. Тогда*

$$\frac{S_n}{n} \xrightarrow{\mathbb{P}} \mathbb{E}\xi_1 \quad \textit{при} \quad n \to \infty.$$

В сравнении с ЗБЧ в форме Чебышёва, теперь мы требуем одинаковой распределенности случайных величин. Зато мы требуем конечности только первого момента $\mathbb{E}|\xi_j|$, а не второго. Это очень хорошо, потому что оптимально. Действительно, задаче 8.7 семинаров рассматривается случай, когда $\xi_j \sim Cauchy(t)$ и следовательно, $\mathbb{E}|\xi_1| = \infty$. Доказывается, что $\frac{S_n}{n} \sim Cauchy(t)$ для всех $n \geq 1$, так что никакого ЗБЧ быть не может.

Доказательство. Вычислим характеристическую функцию случайной величины S_n/n:

$$\varphi_{\frac{S_n}{n}}(t) = \mathbb{E}e^{it\frac{\xi_1+\cdots+\xi_n}{n}} = \mathbb{E}\prod_{j=1}^{n} e^{it\frac{\xi_j}{n}} = \prod_{j=1}^{n} \mathbb{E}e^{it\frac{\xi_j}{n}} = \prod_{j=1}^{n} \varphi_{\xi_j}\Big(\frac{t}{n}\Big) = \Big[\varphi_{\xi_1}\Big(\frac{t}{n}\Big)\Big]^n, \tag{13.1}$$

где, чтобы обменять местами матожидание и произведение, мы использовали независимость ξ_j, а в последнем равенстве мы использовали одинаковую распределенность случайных величин ξ_j, так что $\varphi_{\xi_j} = \varphi_{\xi_1}$ для всех j.

Так как случайная величина ξ_1 интегрируема, она удовлетворяет условиям леммы 11.4(1) с $n = 1$. Значит, $\varphi_{\xi_1}(s) = 1 + ism + o(s)$, где мы обозначили $m := \mathbb{E}\xi_1$. Подставляя эту формулу в правую часть (13.1), при каждом фиксированном $t \in \mathbb{R}$ находим

$$\varphi_{\frac{S_n}{n}}(t) = \Big[1 + \frac{itm}{n} + o\Big(\frac{1}{n}\Big)\Big]^n \to e^{itm} \quad \text{при} \quad n \to \infty,$$

так как это в точности второй замечательный предел. Так как характеристическая функция константной случайной величины $\xi(\omega) \equiv m$ имеет вид $\varphi_m(t) = \mathbb{E}e^{itm} = e^{itm}$, мы видим, что $\varphi_{\frac{S_n}{n}}(t) \to \varphi_m(t)$ для всех $t \in \mathbb{R}$. Но теорема 12.12 утверждает, что поточечная сходимость характеристических функций эквивалентна сходимости по распределению, то есть

$$\frac{S_n}{n} \xrightarrow{d} m \quad \text{при} \quad n \to \infty.$$

Остается применить лемму 14.3. ∎

Перейдем теперь к доказательству центральной предельной теоремы. Оно проводится по той же схеме, что и доказательство ЗБЧ в форме Хинчина.

Доказательство теоремы 8.13

Запишем случайную величину η_n в виде

$$\eta_n = \sum_{j=1}^n \frac{\xi_j - m}{\sigma\sqrt{n}}.$$

Вычислим ее характеристическую функцию. Используя независимость с.в. ξ_j, находим

$$\varphi_{\eta_n}(t) = \mathbb{E}\prod_{j=1}^n e^{it\frac{\xi_j - m}{\sigma\sqrt{n}}} = \prod_{j=1}^n \mathbb{E}e^{it\frac{\xi_j - m}{\sigma\sqrt{n}}} = \prod_{j=1}^n \varphi_{\xi_j - m}\Big(\frac{t}{\sigma\sqrt{n}}\Big) = \Big[\varphi_{\xi_1 - m}\Big(\frac{t}{\sigma\sqrt{n}}\Big)\Big]^n,$$

где в последнем равенстве использована одинаковая распределенность с.в. ξ_j. Заметим, что

$$\mathbb{E}(\xi_1 - m) = 0 \quad \text{и} \quad \mathbb{E}(\xi_1 - m) = \sigma^2.$$

Следовательно, лемма 11.4(1) с $n = 2$ влечет, что

$$\varphi_{\xi_1 - m}(s) = 1 - \frac{\sigma^2 s}{2} + o(s).$$

Тогда, подставляя $s = t/\sigma\sqrt{n}$, при каждом фиксированном $t \in \mathbb{R}$ находим

$$\varphi_{\eta_n}(t) = \Big[1 - \frac{t^2}{n} + o\Big(\frac{1}{n}\Big)\Big]^n \to e^{-\frac{t^2}{2}} \quad \text{при} \quad n \to \infty.$$

Так как функция $e^{-t^2/2}$ является характеристической для с.в. $\eta \sim \mathcal{N}(0,1)$, согласно теореме 12.12 находим $\eta_n \xrightarrow{d} \eta$. $\square$

13.0.1 Объяснение ЦПТ с точки зрения теории динамических систем

Для интересующихся очень рекомендуем посмотреть [KS, §10.3]: там дается другой взгляд на центральную предельную теорему, поясняющий почему всегда возникает именно гауссовское распределение. Оказывается, гауссовское распределение есть устойчивая неподвижная точка некоторого оператора.

14 Виды сходимостей случайных величин и их связи

Случайная величина (или вектор) — это функция, а функции умеют сходиться в самых разных смыслах. Перечислим наиболее используемые типы сходимостей последовательности случайных векторов $\xi_n \in \mathbb{R}^d$ к случайному вектору $\xi \in \mathbb{R}^d$. Далее предполагается, что все с.в. живут на одном и том же вероятностном пространстве $(\Omega, \mathcal{F}, \mathbb{P})$, то есть $\xi_n, \xi : \Omega \mapsto \mathbb{R}^d$.

1. **Сходимость почти наверное**, $\xi_n \to \xi$ п.н. (или почти всюду, п.в.):

   $$\mathbb{P}\{\omega \in \Omega : \xi_n(\omega) \to \xi(\omega) \text{ при } n \to \infty\} = 1.$$

   При каждом ω, $(\xi_n(\omega))_{n\geq 1}$ — числовая последовательность, сходимость которой понимается в обычном смысле.

2. **Сходимость в** L_p, $p>0$, $\xi_n \xrightarrow{L_p} \xi$: это известная из анализа сходимость в пространстве лебега $L_p(\Omega,\mathbb{P})$,

$$\|\xi_n-\xi\|_{L_p}^p = \int_\Omega |\xi_n(\omega)-\xi(\omega)|^p \, d\mathbb{P}(\omega) = \mathbb{E}|\xi_n-\xi|^p \to 0 \quad \text{при} \quad n\to\infty.$$

   Здесь и далее $|\cdot|$ обозначает стандартную евклидову норму в $\mathbb{R}^d$. Отметим, что при $p\geq 1$, $\|\cdot\|_{L_p}$ — норма, а пространство $L_p(\Omega,\mathbb{P})$ — банахово (при $p=2$ оно еще и гильбертово).

3. **Сходимость по вероятности**, $\xi_n \xrightarrow{\mathbb{P}} \xi$: это уже обсуждалось,

$$\mathbb{P}\big\{\omega\in\Omega:\, |\xi_n(\omega)-\xi(\omega)|\geq\varepsilon\big\}\to 0 \quad \forall\varepsilon>0.$$

4. **Сходимость по распределению**, $\xi_n \xrightarrow{d} \xi$, которая тоже уже обсуждалась:

$$\mathbb{E}f(\xi_n)\to\mathbb{E}f(\xi) \quad \forall f\in C_b(\mathbb{R}^d).$$

   Заметим, что сходящиеся по распределению случайные величины вообще говоря могут жить на разных вероятностных пространствах.

Теорема 14.1. *1. Имеются следующие импликации между сходимостями:*

$$\textit{п.н.}\Rightarrow\mathbb{P}, \quad L_p\Rightarrow\mathbb{P}, \quad \mathbb{P}\Rightarrow d.$$

Таким образом, сходимости п.н. и в L_p — самые сильные, и они друг из друга не следуют.

2. Обратные импликации не верны, однако:

(a) Из последовательности, сходящейся по вероятности, можно выбрать подпоследовательность, сходящуюся почти наверное.

(b) Если $\xi_n \xrightarrow{d} \xi$, то существует новое вероятностное пространство $(\widetilde\Omega,\widetilde{\mathcal{F}},\widetilde{\mathbb{P}})$ и случайные величины $(\widetilde\xi_n)_{n\geq1}$ и $\widetilde\xi$ на нем такие, что их распределения совпадают с распределениями $(\xi_n)_{n\geq1}$ и ξ ($\mathbb{P}_{\widetilde\xi_n}=\mathbb{P}_{\xi_n}$ для всех $n\geq1$ и $\mathbb{P}_{\widetilde\xi}=\mathbb{P}_\xi$), и при этом

$$\widetilde\xi_n\to\widetilde\xi \quad \textit{п.н.}.$$

Доказательство. (1) Импликация *п.н.* $\Rightarrow\mathbb{P}$ известна из курса теории меры и составляет задачу A листка 3. Импликация $L_p\Rightarrow\mathbb{P}$ немедленно следует из неравенства Маркова:

$$\mathbb{P}\big\{|\xi_n-\xi|\geq\varepsilon\big\}=\mathbb{P}\big\{|\xi_n-\xi|^p\geq\varepsilon^p\big\}\leq\frac{\mathbb{E}|\xi_n-\xi|^p}{\varepsilon^p}\to 0 \quad \text{при} \quad n\to\infty.$$

Докажем импликацию $\mathbb{P}\Rightarrow d$. Для δ, $R>0$ рассмотрим события

$$A_\delta:=\{|\xi_n-\xi|\leq\delta\} \quad \text{и} \quad B_R:=\{|\xi|\leq R\}.$$

Так как $\Omega=(A_\delta\cap B_R)\cup(A_\delta\cap B_R^c)\cup A_\delta^c$, имеем $\mathbb{I}_{A_\delta\cap B_R}+\mathbb{I}_{A_\delta\cap B_R^c}+\mathbb{I}_{A_\delta^c}=1$. Следовательно, обозначая $\Delta_n:=|f(\xi_n)-f(\xi)|$, находим

$$|\mathbb{E}f(\xi_n)-\mathbb{E}f(\xi)|\leq\mathbb{E}\Delta_n=\mathbb{E}\big(\Delta_n\mathbb{I}_{A_\delta\cap B_R}\big)+\mathbb{E}\big(\Delta_n\mathbb{I}_{A_\delta\cap B_R^c}\big)+\mathbb{E}\big(\Delta_n\mathbb{I}_{A_\delta^c}\big)=:I_1+I_2+I_3.$$

Зафиксируем произвольный $\varepsilon > 0$. Выберем R настолько большим, что $\mathbb{P}(B_R^c) < \varepsilon$. Так как $\Delta_n \le 2\|f\|_\infty$, где $\|f\|_\infty := \sup\limits_{x\in\mathbb{R}^d} |f(x)|$, находим

$$|I_2| < 2\|f\|_\infty \varepsilon.$$

Так как функция f непрерывна, а значит равномерно непрерывна на компакте, существует δ такой, что при $|x - y| \le \delta$ и $|y| \le R$ выполнено $|f(x) - f(y)| < \varepsilon$. Следовательно,

$$I_1 < \varepsilon.$$

Так как $\xi_n \xrightarrow{\mathbb{P}} \xi$, можно выбрать настолько большой N, что при $n \ge N$ выполнено $\mathbb{P}(A_\delta^c) < \varepsilon$, а значит

$$I_3 < 2\|f\|_\infty \varepsilon.$$

Итого, для произвольного $\varepsilon > 0$ мы нашли N такое, что при любом $n \ge N$,

$$|\mathbb{E}f(\xi_n) - \mathbb{E}f(\xi)| < 2\|f\|_\infty\varepsilon + \varepsilon + 2\|f\|_\infty\varepsilon = \varepsilon(4\|f\|_\infty + 1).$$

Доказательство окончено.

Замечание. Как обсуждалось в последнем разделе §12.1, в слабой сходимости достаточно рассматривать не все непрерывные функции, а только липшицевы. Использование этого факта упростило бы доказательство: не нужно было бы вводить множество B_R.

Тот факт, что сходимости п.н. и в L_p не влекут друг друга, а также, что импликации, обратные к доказанным, не верны, доказывается несложными контрпримерами, см. задачу 9.5 семинаров. Все контрпримеры можно построить на множестве элементарных исходов $(\Omega, \mathcal{F}, \mathbb{P}) = ([0,1], \mathcal{B}([0,1]), Leb)$.

Пункт (2а) теоремы доказывался в курсе меры. Остается доказать пункт (2б). Нам понадобится следующее утверждение (задача А листка 2.)

Лемма 14.2. *Пусть F_ξ —функция распределения произвольной случайной величины ξ, а $\eta \sim Uniform([0,1])$. Положим*

$$\hat{F}_\xi^{-1}(x) = \inf\{t \in \mathbb{R} : F_\xi(t) \ge x\}. \tag{14.1}$$

Тогда случайная величина $\hat{F}_\xi^{-1}(\eta)$ имеет то же распределение, что и ξ.

Заметим, что если функция F_ξ непрерывна и строго монотонна, а значит обратима, имеем

$$\hat{F}_\xi^{-1} = F_\xi^{-1}. \tag{14.2}$$

Доказательство. Достаточно убедиться, что $F_{\hat{F}_\xi^{-1}(\eta)} = F_\xi$. Для иллюстрации проведем сперва доказательство в случае, когда F_ξ непрерывна и строго монотонна, так что выполнено (14.2):

$$F_{\hat{F}_\xi^{-1}(\eta)}(t) = \mathbb{P}\{F_\xi^{-1}(\eta) \le t\} = \mathbb{P}\{\eta \le F_\xi(t)\} = F_\xi(t).$$

В общем случае рассуждение совершенно аналогично: непосредственно проверяется, что для каждого $x \in \mathbb{R}$

$$\{y \in [0,1] : \hat{F}_\xi^{-1}(y) \le x\} = \{y \in [0,1] : y \le F_\xi(x)\},$$

а значит

$$F_{\hat{F}_\xi^{-1}(\eta)}(t) = \mathbb{P}\{\hat{F}_\xi^{-1}(\eta) \le t\} = \mathbb{P}\{\eta \le F_\xi(x)\} = F_\xi(x).$$

$\square$

(2б) Мы проведем доказательство только в размерности $d = 1$. Выберем $\widetilde{\Omega} = [0,1]$, $\mathcal{F} = \mathcal{B}([0,1])$ и $\mathbb{P} = Leb$. Положим

$$\widetilde{\xi}_n(\omega) = \hat{F}^{-1}_{\xi_n}(\omega) \quad \text{и} \quad \widetilde{\xi}(\omega) = \hat{F}^{-1}_{\xi}(\omega) \qquad \omega \in \widetilde{\Omega}$$

где функции $\hat{F}^{-1}$ определены в (14.1), так что, согласно лемме 14.2, $\mathbb{P}_{\widetilde{\xi}_n} = \mathbb{P}_{\xi_n}$ и $\mathbb{P}_{\widetilde{\xi}} = \mathbb{P}_{\xi}$ (так как случайная величина $\eta(\omega) = \omega$ имеет требуемое равномерное распределение).

Для иллюстрации происходящего сперва проведем доказательство в частном случае, когда функции распределения F_{ξ_n} и F_{ξ} непрерывны и строго монотонны, так что имеет место (14.2) (для ξ и ξ_n).

Согласно теореме 12.11, сходимость $\xi_n \xrightarrow{d} \xi$ влечет сходимость $F_{\xi_n}(t) \to F_{\xi}(t)$ для всех t, а значит то же верно и для функций, обратных к функциям распределения (**докажите это!**)

$$\widetilde{\xi}_n(\omega) = F^{-1}_{\xi_n}(\omega) \to F^{-1}_{\xi}(\omega) = \widetilde{\xi}(\omega) \qquad \forall \omega \in \widetilde{\Omega} = [0,1].$$

Перейдем теперь к доказательству общего случая. Достаточно проверить, что

$$\limsup_{n\to\infty} \hat{F}^{-1}_{\xi_n}(\omega) \le \hat{F}^{-1}_{\xi}(\omega) \le \liminf_{n\to\infty} \hat{F}^{-1}_{\xi_n}(\omega) \quad \forall \omega \in [0,1]. \tag{14.3}$$

Будем рассуждать от противного: пусть $\liminf\limits_{n\to\infty} \hat{F}^{-1}_{\xi_n}(\omega) < \hat{F}^{-1}_{\xi}(\omega)$. Тогда существует подпоследовательность ξ_{n_k} такая, что для любого t, удовлетворяющего $\liminf\limits_{n\to\infty} \hat{F}^{-1}_{\xi_n}(\omega) < t < \hat{F}^{-1}_{\xi}(\omega)$, и достаточно больших k и имеем $\hat{F}^{-1}_{\xi_{n_k}}(\omega) < t < \hat{F}^{-1}_{\xi}(\omega)$. Следовательно, $F_{\xi}(t) < \omega \le F_{\xi_{n_k}}(t)$. Так как монотонная функция имеет не более, чем счетное число разрывов, можно выбрать t точкой непрерывности функции F_{ξ} . Тогда, по теореме 12.11 имеем $F_{\xi_n}(t) \to F_{\xi}(t)$, что приводит к противоречию: $F_{\xi}(t) < \omega \le F_{\xi}(t)$. Второе неравенство в (14.3) доказано. Первое доказывается так же. Доказательство теоремы окончено. □

Как обсудили видели, сходимость по распределению не влечет сходимости по вероятности. Но если предельная случайная величина — константная, то из сходимости по распределению таки следует сходимость по вероятности:

Лемма 14.3. *Если $\xi_n \xrightarrow{d} a$, где $a \in \mathbb{R}^d$ — константа, то $\xi_n \xrightarrow{\mathbb{P}} a$.*

Доказательство. Выберем произвольный $\varepsilon > 0$ и обозначим через $B_{\varepsilon}(a)$ открытый шар в $\mathbb{R}^d$ радиуса ε с центром в точке a. Тогда

$$\mathbb{P}\{|\xi_n - a| < \varepsilon\} = \mathbb{P}\{\xi_n \in B_{\varepsilon}(a)\} \overset{n\to\infty}{\to} \mathbb{P}\{a \in B_{\varepsilon}(a)\} = 1,$$

согласно теореме портманто. □

14.0.1 Доказательство импликации ⇐ в теореме 12.11

Предположим, что

$$F_{\xi_n}(x) \to F_{\xi}(x) \quad \text{при} \quad n \to \infty$$

для всех x, в которых функция F_{ξ} непрерывна. Докажем, что $\xi_n \xrightarrow{d} \xi$, воспользовавшись теоремой 14.1(2b). Отметим, что при доказательстве последней мы пользовались только уже обоснованной, обратной к рассматриваемой импликацией теоремы 12.11.

Для любой $f \in C_b(\mathbb{R}^d)$ имеем $|f(\widetilde{\xi}_n)| \le \|f\|_{\infty}$, а потому, согласно теореме Лебега о мажорируемой сходимости,

$$\mathbb{E}f(\xi_n) = \mathbb{E}f(\widetilde{\xi}_n) \to \mathbb{E}f(\widetilde{\xi}) = \mathbb{E}f(\xi),$$

где в первом и последнем равенствах мы пользовались тем, что распределения $\widetilde{\xi}_n$ и ξ_n, а также $\widetilde{\xi}$ и ξ, совпадают. □

15 Усиленный ЗБЧ, а также ЦПТ для разнородных независимых случайных величин

В этом параграфе мы усиливаем доказанные ранее закон больших чисел и центральную предельную теорему.

Усиленный закон больших чисел

Разница усиленного ЗБЧ с законом больших чисел в форме Хинчина (теорема 13.1) состоит лишь в том, что в усиленном ЗБЧ имеет место сходимость *почти наверное*. Обдумайте: это именно то, что хотелось бы видеть в эксперименте.

Теорема 15.1 (Усиленный ЗБЧ). *В теореме 13.1 имеет место сходимость почти наверное.*

На лекции был дан скетч доказательства в дополнительном предположении $\mathbb{E}\xi_j^4 < \infty$. Детали см. в [Shi-2, теорема 4.3.1].

ЦПТ при условии Линдеберга

Имеется много версий для ЦПТ, отличающихся формулировками условий на последовательность $\{\xi_k\}$, выводы касаются сходимости распределения сумм S_n в последовательности к (необязательно стандартному) гауссову закону. В этом разнообразии условий можно выделять те, в которых не требуется одинаковая распределенность ξ_k, а условия независимости сохранены. В этом контексте наиболее известны варианты достаточных условий Линдеберга. Обозначим математические ожидания $m_k = \mathbb{E}\xi_k$, и дисперсии $\sigma_k^2 = \operatorname{Var}\xi_k$, $D_n^2 := \operatorname{Var}(S_n)$:

Теорема 15.2 (Классическое условие Линдеберга). *Пусть для любого $\varepsilon > 0$ и $n \to \infty$*

$$\frac{1}{D_n^2}\sum_{k=1}^{n}\int\limits_{\{x:|x-m_k|\geqslant\varepsilon D_n\}}(x-m_k)^2 d\mathbb{P}_{\xi_k} \longrightarrow 0$$

Тогда справедлива (классическая) Центральная Предельная Теорема:

$$\frac{S_n - \mathbb{E}(S_n)}{D_n} \xrightarrow{d} \mathcal{N}(0,1)$$

Доказательство основано на той же схеме предела $\varphi_n(t) \to e^{-t^2/2} = \varphi_{\mathcal{N}(0,1)}(t)$ характеристических функций $\varphi_n(t)$ случайных величин $\frac{S_n-\mathbb{E}(S_n)}{D_n}$. Доказательство сходимости не очень короткое, см. [Shi-1, теорема 3.4.1] или [KS, теорема 10.3].

Отметим, что условие Линдеберга **не является необходимым** для наличия ЦПТ в случае независимых случайных величин.

Пример

Пусть каждая из независимых с.в. распределена равномерно на симметричном отрезке $[-a_k, a_k]$, т. е. с плотностью $\frac{1}{2a_k}$, тем самым $\sigma_k^2 = \frac{1}{3}a_k^2$. Если все $a_k < const$ а ряд $\sum_k a_k^2$ расходится, то условие Линдеберга для ЦПТ выполнены, потому что интегрирование в формуле условия для достаточно большого показателя n происходит по области, где плотность нулевая.

С другой стороны, если ряд $\sum_k a_k^2$ сходится, то все D_n с ростом n останутся ограниченными и условие может не выполняться и центральная предельная теорема не применима, во всяком случае сходимости случайных величин S_n/D_n именно к гауссовой случайной величине нет. Менее очевидный случай, где центральная предельная теорема не применима, — это случай $a_k^2 = 2^k$, тогда $3D_n^2 = 2^{n+1} - 2 < a_n^2$ и левая часть условия Линдеберга будет больше $1/2$ при достаточно маленьких значениях t. Пока что на этом примере видна возможность нарушения условий Линдеберга.

16 Многомерное нормальное распределение

Благодаря центральной предельной теореме, гауссовское распределение имеет чрезвычайное значение для теории вероятностей. Сейчас мы обсудим его многомерное обобщение, имеющее не меньшее значение благодаря многомерной версии ЦПТ. Определить многомерное нормальное распределение можно несколькими эквивалентными способами. Мы приведем сразу три. В этом параграфе все вектора — столбцы, а T обозначает транспонирование.

Определение 16.1. Говорят, что случайный вектор $\xi = (\xi_1, \dots, \xi_d)$ имеет *многомерное нормальное распределение* (м.н.р.), или является *гауссовским*, или что случайные величины $\xi_1, \dots, \xi_d$ *совместно гауссовские*, если выполнено любое из следующих эквивалентных утверждений:

1. Существует случайный вектор $\widetilde{\xi} \in \mathbb{R}^d$, распределение которого совпадает с распределением ξ, имеющий вид
$$\widetilde{\xi} = A\eta + m, \tag{16.1}$$
где A — $d \times d$-матрица, $m \in \mathbb{R}^d$, а $\eta = (\eta_1, \dots, \eta_d)$ — случайный вектор с независимыми компонентами $\eta_j \sim \mathcal{N}(0, 1)$.

2. Характеристическая функция φ_ξ вектора ξ имеет вид
$$\varphi_\xi(t) = e^{im \cdot t - \frac{1}{2} t^T \Sigma t}, \quad t = (t_1, \dots, t_d), \tag{16.2}$$
где $m \in \mathbb{R}^d$, $m \cdot t$ обозначает (стандартное) скалярное произведение векторов m и t, а Σ — симметричная неотрицательно определенная $d \times d$-матрица.

3. Для любого вектора $v \in \mathbb{R}^d$ случайная величина $v \cdot \xi$ имеет нормальное распределение.

Так как имеется взаимно однозначное соответствиями между характеристическими функциями и распределениями, м.н.р. однозначно задается параметрами m и Σ из пункта 2. Их смысл таков:
$$m = \mathbb{E}\xi, \qquad \Sigma = (\sigma_{ij})_{1 \le i,j \le n}, \quad \text{где} \quad \sigma_{ij} = \mathrm{Cov}(\xi_i, \xi_j). \tag{16.3}$$

Мы проверим (16.3) в ходе доказательства импликации $3 \Rightarrow 2$. Матрица Σ называется *матрица ковариаций* вектора ξ. На ее диагонали стоят дисперсии $\mathrm{Var}\,\xi_j$. В частности, при $d = 1$, $\Sigma = \mathrm{Var}\,\xi$. Пишут
$$\xi \sim \mathcal{N}(m, \Sigma).$$

Параметр m из первого определения тот же самый, $m = \mathbb{E}\widetilde{\xi} = \mathbb{E}\xi$, что очевидно из представления (16.1). Матрица A такова, что $AA^T = \Sigma$. Мы покажем это в ходе доказательства импликации $1 \Rightarrow 2$. Вообще говоря, A определена не однозначно.

Замечания

- В определении 1 нельзя просто потребовать, чтобы $\xi = A\eta + m$, не вводя дополнительную случайную величину $\widetilde{\xi}$. Действительно, аналогично одномерному случаю, тождественно нулевая случайная величина $\xi \equiv 0$ включается в класс гауссовских: можно выбрать некоторое вероятностное пространство $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{\mathbb{P}})$, на котором определены гауссовские случайные величины η_j, и положить $\widetilde{\xi} = A\eta + m$ с $A = m = 0$, так что $\widetilde{\xi} \equiv 0$. Но может оказаться так, что исходное вероятностное пространство $(\Omega, \mathcal{F}, \mathbb{P})$, на котором определена случайная величина $\xi \equiv 0$, слишком бедно, чтобы поселить на нем независимые гауссовские величины η_j (например, если $\Omega = \{0\}$, $\mathbb{P}\{0\} = 1$ и $\xi(0) = 0$).

- Любой подвектор гауссовского вектора тоже гауссовский. Это сразу следует из определения 1.

- Гауссовость отдельных компонент вектора ξ не влечет гауссовость всего вектора, см. задачу из семинаров 8.6. Однако, если компоненты независимы, то влечет — проверьте это, используя определение 1 или 2.

Доказательство эквивалентности трех определений гауссовости

$1 \Rightarrow 2$: Так как $\mathbb{P}_\xi = \mathbb{P}_{\widetilde{\xi}}$, достаточно вычислить характеристическую функцию вектора $\widetilde{\xi}$. Имеем

$$\varphi_{\widetilde{\xi}}(t) = e^{i\big(A\eta+m\big)\cdot t} = \mathbb{E}e^{im\cdot t}\mathbb{E}e^{iA\eta\cdot t} = e^{im\cdot t}\mathbb{E}e^{i\eta\cdot(A^T t)} = e^{im\cdot t}\mathbb{E}\prod_{k=1}^{d} e^{i\eta_k (A^T t)_k},$$

где $(A^T t)_k$ обозначает k-ую компоненту вектора $A^T t$. Так как η_k независимы и имеют распределение $\mathcal{N}(0,1)$, находим

$$\varphi_{\widetilde{\xi}}(t) = e^{im\cdot t}\prod_{k=1}^{d}\varphi_{\eta_k}\big((A^T t)_k\big) = e^{im\cdot t}e^{-\frac{1}{2}\sum_{k=1}^{d}(A^T t)_k^2} = e^{im\cdot t - \frac{1}{2}t^T\Sigma t},$$

где $\Sigma = AA^T$.

$2 \Rightarrow 1$: Известно, что для любой неотрицательно определенной симметричной $d\times d$ матрицы Σ можно подобрать $d\times d$ матрицу A такую, что $\Sigma = AA^T$ (вообще говоря, не единственную). Тогда, проходя цепочку равенств, образующих доказательство импликации $1 \Rightarrow 2$, в обратную сторону, находим, что

$$\varphi_\xi(t) = e^{i\big(A\eta+m\big)\cdot t}$$

для некоторого вектора $\eta = (\eta_1, \dots, \eta_d)$ с независимыми компонентами, имеющими стандартное нормальное распределение $\eta_j \sim \mathcal{N}(0,1)$. Следовательно, распределение вектора ξ совпадает с распределением $\widetilde{\xi} := A\eta + m$.

$2 \Rightarrow 3$: Для произвольного $v \in \mathbb{R}^d$ и $s \in \mathbb{R}$ имеем

$$\varphi_{v\cdot\xi}(s) = \mathbb{E}e^{is(v\cdot\xi)} = \varphi_\xi(sv) = e^{is(m\cdot v) - \frac{s^2}{2}v^T\Sigma v}.$$

Согласно (11.5), $\varphi_{v\cdot\xi}(s)$ имеет вид характеристической функции нормального распределения $\mathcal{N}(m\cdot v, v^T\Sigma v)$. Остается напомнить, что характеристическая функция однозначно определяет распределение.

$3 \Rightarrow 2$: Обозначим матожидание и матрицу ковариаций вектора ξ как в (16.3). Заметим, что $\mathbb{E}(v \cdot \xi) = v \cdot m$, а

$$\operatorname{Var}(v \cdot \xi) = \sum_{1 \le i,j \le d} v_i v_j \operatorname{Cov}(\xi_i, \xi_j) = v^T \Sigma v.$$

Тогда, по предположению, $v \cdot \xi \sim \mathcal{N}(v \cdot m, v^T \Sigma v)$, так что

$$\varphi_{v \cdot \xi}(1) = e^{iv \cdot m - \frac{1}{2} v^T \Sigma v}.$$

Остается заметить, что

$$\varphi_{v \cdot \xi}(1) = \mathbb{E} e^{iv \cdot \xi} = \varphi_\xi(v).$$

□

Некоторые свойства гауссовских векторов

Сумма совместно гауссовских случайных величин снова гауссовская случайная величина, согласно определению 3. Более общо, класс гауссовских распределений инвариантен относительно аффинных преобразований:

Лемма 16.2. *Пусть ξ — d-мерный гауссовский вектор, $d \ge 1$, C — матрица размера $k \times d$, $k \ge 1$, а $u \in \mathbb{R}^k$. Тогда вектор $C\xi + u$ имеет гауссовское распределение.*

Доказательство. Для произвольного вектора $v \in \mathbb{R}^d$ имеем $v \cdot (C\xi + u) = \xi \cdot (C^T v) + v \cdot u$. Согласно определению 3, $\xi \cdot (C^T v)$ имеет нормальное распределение, а сдвиг на константу $v \cdot u$ лишь сдвигает его матожидание. □

Задача 16.3. Найдите характеристики вектора $C\xi + u$ из предыдущей леммы.

Как мы знаем, если случайные величины α и β независимы, то $\operatorname{Cov}(\alpha, \beta) = 0$, однако обратное, вообще говоря, не верно. Но только не для гауссовского распределения:

Лемма 16.4. *Совместно гауссовские случайные величины (α, β) независимы тогда и только тогда, когда* $\operatorname{Cov}(\alpha, \beta) = 0$.

Доказательство. Пусть случайные величины α и β некоррелированы. Так как матрица ковариаций вектора (α, β) диагональна, его характеристическая функция (16.2) принимает вид

$$\varphi_{(\alpha,\beta)}(t_1, t_2) = \varphi_\alpha(t_1) \varphi_\beta(t_2).$$

Согласно теореме 11.8, случайные величины (α, β) независимы. □

Отметим, что гауссовости каждой из случайных величин α и β по отдельности недостаточно: необходимо требовать именно *совместную* гауссовость (приведите контрпример!)

Задача 16.5. Докажите, что если $(\xi_1, \dots, \xi_d)$ — гауссовский вектор, то для любых фиксированных i, j компоненты ξ_i и ξ_j независимы тогда и только тогда, когда $\operatorname{Cov}(\xi_i, \xi_j) = 0$.

Плотность в невырожденном случае

Если матрица ковариаций Σ невырождена, то такое гауссовское распределение называют невырожденным. В этом случае гауссовский случайный вектор $\xi = (\xi_1, \dots, \xi_d)$ имеет плотность, по форме аналогичную плотности одномерного нормального распределения:

$$\rho_\xi(x) = \frac{1}{(2\pi)^{d/2} \sqrt{\det \Sigma}} \exp\Big(-\frac{1}{2} (x - m)^T \Sigma^{-1} (x - m) \Big), \qquad x = (x_1, \dots, x_d).$$

Это проверяется непосредственно (проделайте вычисления!): для вектора η из определения 1 гауссовости плотность имеет вид $p_\eta(x) = p_{\eta_1}(x_1)\cdots p_{\eta_n}(x_n)$ в силу независимости его компонент, где $p_{\eta_j}(x_j)$ — плотность стандартного нормального распределения $\mathcal{N}(0,1)$. Далее нужно заменить переменные в интеграле. Отметим, что матрица A из определения 1 также невырождена, так как $AA^T = \Sigma$.

Детали этого доказательства, а также больше информации про многомерное нормальное распределение, можно найти, например, в [Shi-1].

A Напоминание: комбинаторика и простейшие формулы

A.1 Простейшие базовые формулы

- **Формула включений и исключений**. Пусть $|A|$ обозначает число элементов в конечном множестве A. Тогда для (конечного) семейства конечных множеств A_i, $i = 1, 2, \ldots n$ выполнено:

$$|\bigcup_{i=1}^n A_i| = \sum_i^n |A_i| - \sum_{i_1<i_2} |A_{i_1} \cap A_{i_2}| + \sum_{i_1<i_2<i_3} |A_{i_1} \cap A_{i_2} \cap A_{i_3}| - \ldots + (-1)^{n-1} |A_{i_1} \cap A_1 \cap A_2 \cap \ldots \cap A_n|$$

  Доказательство математической индукцией по подсчету вхождений каждого элемента из $\bigcup_{i=1}^n A_i$ в множества из правой части равенства.

- **Таблицы свойств**: пусть имеется k списков, в каждом таком списке перечислено N_k характеристик, надо сосчитать сколько возможных вариантов набора характеристик (по одной из каждого списка). Например, если списков два, то варианты удобно представлять в виде позиций в прямоугольной таблице – это дает ответ $N_1 \cdot N_2$. Многомерный аналог этого подхода дает ответ $N_1 \cdot N_2 \cdot \ldots N_k$.

- **Размещения**: сколько способов построить n солдат в шеренгу длины k. Для правого фланга есть n вариантов, следующее по порядку место заполняется $n-1$ способами итд. Всего вариантов получается $n \cdot (n-1) \cdot (n-2) \cdot \ldots (n-k+1)$, что традиционно обозначается как A_n^k.

- **Перестановки** Размещения с $n = k$ традиционно называются перестановками, обозначаются $\prod_n = n!$, где по определению $0! = 1$.

- **Сочетания**: сколько способов собрать из n солдат взвод в k человек. Ясно, что такой взвод можно построить в шеренгу $\prod_k$ способами, откуда вытекает число способов формирования взвода $\frac{A_n^k}{\prod_k} = \frac{n!}{k!(n-k)!}$, оно называется числом сочетаний или биномиальным коэффициентом, обозначалось ранее как C_n^k, но в современной литературе [32] как $\binom{n}{k}$.

A.2 Некоторые примеры

Перестановки разнотипных объектов или перестановки с повторениями

Пусть имеются предметы k различных типов. Сколько различных перестановок можно сделать в последовательности из n_1 предметов первого типа, n_2 предметов второго типа, $\ldots$, n_k предметов k -го типа? Число элементов в последовательности равно $n = \sum_{i=1}^k n_k$. Поскольку различимы элементы только по их типам, то общее число перестановок будет меньше $\prod_n$: некоторые перестановки надо отождествить. В самом деле, однотипные

[32] кроме того для любого действительного x полагают $\binom{x}{k} = \frac{x(x-1)(x-2)\ldots(x-k+1)}{k!}$

предметы можно переставлять между собой и это даст неотличимую от исходной последовательность. Для разных типов это можно делать независимо и таких «внутренних» перестановок будет соответственно $n_1!$ в первом типе, $n_2!$ во втором итд. Таким образом общее число действительно различных перестановок получится таким:

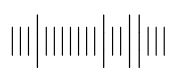
$$\frac{(n_1 + n_2 + \ldots + n_k)!}{n_1! \cdot n_2! \cdot \ldots \cdot n_k!}$$

A.2.1 Разложения шаров по ящикам. Различимые и неразличимые объекты

Пусть имеются n шаров и $k \leqslant n$ ящиков, мы собираемся сосчитать количество конфигураций, когда все шары разложены по ящикам. Совершенно очевидно, что прежде всего необходимо договориться, какие кофигурации следует считать разными, это приводит к следующщим задачам с разными условиями:

1. **Все ящики различимы между собой и все шары различимы между собой**. Условие означает, что можно пронумеровать предметы и тогда для каждого шара возникает ровно k возможностей, по формуле таблиц общее число раскладок будет k^n

2. **Все ящики различимы между собой, а все шары неразличимы между собой**. Можно закодировать каждую такую раскладку схематической картинкой: палочки, обозначающие стенки плотно стоящих ящиков слева направо и шары между ними. Достаточно укзать только $k-1$ стенку – самая левая и самая правая стенки ничего не прибавляют к знанию о раскладке. А теперь, если шары тоже нарисовать палочками, но поменьше размером, то получится такая, например, картинка, кодирующая раскладку 3,7,2,0,4 шестнадцати одинаковых шаров по пяти различимым (по их порядку) ящикам:



В общем случае всего палочек $n+k-1$ и из них $k-1$ длинных. Всего таких картинок (и раскладок!) получается $\binom{n+k-1}{k-1}$.

3. **Все ящики неразличимы между собой и все шары неразличимы между собой**. Здесь все кодируется числом способов разбиения числа n на не более, чем $k \leqslant n$ ненулевых слагаемых, порядок которых неважен. Поэтому всякому такому разбиению $n = n_1 + n_2 + \ldots + n_j$ можно сопоставить картинку: $j \leqslant k$ выровненных по левой границе горизонтальных полосок (каждая из некоторого числа клеточек, всего клеточек n), нарисованных друг под другом в порядке (нестрогого) убывания длин [33]. Такая картинка см. Рис.3 из клетчатых полосок называется диаграммой Юнга. Нас интересует подсчет всех таких табличек с общим числом клеток n и числом строк не более k.

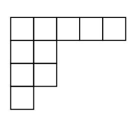


Рис. 3: Пример диаграммы Юнга

Первым делом заметим, что каждую такую табличку можно *транспонировать* – строки по порядку выписать столбцами – тогда опять получится диаграмма Юнга

[33]французские математики предпочитают рисовать их по возрастанию

из n клеток уже с любым (естественно, не превосходящим n) количеством строк, но у которой длина любой строки теперь не превосходит k. Мы свели исходную задачу к подсчету $p(n,k)$ – количества разбиений числа n на слагаемые, *каждое из которых* не превосходит k. Продолжение вычислений с диаграммами Юнга см. в разделе методов A.3.1.

4. **Все ящики неразличимы между собой, а все шары различимы между собой**. То есть речь идет о подсчете количества неупорядоченных разбиений множества $\{1, 2, \ldots n\}$ на не более чем k подмножеств. Количество разбиений $\{1, 2, \ldots n\}$ *в точности на i непустых множеств* может быть явно указано, оно называется *числом Стирлинга второго рода* и обычно обозначается как $\left\{ n \atop i \right\}$. Поэтому в нашей задаче про разложение различимых шаров в неразличимые ящики окончательный ответ дается суммой чисел Стирлинга $\left\{ n \atop i \right\}$ по i от единицы до k. Явное вычисление чисел Стирлинга см. далее разделе методов A.3.1.

A.3 Методы вычислений

A.3.1 Рекуррентные соотношения. Треугольник Паскаля

Биномиальные коэффициенты имеют много тождеств и рекуррентных соотношений, например:

$$\binom{n}{r} = \binom{n-1}{r-1} + \binom{n-1}{r}$$

Это соотношение порождает возможность вычислять биномиальные коэффициенты последовательно выписывая значения, выдаваемые рекуррентным соотношением, в треугольную таблицу (треугольник Паскаля), строки которой принято нумеровать от нуля.

$$\begin{array}{ccccccccc}
 & & & & 1 & & & & \\
 & & & 1 & & 1 & & & \\
 & & 1 & & 2 & & 1 & & \\
 & 1 & & 3 & & 3 & & 1 & \\
1 & & 4 & & 6 & & 4 & & 1 \\
. & . & . & . & . & . & . & . & .
\end{array}$$

Из треугольника Паскаля возможно вычислить и множество других очень полезных комбинаторных формул. Индукцией по номеру строки несложно проверить, что в треугольнике Паскаля k стоят коэффициенты многочлена $(a+b)^n$ последовательно возрастающие как раз в соответствии с рекуррентным соотношением. В частности, при $a = b = 1$ возникает тождество:

$$\binom{n}{0} + \binom{n}{1} + \ldots + \binom{n}{n} = 2^n$$

В докомпьютерную эпоху последовательные вычисления при больших n, r биномиальных коэффициентов $\binom{n}{r}$ даже и с помощью треугольника Паскаля представляли трудность, поэтому возникли асимптотические формулы для факториалов, дающие приблизительные вычисления по формуле $\binom{n}{k} = \frac{n!}{k!(n-k)!}$. Часто встречается асимптотическая формула Стирлинга, объяснение которой приведено в разделе про биномиальное распределение.

Производящие функции

В математике часто встречается следующий прием для вычислений. Пусть $a_0, a_1, a_2, \ldots$ — произвольная числовая последовательность $\{a_n\}$. Формальный степенной ряд $\sum_{n=0}^{\infty} a_n q^n$ от переменной q называется производящей функцией для этой последовательности. Например, для последовательности из одних единиц ее производящая функция геометрическая прогрессия — формальный ряд для $\frac{1}{1-q}$, а для (конечной) последовательности биномиальных коэффициентов $\binom{n}{0}, \binom{n}{1} \ldots \binom{n}{n}$ производящая функция есть $(1+q)^n$ итп. Явное знание вида производящей функции позволяет дифференциированием находить ее ряд Маклорена, коэффициенты которого дадут исходную последовательность. Мы применим производящие функции к поставленным выше задачам о разложениях шаров в ящики, больше примеров найдется, например, в книге С.Ландо "Лекции по комбинаторике".

Диаграммы Юнга. Подсчет числа $p(n,k)$ для разбиений

Мы сосредоточимся на вычислении производящей функции $P_k(q)$ для последовательности $\{p(n,k)\}$ с фиксированным k. Ясно что $P_1(q) = \frac{1}{1-q}$ потому что каждое число единственным образом разбивается в сумму единиц. Теперь заметим, что количество способов разбить число n в сумму слагаемых, *каждое из которых* равно двум — это либо 1, если n четно, либо 0, если n нечетно. Чередование единиц и нулей отвечает производящей функции $\frac{1}{1-q^2}$, а $P_2(q) = \frac{P_1(q)}{1-q^2}$. Действительно, раскроем скобки, но не будем пока приводить подобные члены в произведении рядов

$$(1+q+q^2+q^3+\ldots)(1+q^2+q^4+q^6+\ldots)$$

Каждое слагаемое после раскрытия скобок имеет вид $q^r q^{2s}$ и каждому такому слагаемому можно сопоставить разбиение числа $r+2s$ в сумму r единиц и s двоек. А после приведения подобных членов коэффициент при q^n окажется как раз $p(n,2)$. Рассуждая аналогичным образом получаем, что производящая функция $P_k(q) = \sum_n p(n,k)q^n$ равна

$$\frac{1}{(1-q)(1-q^2)\ldots(1-q^k)} = \prod_{m=1}^{k}(1-q^m)^{-1}$$

Последовательным дифференциированием правой части мы найдем ее ряд Маклорена и тем самым необходимый коэффициент $p(n,k)$ — достаточно сложный путь!

Вычисление чисел Стирлинга второго рода

Рассмотрим Y – все *эпиморфные* отображения f множества $S_1 = \{1,2,\ldots n\}$ на множество $S_2 = \{1,2,\ldots k\}$. Каждое такое отображение делит S_1 в точности на k кусков P_i так, что $f(P_i) = i$. Ясно, что поскольку порядок кусков нам не важен, то искомых отображений будет $k!\begin{Bmatrix} n \\ k \end{Bmatrix} = |Y|$. С другой стороны по формуле таблиц все (то есть уже не обязательно эпиморфные) такиие функции составляют множество X из в точности k^n элементов, обозначим множество тех из них, которые в образе не содержат j через X_j. Понятно, что $Y = \bigcap_{j=1}^{k}(X \setminus X_j)$, а потому

$$|Y| = |X - \bigcup_{j=1}^{k} X_j| = k^n - |\bigcup_{j=1}^{k} X_j|$$

При этом ясно, что $|X_j| = (k-1)^n$ и опять-таки по формуле таблиц

$$\sum_{1\leqslant i_1<i_2<\ldots<i_j\leqslant k} |X_{i_1} \cap X_{i_2} \cap \ldots X_{i_j}| = \binom{k}{j}(k-j)^n$$

Применяя формулу включений и исключений получаем

$$k!\left\{ n \atop k \right\} = |Y| = k^n - \left[\sum_{j=1}^{k}(-1)^{j+1}\binom{k}{j}(k-j)^n\right] = \sum_{j=0}^{k}(-1)^j\binom{k}{j}(k-j)^n$$

откуда уже и получается итоговая формула

$$\left\{ n \atop k \right\} = |Y| = \frac{1}{k!}\sum_{j=0}^{k}(-1)^j\binom{k}{j}(k-j)^n$$

A.4 Асимптотики и формула Стирлинга

Начертим график функции $y = \ln x$ и заметим, что величина $I_n = \ln 2 + \ln 3 + \ldots + \frac{1}{2}\ln n$ выражает формулу трапеций для интегральной суммы, оценивающей площадь под графиком функции $y = \ln x$ на отрезке $[1, n]$. Эта интегральная сумма по построению меньше площади под кривой на отрезке $[1, n]$. С другой стороны, для $1 < k < n$ число $\ln k$ в точности равно площади трапеции, образованной осью абцисс, вертикальными прямыми $x = k - \frac{1}{2}$, $x = k + \frac{1}{2}$ и зажатым меж ними отрезком касательной к графику в точке $(k, \ln k)$ – сделайте рисунок всех упомянутых фигур сами! Поэтому

$$\ln\Big[(n-1)!\Big] = \sum_{k=1}^{n-1}\ln k > \int\limits_{[\frac{3}{2},n-\frac{1}{2}]}\ln x\ dx \quad \text{и} \quad \int\limits_{[n-\frac{1}{2},n]}\ln x\ dx < \frac{\ln n}{2} \Longrightarrow I_n > \int\limits_{[\frac{3}{2},n]}\ln x\ dx$$

Имеем неравенства с определенными интегралами от логарифмической функции, которые вычисляются явно так как $\int \ln x\ dx = x\ln x - x$,

$$\int\limits_{[\frac{3}{2},n]}\ln x\ dx < I_n < \int\limits_{[1,n]}\ln x\ dx$$

$$\left(n+\frac{1}{2}\right)\ln n - n + \frac{3}{2}\left(1-\ln\frac{3}{2}\right) < \ln n! < \left(n+\frac{1}{2}\right)\ln n - n + 1$$

$$n^{\left(n+\frac{1}{2}\right)}e^{-n}\cdot\beta < n! < n^{\left(n+\frac{1}{2}\right)}e^{-n}\cdot e$$

$$\beta < \frac{n!}{n^{\left(n+\frac{1}{2}\right)}e^{-n}} < e$$

для $\beta = \exp\left(\frac{3}{2}\left(1-\ln\frac{3}{2}\right)\right) \simeq 1.23586$. То есть мы указали выражение того же порядка величины, что и $n!$, ибо их отношение лежит между 1.23 и 2.7183. Несколько более детальные рассуждения показывают (см. например В.Феллер "Введение в теорию вероятностей и ее приложения том 1, откуда и взято приводимое выше рассуждение), что это отношение с ростом n быстро стремится к пределу, равному $\sqrt{2\pi} \simeq 2.507$ – возникает формула Стирлинга для асимптотики значения факториала. С ее помощью, например, выясняется, что $\binom{2n}{n} \sim \frac{2^{2n}}{\sqrt{\pi n}}$.

B Базовые объекты и конструкции теории меры

В этом аппендиксе мы напоминаем основны теории меры. Доказательств мы не приводим. Зафиксируем некоторое множество Ω.

B.1 Алгебры и сигма-алгебры

Определение B.1. Набор подмножеств $\mathcal{A}$ множества Ω называется *алгеброй*, если
1) $\Omega \in \mathcal{A}$,
2) $A \in \mathcal{A} \quad \Rightarrow \quad A^c := \Omega \setminus A \in \mathcal{A}$,
3) $A_1, A_2 \in \mathcal{A} \Rightarrow A_1 \cup A_2 \in \mathcal{A}$.

Лемма B.2. *Пусть $\mathcal{A}$ — алгебра. Тогда*
1) $\emptyset \in \Omega$,
2) $A_1, \dots, A_n \in \mathcal{A} \quad \Rightarrow \quad \cup_{i=1}^n A_i \in \mathcal{A}$, $\cap_{i=1}^n A_i \in \mathcal{A}$,
3) $A, B \in \mathcal{A} \quad \Rightarrow \quad A \setminus B \in \mathcal{A}$.

Пример B.3. Следующие наборы подмножеств Ω образуют алгебру:
1) $\{\emptyset, \Omega\}$,
2) $2^\Omega :=$ множество всех подмножеств Ω.
3) В случае $\Omega = \mathbb{R}$, набор всевозможных конечных объединений множеств вида (a, b), $(a, b]$, $[a, b]$, $[a, b)$, где $a, b \in \mathbb{R}$.
4) Набор всех конечных подмножеств Ω.
5) Набор всех не более, чем счетных подмножеств Ω.

Определение B.4. Набор подмножеств $\mathcal{F}$ множества Ω называется *σ-алгеброй*, если $\mathcal{F}$ — алгебра, замкнутая относительно операции счетного объединения. То есть, если для любых $A_i \in \mathcal{F}$, $i \geq 1$, выполнено $\cup_{i=1}^\infty A_i \in \mathcal{F}$. Пару $(\Omega, \mathcal{F})$, где $\mathcal{F}$ —σ-алгебра, называют *измеримым пространством.*

Лемма B.5. *Если в определении B.4 $\cup_{i=1}^\infty$ заменить на $\cap_{i=1}^\infty$, то получится эквивалентное определение σ-алгебры.*

Алгебры из примеров B.3(1,2,5) являются σ-алгебрами, а из примеров B.3(3,4) —- вообще говоря нет (если в примере B.3(4) предположить, что множество Ω бесконечно).

Можно показать, что для всякого набора $\mathcal{C}$ подмножеств Ω существует единственная наименьшая сигма-алгебра $\mathcal{F}$ на Ω, содержащая $\mathcal{C}$. Это значит, что если $\widetilde{\mathcal{F}}$ — другая σ-алгебра, содержащая $\mathcal{C}$, то она содержит и $\mathcal{F}$. То есть, $\forall A \in \mathcal{F}$ удовлетворяет $A \in \widetilde{\mathcal{F}}$. Наименьшая сигма-алгебра $\mathcal{F}$ обозначается $\sigma(\mathcal{C})$.

Определение B.6. Если Ω — топологическое пространство, то *борелевской* сигма-алгеброй $\mathcal{B}(\Omega)$ называется наименьшая сигма-алгебра, содержащая все открытые подмножества Ω. Эквивалентно, содержащая все замкнутые подмножества. Подмножества $A \in \mathcal{B}(\Omega)$ называются *борелевскими.*

Например, $\mathcal{B}(\mathbb{R}) = \sigma($все интервалы$) = \sigma($все множества из примера B.3(3)$)$.

B.2 Мера

Напомним, что функция множеств μ называется *аддитивной*, если для любых не пересекающихся множеств $A_1, \dots, A_n$ из ее области определения, $A_i \cap A_j = \emptyset$ при $i \neq j$, выполнено $\mu(\cup_{i=1}^n A_i) = \sum_{i=1}^n \mu(A_i)$. Разумеется, выполнение последнего равенства для $n = 2$ влечет его выполнение для произвольного n.

Определение B.7. Пусть дано измеримое пространство $(\Omega, \mathcal{F})$. Аддитивная функция множеств $\mu : \mathcal{F} \mapsto [0, \infty)$ называется *мерой* (конечной, неотрицательной), если она обладает свойством счетной аддитивности. То есть, для любых не пересекающихся подмножеств $A_i \in \mathcal{F}$, $i \geq 1$, $A_i \cap A_j = \emptyset$ при $i \neq j$, выполнено

$$\mu\Big(\bigcup_{i=1}^{\infty} A_i\Big) = \sum_{i=1}^{\infty} \mu(A_i).$$

Для краткости, говоря "мера" мы обычно мы будем подразумевать конечную и неотрицательную меру. Отметим, что в некоторых источниках σ-аддитивность не включается в определение меры, откуда появляется выражение "σ-аддитивная мера".

Борелевские сигма-алгебры являются самыми типичными областями определения мер.

Пример B.8. 1) Мера Лебега на измеримом пространстве $\big([0,1], \mathcal{B}[0,1]\big)$ — мера.

2) Пусть Ω — не более, чем счетное множество, а функция $p : \Omega \mapsto [0, \infty)$ удовлетворяет соотношению $\sum_{\omega \in \Omega} p(\omega) < \infty$. Тогда функция множеств μ, определенная равенством

$$\mu(A) := \sum_{\omega \in \Omega} p(\omega), \quad A \subset \Omega,$$

задает меру на измеримом пространстве $(\Omega, 2^{\Omega})$.

3) Пусть множество Ω — счетно, а $\mathcal{A}$ — алгебра всех конечных подмножеств в Ω и их дополнений. Рассмотрим функцию множеств m, определенную равенствами

$$m(A) = 0, \quad m(A^c) = 1, \quad \text{для любого конечного подмножества } A. \tag{B.1}$$

Эта функция аддитивна, но не счетно-аддитивно, так что она не является мерой. Да и определена она не на сигма-алгебре, а лишь на алгебре.

4) Пусть множество Ω — несчетно, а $\mathcal{A}$ — алгебра всех счетных подмножеств в Ω и их дополнений. Заметим, что $\mathcal{A}$ является σ-алгеброй. Рассмотрим функцию множеств m, определенную соотношением (B.1), где A — счетные подмножества. Она счетно-аддитивна, а значит задает меру.

Важную роль в теории вероятностей играет свойство мер, называемое *непрерывностью* — см. пункты (2-4) следующей леммы.

Лемма B.9. *Пусть $(\Omega, \mathcal{F})$ — измеримое пространство, а $\mu : \mathcal{F} \mapsto [0, \infty)$ — аддитивная функция множеств. Следующие свойства эквивалентны:*

1) [Счетная аддитивность] *μ — счетно-аддитивна, то есть является мерой.*

2) [Непрерывность снизу] *Для любой возрастающей последовательности множеств A_i, $i \geq 1$ (то есть такой, что $A_i \subset A_{i+1}$) выполнено $\mu(A_i) \to \mu(A)$ при $i \to \infty$, где $A := \cup_{i=1}^{\infty} A_i$. Это можно записать так:*

$$A_i \uparrow A \quad \Rightarrow \quad \mu(A_i) \to \mu(A) \quad \textit{при} \quad i \to \infty.$$

3) [Непрерывность сверху] *Для любой убывающей последовательности множеств B_i, $i \geq 1$ (то есть такой, что $B_i \supset B_{i+1}$), выполнено $\mu(B_i) \to \mu(B)$ при $i \to \infty$, где $B := \cap_{i=1}^{\infty} B_i$. Это можно записать так:*

$$B_i \downarrow B \quad \Rightarrow \quad \mu(B_i) \to \mu(B) \quad \textit{при} \quad i \to \infty.$$

4) [Непрерывность в нуле] *Для любой убывающей к нулю последовательности множеств C_i, $i \geq 1$ (то есть такой, что $C_i \supset C_{i+1}$ и $\cap_{i=1}^{\infty} C_i = \emptyset$), выполнено $\mu(C_i) \to 0$ при $i \to \infty$. Это можно записать так:*

$$C_i \downarrow \emptyset \quad \Rightarrow \quad \mu(C_i) \to 0 \quad \textit{при} \quad i \to \infty.$$

Задача B.10. Докажите лемму.

B.3 Прямое произведение пространств с мерой

Пусть $(\Omega_i, \mathcal{F}_i)$ — измеримые пространства с мерами μ_i, $1 \le i \le n$.

Определение B.11. *Прямым произведением* измеримых пространств $(\Omega_i, \mathcal{F}_i)$ называется измеримое пространство $(\Omega, \mathcal{F})$, где $\Omega = \Omega_1 \times \cdots \times \Omega_n$, а $\mathcal{F} = \sigma(\mathcal{F}_1 \times \cdots \times \mathcal{F}_n)$. Здесь $\mathcal{F}_1 \times \cdots \times \mathcal{F}_n$ обозначает набор множеств вида вида $A_1 \times \cdots \times A_n$ с $A_i \in \mathcal{F}_i$.

Чтобы пояснить это определение, отметим, что прямое произведение $\mathcal{F}_1 \times \cdots \times \mathcal{F}_n$ сигма-алгебр $\mathcal{F}_i$ вообще говоря уже не является σ-алгеброй. Чтобы это увидеть, достаточно положить $n = 2$, $\Omega_i = \mathbb{R}$ и $\mathcal{F}_i = \mathcal{B}(\mathbb{R})$, и заметить, что разность прямоугольника со вложенным в него меньшим прямоугольником уже не прямоугольник.

Рассмотрим функцию множеств $\mu = \mu_1 \times \cdots \times \mu_n$, определенную на наборе множеств $\mathcal{F}_1 \times \cdots \times \mathcal{F}_n$ равенством

$$\mu(A_1 \times \cdots \times A_n) = \mu_1(A_1) \cdots \mu_n(A_n), \qquad \forall A_i \in \mathcal{F}_i.$$

Возникает естественный вопрос, можно ли ее продолжить до меры на измеримом пространстве $(\Omega, \mathcal{F})$? Оказывается, можно.

Определение B.12. Набор множеств $\mathcal{R}$ называется *полукольцом*, если $\emptyset \in \mathcal{R}$,

(1) $A \cap B \in \mathcal{R}$ для всех $A, B \in \mathcal{R}$,

(2) Если $A, B \in \mathcal{R}$ и $A \subset B$, то существуют непересекающиеся множества $C_1, \ldots, C_k \in \mathcal{R}$, такие что $B \setminus A = C_1 \cup \cdots \cup C_k$.

Легко проверить, что $\mathcal{F}_1 \times \cdots \times \mathcal{F}_n$ является полукольцом (удобно представить себе разность вложенных прямоугольников на плоскости). А теперь сформулируем знаменитую

Теорема B.13 (теорема Каратеодори о продолжении меры)**.** *Пусть счетно-аддитивная функция множеств $\nu : \mathcal{R} \mapsto [0, \infty)$ определена на полукольце $\mathcal{R}$. Тогда она однозначно продолжается до меры на минимальной сигма-алгебре $\sigma(\mathcal{R})$, содержащей $\mathcal{R}$.*

Ниже, в лемме B.16, мы проверяем счетную аддитивность функции множеств μ. Тогда, применяя теорему Каратеодори, мы видим, что μ однозначно продолжается до меры на измеримом пространстве $(\Omega, \mathcal{F})$, которую мы будет обозначать тем же символом μ.

Определение B.14. Измеримое пространство $(\Omega, \mathcal{F})$, снабженное мерой μ, называется *прямым произведением* измеримых пространств $(\Omega_i, \mathcal{F}_i)$, снабженных мерами μ_i.

Пример B.15. Пусть $\Omega_i = [a_i, b_i]$, с действительными $a_i < b_i$, $\mathcal{F}_i = \mathcal{B}([a_i, b_i])$ — борелевская σ-алгебра, а μ_i — мера Лебега на $\mathcal{B}([a_i, b_i])$. Тогда Ω — прямоугольник в $\mathbb{R}^n$, и можно доказать, что $\mathcal{F} = \mathcal{B}(\Omega)$, а μ — мера Лебега на $\mathcal{B}(\Omega)$.

Отметим, что если меры μ_i были вероятностными, то и мера μ также окажется вероятностной:

$$\mu(\Omega) = \prod_{i=1}^{n} \mu_i(\Omega_i) = 1.$$

Лемма B.16. *Функция μ, определенная на полукольце множеств $\mathcal{R} := \mathcal{F}_1 \times \cdots \times \mathcal{F}_n$, σ-аддитивна. То есть, для любого набора не пересекающихся множеств $A_i \in \mathcal{R}$, таких что $\cup_{i=1}^{\infty} A_i \in \mathcal{R}$, выполнено*

$$\mu(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu(A_i).$$

Доказательство. Проведем доказательство для $n = 2$. Пусть $A_i = B_i \times C_i$, где $B_i \in \mathcal{F}_1$, $C_i \in \mathcal{F}_2$, и $\cup_{i=1}^{\infty} B_i \times C_i = B \times C$, где $B \in \mathcal{F}_1$, $C \in \mathcal{F}_2$. Рассмотрим последовательность функций

$$f_k : \Omega_1 \mapsto \mathbb{R}, \quad f_k(\omega_1) = \sum_{i=1}^{k} \mathbb{I}_{B_i}(\omega_1)\mu_2(C_i),$$

где $\mathbb{I}_{B_i}$ обозначает индикаторную функцию множества B_i. Аналогично положим $f(\omega_1) = \mathbb{I}_B(\omega_1)\mu_2(C)$. Заметим, что $f_k \le \mu_2(B)$ и $f_k(\omega_1) \to f(\omega_1)$ при $k \to \infty$. Тогда, согласно теореме Лебега о мажорируемой сходимости,

$$\mu\big(\cup_{i=1}^{k} B_i \times C_i\big) = \sum_{i=1}^{k} \mu(B_i)\mu(C_i) = \int_{\Omega_1} f_k(\omega_1)\, d\mu_1(\omega_1) \overset{k\to\infty}{\to} \int_{\Omega_1} f(\omega_1)\, d\mu_1(\omega_1)$$
$$= \mu_1(B)\mu_2(C) = \mu(B \times C).$$

Доказательство для произвольного n можно провести по индукции. □

B.4 Образ меры при отображении

Пусть имеется два измеримых пространства $(X, \mathcal{F})$ и $(Y, \mathcal{G})$, и на пространстве $(X, \mathcal{F})$ живет мера μ. Пусть $f : (X, \mathcal{F}) \mapsto (Y, \mathcal{G})$ — измеримое отображение. Тогда на пространстве $(Y, \mathcal{G})$ определена мера $f_*(\mu)$, называемая *образом меры μ при отображении f* (pushforward меры), заданная соотношением

$$f_*(\mu)(A) := \mu\big(f^{-1}(A)\big), \quad \forall A \in \mathcal{G}. \tag{B.2}$$

B.5 Типы мер

B.5.1 Сингулярные меры

Определение B.17. Мера μ на $\mathcal{B}(\mathbb{R}^n)$ называется *сингулярной*, если существует множество $A \in \mathcal{B}(\mathbb{R}^n)$ нулевой меры Лебега, такое что $\mu(A) = \mu(\mathbb{R}^n)$.

Другими словами, мера μ сосредоточена на множестве нулевой меры Лебега. Выделяют два типа сингулярных мер: дискретные и непрерывные.

• **Дискретные меры.** Это меры μ на $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$, сосредоточенные на не более, чем счетных подмножествах $\mathbb{R}^n$. То есть, существует не более, чем счетное множество $X = \{x_1, x_2, \dots\} \subset \mathbb{R}^n$ для которого $\mu(X) = \mu(\mathbb{R}^n)$.

Напомним, что *дельта-мерой* в точке $x \in \mathbb{R}^n$ называется (вероятностная) мера δ_x, такая что

$$\delta_x(A) = \begin{cases} 1, & \text{если } x \in A, \\ 0, & \text{если } x \notin A, \end{cases} \quad \text{для всех} \quad A \in \mathcal{B}(\mathbb{R}^n). \tag{B.3}$$

Определение выше означает, что дискретная мера μ представляется в виде

$$\mu = \sum_j p_j \delta_{x_j}, \tag{B.4}$$

для некоторых чисел $p_j \ge 0$, таких что $\sum_j p_j < \infty$. Говоря более подробно, (B.4) означает, что

$$\mu(A) = \sum_j p_j \delta_{x_j}(A) = \sum_{j:\, x_j \in A} p_j, \quad \forall A \in \mathcal{B}(\mathbb{R}^n).$$

Мера μ оказывается вероятностной, если $\sum_j p_j = 1$.

В случае $n = 1$ дискретные меры удобно записывать в виде таблиц с двумя строками, в перовой из которых находятся точки x_j, а во второй числа p_j.

• **Непрерывные сингулярные меры.** Это такие сингулярные меры μ на $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$, что $\mu(x) = 0$ для любого $x \in \mathbb{R}^n$. Такие меры существуют, однако это экзотика. Для примера см. конец параграфа 4.3, про сингулярные случайные величины.

Нетрудно показать, что любая сингулярная мера μ единственным образом представляется в виде суммы $\mu_d + \mu_s$ дискретной и сингулярной непрерывной мер.

B.5.2 Абсолютно непрерывные меры

Определение B.18. Напомним, что мера μ на $\mathcal{B}(\mathbb{R}^n)$ называется *абсолютно непрерывной*, если для любого множества $A \in \mathcal{B}(\mathbb{R}^n)$ нулевой меры Лебега выполнено $\mu(A) = 0$.

Теорема Радона-Никодима утверждает, что всякая абсолютно непрерывная мера μ имеет *плотность*: существует интегрируемая по Лебегу функция $\rho : \mathbb{R}^n \mapsto [0, \infty)$, такая что

$$\mu(A) = \int_A \rho(x)\, dx, \quad \forall A \in \mathcal{B}(\mathbb{R}^n). \tag{B.5}$$

Это принято обозначать следующим образом:

$$\mu(dx) = \rho(x)\, dx \quad \text{или} \quad d\mu = \rho\, dx.$$

Пусть теперь, наоборот, дана интегрируемая по Лебегу функция $\rho : \mathbb{R}^n \mapsto [0, \infty)$. Из сигма-аддитивности интеграла Лебега следует, что функция множеств μ, определенная соотношением (B.5), является мерой, а значит ρ является ее плотностью. Таким образом, имеется взаимно-однозначное соответствие между интегрируемыми по Лебегу неотрицательными функциями (плотностями) и мерами (с точностью до эквивалентности, обычной для пространств Лебега). Отметим, что, если мера μ — вероятностная, то ее плотность ρ удовлетворяет

$$\int_{\mathbb{R}^n} \rho(x)\, dx = 1.$$

B.5.3 Теорема о разложении меры

Оказывается, тремя типами мер, перечисленными выше, все и исчерпывается.

Теорема B.19. *Для любой меры μ на $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ существуют и единственны дискретная, сингулярная непрерывная и абсолютно непрерывная меры μ_d, μ_s и μ_a, такие что*

$$\mu = \mu_d + \mu_s + \mu_a.$$

Доказательство можно найти в книге [KS, Теорема 3.16]. Оно дано в случае $n = 1$, но легко видеть, что оно работает и для произвольного $n \geq 1$.

B.6 Интеграл Лебега

Пусть $(\Omega, \mathcal{F})$ — измеримое пространство, μ — мера на нем, а $f : (\Omega, \mathcal{F}) \mapsto (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ — измеримая функция. Цель этого параграфа — напомнить как строится интеграл

$$\int_\Omega f(\omega)\, d\mu(\omega)$$

и некоторые его свойства. См. [KS, параграф 3.1] за более подробным изложением.

B.6.1 Построение интеграла

Шаг 1. Пусть сперва функция f принимает не более, чем счетное число значений, то есть

$$f(x) = \sum_{k=1}^{\infty} c_k \mathbb{I}_{A_k}(x), \quad A_k \in \mathcal{F}, \quad c_k \in \mathbb{R}.$$

Будем называть такую функцию измеримой *простой* (легко видеть, что она действительно измерима). [34] Допустим, что $f \geq 0$. Положим

$$\int_\Omega f(\omega)\, d\mu(\omega) := \sum_{k=1}^{\infty} c_k \mu(A_k), \tag{B.6}$$

если ряд в правой части сходится. Если он расходится, положим $\int_\Omega f\, d\mu := \infty$.

Из определения (B.6) сразу же следует

Лемма B.20. *Если f_1, f_2 — простые неотрицательные функции, то*

1. $\displaystyle\int_\Omega f_1\, d\mu \geq \int_\Omega f_2\, d\mu,$ *если* $f_1 \geq f_2$.
2. $\displaystyle\int_\Omega (a_1 f_1 + a_2 f_2)\, d\mu = a_1 \int_\Omega f_1\, d\mu + a_2 \int_\Omega f_2\, d\mu$ *для произвольных констант* a_1, a_2.

Заметим, что утверждения леммы верны и в случае расходящихся интегралов (напомним, что мы пока рассматриваем только неотрицательные функции).

Шаг 2. Обратимся теперь к случаю, когда $f \geq 0$ — произвольная неотрицательная измеримая функция.

Лемма B.21. *Для произвольной измеримой функции $g : \Omega \mapsto \mathbb{R}$ существует последовательность $(g_n)_{n\geq 1}$ простых измеримых функций, таких что $g_n \leq g_{n+1} \leq \cdots \leq g$ и*

$$|g(\omega) - g_n(\omega)| \leq 2^{-n} \qquad \forall \omega \in \Omega.$$

В частности, $g_n \rightrightarrows g$ при $n \to \infty$. Кроме того, если $g \geq 0$, то и $g_n \geq 0$.

Доказательство. Положите

$$g_n(\omega) := \frac{k}{2^n}, \quad \text{если} \quad \frac{k}{2^n} \leq g(\omega) \leq \frac{k+1}{2^n},$$

и пробегите все $k \geq 0$. □

Рассмотрим произвольную последовательность $(f_n)_{n\geq 1}$ неотрицательных простых измеримых функций, таких что $f_n \leq f_{n+1} \leq \cdots \leq f$ и $f_n \rightrightarrows f$ при $n \to \infty$. Тогда, согласно лемме B.20(1), последовательность $\int f_n\, d\mu$ не убывает, а значит сходится к пределу, который мы и назовем интегралом функции f:

$$\int_\Omega f\, d\mu := \lim_{n\to\infty} \int_\Omega f_n\, d\mu. \tag{B.7}$$

Конечно, чтобы такое определение интеграла было корректным, необходима следующая

Теорема B.22. *Предел (B.7) не зависит от выбора последовательности f_n.*

[34] В литературе простыми часто называют функции, принимающие лишь конечное число значений.

Доказательство этого результата давалось в курсе теории меры. См. также [KS, Теорема 3.4].

Шаг 3. Остается избавиться от условия неотрицательности функции f. Для произвольной измеримой функции f, функции $f_+ := f\mathbb{I}_{f>0}$ и $f_- := -f\mathbb{I}_{f<0}$ — неотрицательны и измеримы, и
$$f = f_+ - f_-.$$
Если интегралы $\int f_+\, d\mu$ и $\int f_-\, d\mu$ конечны, то положим
$$\int_\Omega f\, d\mu = \int_\Omega f_+\, d\mu - \int_\Omega f_-\, d\mu.$$
В этом случае говорят, что функция f интегрируема и пишут $f \in L_1(\Omega, \mu)$.

B.6.2 Свойства интеграла Лебега

Если же функция f не интегрируема, причем интеграл $\int f_+\, d\mu = \infty$, а $\int f_-\, d\mu < \infty$, то говорят, что $\int f\, d\mu = \infty$. Если, наоборот, $\int f_+\, d\mu < \infty$, а $\int f_-\, d\mu = \infty$, то $\int f\, d\mu = -\infty$. Наконец, если оба интеграла бесконечны, то интеграл $\int f\, d\mu$ не определен.

Переходом к пределу легко проверяется следующий результат:

Лемма B.23. *Утверждения леммы B.20 верны для произвольных измеримых функций f_1, f_2. При этом в пункте (2) достаточно, чтобы конечен был хотя бы один интеграл из правой части.*

Так как $|f| = f_+ + f_-$, конечность интеграла $\int f\, d\mu$ эквивалентна конечности интеграла $\int |f|\, d\mu$, то есть *интегрируемость и абсолютная интегрируемость для интеграла Лебега эквивалентны.* Действительно, если интегралы $\int f_\pm\, d\mu$ конечны, то, согласно лемме B.23, $\int (f_+ + f_-)\, d\mu$ тоже конечен. Обратно, если $\int |f|\, d\mu$ конечен, то, так как $-|f| \le f \le |f|$, опять согласно лемме B.23 интеграл $\int |f|\, d\mu$ конечен.

Пусть $A \in \mathcal{F}$. Как определить интеграл $\int_A f\, d\mu$? На данный момент мы это сделали только для $A = \Omega$. Первый способ — заменить множество Ω множеством A, и повторить построение выше. Другой способ такой:
$$\int_A f\, d\mu := \int_\Omega f\mathbb{I}_A\, d\mu.$$
Легко увидеть, что оба способа дают один и тот же результат. Отметим важное свойство интеграла Лебега — счетную аддитивность. А именно, если $A = \cup_{i=1}^\infty A_i$, $A_i \in \mathcal{F}$ — не пересекающиеся множества, то
$$\int_A f\, d\mu = \sum_{i=1}^\infty \int_{A_i} f\, d\mu$$
для любой измеримой функции f, для которой конечен интеграл из левой части.

B.6.3 Замена переменной в интеграле Лебега

Пусть имеется второе измеримое пространство $(X, \mathcal{G})$ и измеримая функция $h : (\Omega, \mathcal{F}) \mapsto (X, \mathcal{G})$. Напомним, что на $(X, \mathcal{G})$ определена мера $h_*\mu$, перенесенная с помощью отображения h, см. (B.2).

Теорема B.24. *Для любой измеримой функции $f: X \mapsto \mathbb{R}$,*

$$\int\limits_{\Omega} f(h(\omega))\, d\mu(\omega) = \int\limits_{X} f(x)\, df_*\mu(x).$$

Интеграл в правой части определен в том и только том случае, когда определен интеграл в левой части.

Доказательство. Не ограничивая общности будем считать, что функция $f \geq 0$. Если f— простая, теорема следует из определения перенесенной меры $f_*\mu$. В случае произвольной измеримой f теорема следует из того, что f можно приблизить неубывающей последовательностью простых функций, как это делалось выше. □

Пример B.25. Пусть $\Omega = \{\omega_1, \omega_2, \dots\}$ — не более, чем счетное множество, $\mathcal{F} = 2^{\Omega}$ и $\mu_i := \mu(\{\omega_i\})$. Тогда любая функция $f : \Omega \mapsto \mathbb{R}$ является простой, и, если она интегрируема, то

$$\int_{\Omega} f(\omega)\, d\mu(\omega) = \sum_j \mu_j f(\omega_j).$$

Пусть $\{a_1, a_2, \dots\}$ —множество значений функции f, а

$$\nu_i := \mu\{\omega : f(\omega) = a_i\} = f_*\mu(\{a_i\}).$$

Тогда, согласно теореме B.24, (в данном случае это элементарное вычисление, проведите его!)

$$\int_{\Omega} f(\omega)\, d\mu(\omega) = \sum_i a_i \nu_i.$$

B.6.4 Нужно прокинуть предел внутрь матожидания? Вам сюда

Стандартные средства для этой цели — теорема Лебега о мажорируемой сходимости, лемма Фату и теорема Б. Леви о монотонной сходимости. См, например, [KS, §3.5] и [KF, §5.5.5]. О существовании этих инструментов должен иметь представление любой специалист. Мы же сформулируем только теорему Лебега о мажорируемой сходимости, поскольку пока только ей и пользуемся.

Пусть $(\Omega, \mathcal{F})$ — измеримое пространство с мерой μ.

Теорема B.26. *Пусть последовательность измеримых функций $f_n \to f$ п.в. для измеримой функции f и $|f_n| \leq \varphi$, где функция φ интегрируема, т.е. $\int_{\Omega} \varphi\, d\mu < \infty$. Тогда*

$$\lim_{n\to\infty} \int_{\Omega} f_n\, d\mu \to \int_{\Omega} f\, d\mu.$$

B.6.5 Пункт обмена интегралов местами

Для этой цели используется теорема Фубини. Ниже мы формулируем ее немного упрощенную версию, единственно которой мы и пользуемся в нашем курсе.

Пусть μ_1 и μ_2 — меры на измеримых пространствах $(\Omega_1, \mathcal{F}_1)$ и $(\Omega_2, \mathcal{F}_2)$ соответственно. Обозначим через μ прямое произведение мер μ_1 и μ_2 — это мера на прямом произведении пространств $(\Omega_1, \mathcal{F}_1)$ и $(\Omega_2, \mathcal{F}_2)$.

Теорема B.27 (Фубини)**.** *(1) Пусть функция $f : \Omega_1 \times \Omega_2 \mapsto \mathbb{R}$ интегрируема по мере μ на множестве $A_1 \times A_2$, где $A_i \in \mathcal{F}_i$. Тогда*

$$\int_{A_1 \times A_2} f(x,y)\,d\mu(x,y) = \int_{A_1} \Big[\int_{A_2} f(x,y) d\mu_2(y) \Big] d\mu_1(x) = \int_{A_2} \Big[\int_{A_1} f(x,y) d\mu_1(x) \Big] d\mu_2(y),$$

причем все три интеграла существуют.

(2) Достаточное условие: если хотя бы один из интегралов

$$\int_{A_1} \Big[\int_{A_2} |f(x,y)| d\mu_2(y) \Big] d\mu_1(x) \quad \text{или} \quad \int_{A_2} \Big[\int_{A_1} |f(x,y)| d\mu_1(x) \Big] d\mu_2(y) \tag{B.8}$$

существует, то f интегрируема по мере μ на $A_1 \times A_2$, так что выполнено утверждение пункта (1).

См. [KF, теорема 5.6.5 и замечание после ее доказательства]. В частности, по этой ссылке можно найти пример, в котором каждый из повторных интегралов пункта (1) существует, но условие пункта (2) не выполнено, а функция f по мере μ не интегрируема.

Чтобы применить теорему Фубини на практике, как правило проверяют выполнение достаточного условия (2).

C Список учебников

По теории вероятностей имеется масса книг. Вот некоторые из них. Основная часть списка с комментариями ниже взята из конспекта лекций В.И. Богачева 2017 года.

Я бы рекомендовал (но это мои личные предпочтения, они у всех разные!) брать книгу Кораллова-Синая, и вместе с ней листать книгу Ширяева, иначе может быть тяжеловато от суховатости изложения. Можно обойтись и одним Ширяевым, но там уж очень много всего. Если хочется еще больше примеров, можно смотреть Феллера и Тутубалина. Все или почти все книги легко скачиваются.

1. Кораллов Л.Б., Синай Я.Г. Теория вероятностей и случайные процессы. МЦНМО, М., 2013 — первая половина книги содержит компактный, четкий и интенсивный курс теории вероятностей (более объемный, чем семестровый). Наши лекции в некоторой степени следуют ей.

2. Ламперти Дж. Вероятность. Наука, М., 1973 — хорошо написанное краткое пособие для математиков.

3. 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,

4. 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. Книга всем хороша, кроме того, что там очень много всего, и от этого может создаться некоторое ощущение хаотичности.

5. 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. В книге рассматривается масса примеров.

6. 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.

7. 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:

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

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

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

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

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

Список литературы

[KF] А.Н. Колмогоров, С.В. Фомин, *Элементы теории функций и функционального анализа*, Наука 1976.

[KS] Л.Б. Коралов, Я.Г. Синай, *Теория вероятностей и случайные процессы*, МЦНМО 2013.

[Shi-1] А. Н. Ширяев, *Вероятность - 1*, МЦНМО 2011.

[Shi-2] А. Н. Ширяев, *Вероятность - 2*, МЦНМО 2011.

[Dud] R.M. Dudley, *Real analysis and probability*, CUP 2004.