Напоминание: комбинаторика и простейшие формулы

◦ Простейшие базовые формулы

- **Формула включений и исключений**. Пусть $|A|$ обозначает число элементов в конечном множестве A. Тогда для (конечного) семейства конечных множеств A_i, $i = 1, 2, \dots n$ 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 [1] как $\binom{n}{k}$.

◦ 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, ..., 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 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:

$$\frac{(n_1 + n_2 + \ldots + n_k)!}{n_1!\cdot n_2!\cdot\ldots\cdot n_k!}$$

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

[1] кроме того для любого действительного x полагают $\binom{x}{k} = \frac{x(x-1)(x-2)\ldots(x-k+1)}{k!}$

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 [2]. Такая картинка см. Рис.1 из клетчатых полосок называется диаграммой Юнга. Нас интересует подсчет всех таких табличек с общим числом клеток n и числом строк не более k.

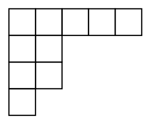

Рис. 1: Пример диаграммы Юнга

Первым делом заметим, что каждую такую табличку можно *транспонировать* – строки по порядку выписать столбцами – тогда опять получится диаграмма Юнга 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 .

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 .

○ Методы вычислений

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.

$$\begin{array}{ccccccccc}
 & & & & 1 & & & & \\
 & & & 1 & & 1 & & & \\
 & & 1 & & 2 & & 1 & & \\
 & 1 & & 3 & & 3 & & 1 & \\
1 & & 4 & & 6 & & 4 & & 1 \\
. & . & . & . & . & . & . & . & .
\end{array}$$

[2]французские математики предпочитают рисовать их по возрастанию

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 $k!\left\{\begin{matrix} n \\ k \end{matrix}\right\}=|Y|$. 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

$$\left\{ {n \atop k} \right\} = |Y| = \frac{1}{k!}\sum_{j=0}^{k}(-1)^j\binom{k}{j}(k-j)^n$$

□ 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 абсцисс, 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 $\int \ln x\ dx = x\ln x - x$,

$$\int\limits_{[\frac{3}{2}, n]}\ln x\ dx < I_n < \int\limits_{[1, n]}\ln x\ dx$$

$$\left(n + \frac{1}{2}\right)\ln n - n + \frac{3}{2}\left(1 - \ln\frac{3}{2}\right) < \ln n! < \left(n + \frac{1}{2}\right)\ln n - n + 1$$

$$n^{\left(n+\frac{1}{2}\right)}e^{-n}\cdot\beta < n! < n^{\left(n+\frac{1}{2}\right)}e^{-n}\cdot e$$

$$\beta < \frac{n!}{n^{(n+\frac{1}{2})}e^{-n}} < e$$

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 $\binom{2n}{n} \sim \frac{2^{2n}}{\sqrt{\pi n}}$

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. Кроме того, в Теорию Вероятностей часто включают фрагменты Математической Статистики. При этом 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"). При начальном изучении такая 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.

Поэтому представляется разумным 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

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 даже 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. При этом необходимо придать и строгий смысл другому общеупотребительному термину – **независимости** измерений. 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, 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 котором вероятность понимается как неотрицательная вещественнозначная мера $0 \leqslant P \leqslant 1$ на подходящем множестве, а независимость определяется как соотношение на возникающие числовые значения.

- фриквентистский (частотный) подход Рихарда фон Мизеса, который основательно (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 $0 \leqslant P \leqslant 1$: к повторениям эксперимента или к единичному эксперименту.

- эвристический и свободный от какой-либо традиционно понимаемой аксиоматики байесианский подход[3], который однако же в конкретных задачах математической статистики иногда быстро приводит к ответам.

[3] 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» считаются осмысленными, так как отражают частную силу веры в истинность предложения.

- 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.

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

Для общего представления можно сказать, что колмогоровская теория вероятностей содержательно состоит из математических утверждений теории меры P применительно к аксиоматически введенному понятию независимости.

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 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 на как минимум два разных подхода: см. левую и правую картинку на Рис.2: Интуитивный подход к оценке «вероятности того, что монет поровну»

Рис. 2: Два способа подсчета шансов в эксперименте «монеты в карманах»

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. Важный вывод: *численные оценки вероятности имеют смысл только в привязке к конкретно выбранной модели*, а само по себе утверждение о численном значении вероятности смысла не имеет. Таким образом необходима правило создания вероятностных моделей. 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).

Интерпретации теории

Исторически так уж сложилось, что утверждения теории вероятностей широко используется в приложениях. Это означает, что при практических 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 *в рамках выбранной модели можно* *ставить количественный вопрос о вероятности соответствия модели и реальности*: это, вообще говоря, требует дополнительных экспериментов (измерений), а сами 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 постепенно дополняя соответствующие определения 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. В начале изложения мы коснемся кратко методов элементарной теории вероятностей, которая во многом состоит из утверждений комбинаторики, взятых в сочетании с понятием независимости.

2 Основные понятия и их связь с интуицией

2.1 ◦ Классические модели в теории вероятностей

Классические модели возникли на основе (необсуждаемой) веры, что любой эксперимент можно описать в терминах **равнозначных** альтернатив $\omega_1, \omega_2, \ldots \omega_N$, называемых *элементарными исходами данного опыта* все элементарные исходы составляют множество Ω. В этой схеме *событиями* A, B итд называют подмножества в Ω, события, состоящие из одного исхода называют элементарными, а Ω — пространством элементарных событий. Наиболее уязвимым местом здесь является именно симметрия всех альтернатив, непонятно, насколько идея равнозначности альтернатив применима к реальности. Впрочем, имеется набор примеров пространств элементарных событий, по поводу которых в обществе давно сложилось согласие, в частности много таких примеров в §1,§2 Главы 1, том 1 книги В.Феллер "Введение в теорию вероятностей и ее приложения", этот материал настоятельно рекомендуется для самостоятельного ознакомления, хотя бы поверхностного.

Замечание о формулах комбинаторики

Во-первых, следует с самого начала понимать, что не для всех вообще ситуаций бывает короткая формула, позволяющая быстро вычислять количество всех мыслимых вариантов. Далее, если такая формула и существует, то ее обоснование может быть достаточно трудоемко. Вывод здесь такой, что необходимо знать несколько базовых принципов комбинаторики и соответствующих формул, а также четко понимать, как именно формулы могут быть применены.

Обзор этих базовых случаев содержится в разделе , предполагается, что он понадобится на упражнениях.

Классическая схема подсчета вероятностей

Определение Пусть Ω — некоторое множество. Система $\mathcal{A}$ подмножеств в Ω называется алгеброй, если

$\Omega \in \mathcal{A}$

$A \in \mathcal{A} \Longrightarrow \overline{A} \in \mathcal{A}$

$A, B \in \mathcal{A} \Longrightarrow A \bigcup B \in \mathcal{A}, A \bigcap B \in \mathcal{A}$

В классической схеме принято рассматривать все подмножества в Ω, они образуют *алгебру событий*, в которой очень часто для операций $\cup$ и $\cap$ приняты[4] обозначения $+$ и $\cdot$ и соответственно термины *сумма и произведение*, дополнение $\overline{A}$ называется *противоположным* событием, а пустое множество – *невозможным.* Два непересекающихся подмножества называются *несовместными* событиями, набор несовместных событий в сумме дающих все Ω называется *полной группой событий* или *разбиением.* Иногда употребляют известные для множеств операции $A \setminus B$ и $A \triangle B$ между событиями A, B но не вводя никаких специальных терминов при этом.

Между операциями имеется много соотношений, вытекающих из соотношений для операций с подмножествами. Для дальнейшего нам важно упомянуть два из них:

$$A \cdot \left(\bigoplus_i B_i \right) = \bigoplus_i (A \cdot B_i) \qquad \text{и} \qquad A + \prod_i B_i = \prod_i (A + B_i)$$

Все события в классической схеме *составляют булеву алгебру* [5]. Описание различных конечных множеств Ω (называемых чаще пространствами элементарных событий) связано в классическом случае в основном с формулами комбинаторики.

Схема подсчета вероятностей

Определение Пусть $\mathcal{A}$ — алгебра подмножеств Ω. Неотрицательная функция множеств $0 \leqslant \mu(A) \leqslant \infty$ называется конечно-аддитивной мерой, заданной на $\mathcal{A}$, если для любых двух непересекающихся множеств A и B из $\mathcal{A}$ $\mu(A) + \mu(B) = \mu(A \bigcup B)$. При $\mu(\Omega) = 1$ она называется *вероятностной мерой* и обозначается чаще как P.

[4] С обозначением $+$ у разных авторов следует быть предельно аккуратным, поскольку некоторые используют его лишь для непересекающихся между собой подмножеств, а стало быть некоторые соотношения могут иметь скрытый ограничительный смысл

[5] это означает наличие тождеств теории множеств для указанных операций

Абстрагируясь, скажем, что *традиционное* построение вероятностной модели — она называется вероятностным пространством — состоит в задании тройки $(\Omega, \mathcal{A}, P)$ объектов: множества элементарных исходов Ω, алгебры событий-подмножеств $\mathcal{A}$, вероятностной меры на алгебре $\mathcal{A}$. Выделение на слове *традиционное* важно: с развитием науки выяснилось, что интуитивно понятная в конечном случае схема нуждается в уточнениях[6], далее мы еще вернемся к вопросу о разных схемах и тем самым разных теориях вероятностей.

В классической схеме из N исходов событию A приписывается вероятность $0 \leqslant P(A) = m/N \leqslant 1$, где $0 \leqslant m = |A|$ — число элементарных исходов, составивших событие A. Нечеткость описания эксперимента порождает так называемые «парадоксы теории вероятностей», когда численные оценки вероятностей для одного и того же словесного описания в разных моделях различаются. Вероятностные описания создают определенные сложности и в интерпретации таких понятий, как *типичное, лучше-хуже* и т п.

Пример

Рассмотрим кубики, на гранях которых следующие числа: $[4,4,4,4,0,0]$, $[3,3,3,3,3,3]$, $[2,2,2,2,6,6]$, $[1,1,1,5,5,5]$.

Заметим, что первый кубик при парном бросании со вторым выигрывает в большинстве случаев, то есть он «лучше» второго, второй «лучше» третьего, третий «лучше» четвертого, а четвертый – «лучше» первого! Заметим, что в термине «лучше» использованы лишь результаты модели для пар кубиков, которые в общем не определяют модель для всех четырех кубиков.

Замечание о неклассических схемах

Неудобство классической схемы заключается в том, что равновероятные элементарные события непонятно как определить в явно лишенных симметрии ситуациях типа бросаний неровной игральной кости, кривой монеты и т. п. Другим ограничением приложений классической схемы послужило то, что на практике было необходимо разбираться с измерениями числовых (и векторных) величин, то есть когда конечность классической модели оказывается слишком ограничительной. Это породило ряд более сложных подходов к понятию пространства элементарных событий и схеме задания вероятностей событий. Здесь встречаются принципиальные трудности, которые разрешимы лишь в рамках аксиоматического подхода к теории вероятностей. Об этом речь пойдет далее, а пока ограничимся общеинтуитивными пояснениями. Чтобы как-то определиться с терминологией будем ***в этом курсе*** называть рассматриваемые вероятностные модели «традиционными», чтобы отличать их от «классических», то есть таких, где вероятностное пространство конечно и элементарные события равновероятны. Вот две интуитивно ясные модификации классической модели.

Дискретное вероятностное пространство Здесь используются и несимметричность и (возможная) бесконечность.

1. Исходы в вероятностном пространстве отныне не обязательно равновероятны: вероятностное пространство Ω состоит из исходов $\omega \in \Omega$, у каждого определена вероятность $p(\omega) > 0$ так, что сумма всех вероятностей равна единице. При этом классическое определение вероятности события «отношение числа исходов в событии к числу всех исходов» более не работает, а вероятность события $A \subset \Omega$ вычисляют по формуле
$$P(A) = \sum_{\omega \in A} p(\omega)$$
2. Исходы индексированы конечным или бесконечным подмножеством натуральных чисел $\mathbb{N}$, сумма в последнем случае берется как сумма сходящегося ряда с положительными членами.

Типичный пример применения – построение вероятностной модели для эксперимента с подбрасыванием монеты до первого (или k-го) выпадения герба. Такой эксперимент кодируется последовательностью некоторой длины, которая заканчивается выпадением герба, поскольку длина последовательности может быть любой — Ω уже не конечное множество. Сразу ясно, что в этом случае правило вычисления вероятностей будет иным, не сводящимся к простому подсчету количества исходов в событии. В традиционном дискретном подходе предполагается необходимым изначально задать вероятности p_i всех элементарных событий так, чтобы ряд $\sum_{i=1}^{\infty} p_i$ сходился бы к единице. Поскольку любое событие A состоит из элементарных вида $\{\omega_{i_k}\}$, то его вероятность определяется суммой $\sum_k p_{i_k}$.

[6] Например, в аксиоматическом подходе важно, что алгебра $\mathcal{A}$ является вдобавок и σ-алгеброй и для любых непересекающихся $A_i \cap A_j = \varnothing$ подмножеств в $A_1, A_2, \ldots \in \mathcal{A}$ выполнено $\sum_i^{\infty} \mu(A_i) = \mu\left(\bigcup_i A_i\right)$, такая мера называется σ-аддитивной.

Геометрические вероятности Другой вариацией классического подхода является использование в качестве вероятностного пространства Ω геометрического объекта и тех событий, которые также являются геометрическими объектами, в схеме вычислений вероятностей использовать характеристики типа длины-площади-объема. Это пример так называемых *геометрических вероятностей* , которые заодно являются и *недискретными.* Но здесь возникает необходимость рассматривать события, состоящие из тех непустых подмножеств в Ω, у которых вероятность нулевая: например событие, состоящее из единственной точки. С другой стороны, при таком подходе на интуитивном уровне легко находить численные ответы к вопросам вроде:

- Найти вероятность, что случайно выбранное направление составляет с вертикалью угол больший $60°$?

2.2 ○ Зависимость и независимость событий. Связь с интуицией

Вторым важнейшим для теории вероятностей понятием служит **независимость**, оно определено для событий — элементов семейства $\mathcal{A}$. Конкретно, для конечного набора индексов $I = \{i_1, i_2, \dots i_n\}$ *независимость в совокупности* событий $\{A_i\}_{\{i \in I\}}$ означает выполнение тождеств $P(A_{j_1})P(A_{j_2}) \dots P(A_{j_k}) = P(A_{j_1} \cap A_{j_2} \cap \dots A_{j_k})$ для любого подмножества $J \subset I$ индексов $j \in J$. Это определение, на первый взгляд, выглядит загадочно, поэтому сначала рассмотрим его на относительно простом примере двух событий A_1 A_2 конечного вероятностного пространства Ω. С чего бы двум событиям лежать в Ω так, чтобы вероятность их общей части равнялась бы произведению? Следующие примеры показывают, что интуиция здесь казалось бы не при чем: сравните ответы в следующих задачах про эксперименты с игральными кубиками (с цифрами на гранях от 1 до 6) и тетраэдрами (с цифрами на гранях от 1 до 4, показания выбираются по нижней грани.).

Примеры

Событие A – четное число очков, событие B – число очков кратно трем. Вопрос один и тот же для всех вариантов: зависимы ли A и B? Варианты условий:

- Очки считаются при бросании одного кубика.
- Очки считаются при бросании одного тетраэдра.
- Очки считаются в сумме при бросании двух кубиков.
- Очки считаются в сумме при бросании двух тетраэдров.
- Очки считаются в сумме при бросании кубика и тетраэдра.

С другой стороны: рассмотрим теперь опыты *по здравому смыслу независимые*: например, в одной комнате бросают игральную кость, а в другой — тетраэдр и рассматриваются события, относящиеся именно к разным комнатам. Классические математические модели для опыта в каждой комнате (Ω_1, P_{Ω_1}) и (Ω_2, P_{Ω_2}) хорошо известны, а составной опыт, как и в примерах выше, естественно описать классическим вероятностным пространством $\Omega_1 \times \Omega_2$, в этом случае вероятности элементарных исходов (ω_i, ω_j) все одинаковые и, как несложно видеть, равны $\frac{1}{24}$ то есть произведению вероятностей для ω_i и ω_j, вычисленных соответственно в (Ω_1, P_{Ω_1}) и (Ω_2, P_{Ω_2}). Вторая мысль: модель для происходящего в первой (или второй) комнате можно вычислять также и в пространстве $\Omega_1 \times \Omega_2$, обращая внимание только на первый или только на второй элемент пары (ω_i, ω_j), что и покажет формально понятую независимость. Третье соображение, что рассматривать произведение вероятностных пространств и произведение мер в принципе возможно и в более общем, чем классический подход случае: интуитивный смысл независимости в таких моделях будет в согласии с формальным свойством. Вывод тут такой, что для интуитивно понимаемой независимости экспериментов нужно создать вероятностную модель с формально проверяемой независимостью.

- Упражнение: следует ли из попарной независимости событий A, B, C их независимость в совокупности? Приведите конкретный пример вероятностного пространства и трех событий, обосновывающих ответ.

Формальное определение условной вероятности

События в вероятностном пространстве выделяются условиями. При рассмотрении нескольких событий возникает также понятие *условной вероятности*: для события ненулевой вероятности B *условной вероятностью события A при условии события B* называется число $P(A|B) = \frac{P(AB)}{P(B)}$.

2.3 ○ Случайные величины. Связь с интуицией

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.

Ясно, что 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 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 (например, первая формулировка в случае геометрических вероятностей) будут уточнены. Пока ограничим себя представлением связи с.в. и вероятностного пространства картинкой Рис.4

Понятие случайного вектора вводится совершенно аналогично, если заменить элементы-числа на элементы-векторы фиксированной размерности. Наконец, возможен выбор комплекснозначных случайных величин и случайных векторов, если не оговорено специально, то подразумеваются вещественнозначные случайные величины.

Распределения дискретных случайных величин и векторов

Имеются два основных способа представления структуры вероятностей, связанных со случайной величиной ξ:

1. В дискретном случае указывают таблицу значений и их вероятностей, например:

ξ	0	$\dots$	k	$\dots$	n
	$\binom{n}{0}q^n$	$\dots$	$\binom{n}{k}p^k q^{n-k}$	$\dots$	$\binom{n}{n}p^n$

   Здесь использованы параметры $0 \leqslant p, q \leqslant 1$, $p + q = 1$, $n > 0$, а соответствующая с.в. известна под названием **биномиальной или бернуллиевской**. Такая таблица называется **распределением** случайной величины.

   - Упражнение: для биномиальной случайной величины ξ с $n = 3$, $p = \frac{1}{2}$ заполните таблицу распределения случайной величины $\left(\xi - \frac{3}{2}\right)^2$.

2. Семейство вероятностей событий $P(\xi \leqslant x)$ можно представить как функцию аргумента x, принимающую значения на отрезке $[0, 1]$ и вдобавок монотонно неубывающую (потому что при $x_1 < x_2$ $P(\xi \leqslant x_1) \leqslant P(\xi \leqslant x_2)$) — эта функция называется **функцией распределения случайной величины** ξ и обозначается $F_\xi(x)$.

   - Упражнение: для биномиальной случайной величины с $n = 3$, $p = \frac{1}{3}$ нарисуйте график ее функции распределения.

Случайный вектор, очевидно, определяет пару случайных величин-координат и в дискретном случае его можно описать соответствующей таблицей распределения

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

С помощью этой таблицы можно восстановить таблицы распределения с.в. координат, а также некоторых других, с ними связанных. Вот список обязательных упражнений для понимания соответствующей техники.

- Найти вероятность $P(\alpha = \beta)$.
- Найти вероятность $P(\alpha \leqslant \beta)$.
- Найти распределение случайного вектора с компонентами $\alpha + \beta$ и $\alpha - \beta$.

Чтобы понять смысл функции распределения $F_{\alpha,\beta}(x, y)$ двумерного случайного вектора ξ, η достаточно рассмотреть семейство событий $\{\alpha \leqslant x\} \cap \{\beta \leqslant y\}/$

- Упражнение: для приведенного выше случайного вектора (α, β) указать все точки непрерывности функции $F_{\alpha,\beta}(x, y) \to \mathbb{R}$.

Связи вероятностной теории с естествознанием

Вопрос о том, какие вероятностные модели следует использовать для описаний природных процессов чрезвычайно важен и далеко не прост. Например, в классической статистической физике, созданной Максвеллом и Больцманом, объекты считаются различимыми друг от друга. Так себя ведут, например, молекулы газа. 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.

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 выше пример с монетами и карманами.

□ Содержательный пример: симметричное дискретное случайное блуждание

Важная для приложений модель (одномерного) 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 и имеют вероятность 2^{-t}, так что эта модель классическая. Такая траектория 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 $k + l = t$.

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

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

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.3 часть пути до его *первого* попадания на абциссу, это задает требуемую биекцию.

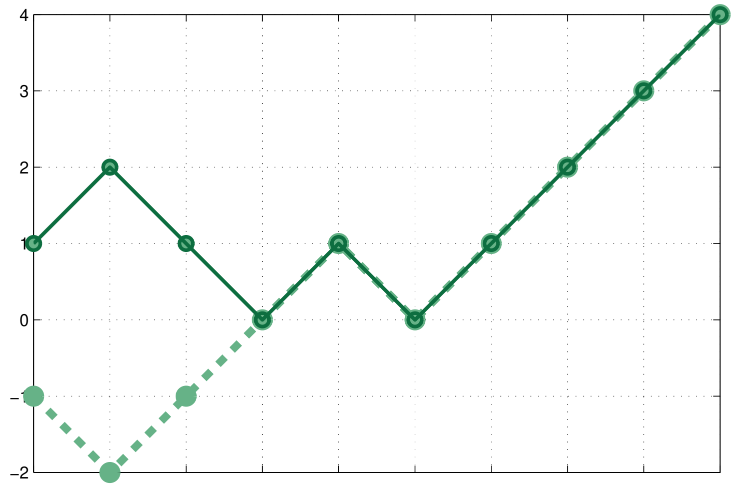

Рис. 3: Биективное соответствие путей, хотя бы раз принимающих значение 0, проходящих из точки с координатами $(t_1, 1)$ в точку $(t_2, 4)$ и всех вообще путей из точки $(t_1, -1)$ в точку $(t_2, 4)$

Таким образом, искомых «строго положительных» путей будет

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

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 $\frac{1}{n} C_{2n-2}^{n-1}$. Подсказка: $\frac{1}{2n-1} C_{2n-1}^{n-1} = \frac{1}{2n-1} C_{2n-1}^{n} = \frac{1}{n} C_{2n-2}^{n-1}$.
- количество путей из начала координат в точку $2n, 0$ *с неотрицательными внутренними вершинами* будет $\frac{1}{n+1} C_{2n}^{n}$. 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 веришинами, которых по предыдущей задаче в точности $\frac{1}{n} C_{2n-2}^{n-1}$, осталось поменять обозначения.

□ Возможные применения методов

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. Указанное интуитивное соображение указывает, 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.

Частичное объяснение этого феномена (он называется Законом арксинуса) перенесено в упражнения, мы не будем здесь приводить аналитического доказательства[7].

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

3 Случайные величины и независимость

3.1 ∘ Дискретные случайные величины

На каждом элементарном исходе ω_k с.в. отвечает конкретное числовое значение $\xi(\omega_k)$, для значения $c \in \mathbb{R}$ возникают события вида $A_c = \{\omega | \xi(\omega) = c\}$. Случайная величина, принимающая всего два значения (как правило, эти значения выбираются как $\{0, 1\}$), называется битом или индикатором. Чаще всего ее обозначают $\mathbf{1}_A$, $A = \{\omega | \mathbf{1}_A(\omega) = 1\}$. Принимающая конечное или счетное число значений случайная величина ξ связана с индикаторами очевидной формулой:

$$\xi(\omega) = \sum_{\{c_k \in \xi(\Omega), k \in \mathbb{N}\}} \mathbf{1}_{\{\xi = c_k\}}$$

Такие случайные величины называются **дискретными**, в дальнейшем будут определены и с.в. не являющиеся дискретными.

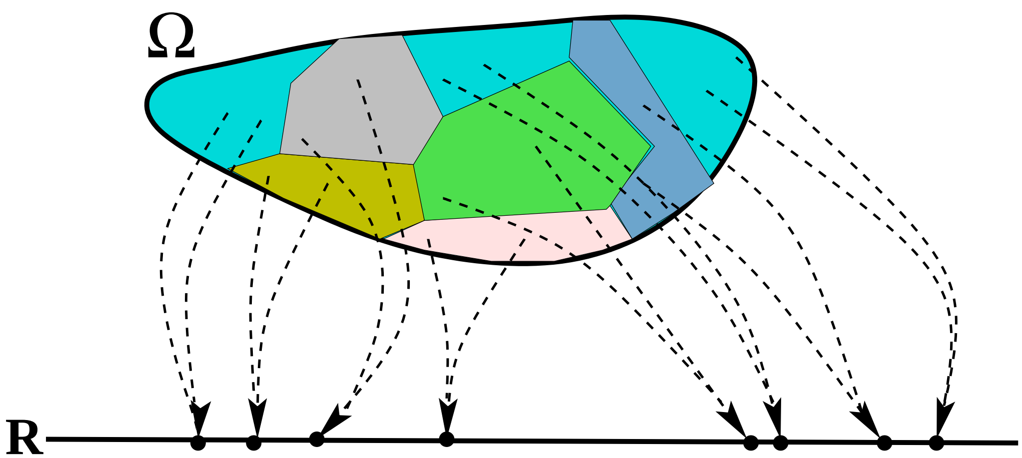

Рис. 4: Дискретная случайная величина $\xi : \Omega \to \mathbb{R}$ порождает полную группу событий вида $\{\xi = a\}$ в Ω

Совершенно аналогично вводится понятие векторнозначной дискретной случайной величины $\xi : \Omega \to Y = \mathbb{R}^n$, см. Рис. 5. Ясно, что композиция дискретной случайной величины ξ и (любой) функции $g : \mathbb{R} \to \mathbb{R}$ задает дискретную случайную величину $\eta = g(\xi)$, такие случайные велЕсли дичины называются функционально зависимыми.

- Верно ли, что любая дискретная с.в. $\eta = g(\xi)$, принимающая конечное множество значений, может быть представлена как композиция $\eta = g(\xi)$, где ξ определена на классическом вероятностном пространстве?

Независимость случайных величин и случайных векторов

Для подмножества $U \subset \mathbb{R}$ его прообраз $\xi^{-1}(U)$ является событием, вероятность которого можно еще обозначать как $P(\xi \in U)$. Сформулируем, что есть независимость двух случайных величин в терминах прообразов подмножеств $U \subset \mathbb{R}$: случайные величины $\beta_1 : \Omega \to \mathbb{R}$ $\beta_2 : \Omega \to \mathbb{R}$ называются попарно независимыми, если для любых подмножеств $U, V \subset \mathbb{R}$ независимы события $\beta_1^{-1}(U)$ и $\beta_2^{-1}(V)$. Аналогичное определение выбирают и для случайных векторов, рассматривая прообразы любых подмножеств $U, V \subset \mathbb{R}^m$.

Совокупная независимость нескольких случайных величин определяется далее очевидным образом: рассмотрением групповой независимости соответствующих проообразов. Независимость случайных величин в модели независимых повторений измерений и случайных величин — проекций на непересекающиеся семейства сомножителей (отвечающих отдельному измерению) заведомо выполнена: дело в том, что формула произведения для вероятностей прообразов положена в основу этой модели. Независимость случайных величин представляется на первый взгляд довольно сильным условием, к тому же возникает отдельное требование, что области определения рассматриваемых случайных величин должны совпадать. С этим вторым требованием можно легко справиться, полагая, что во всех случаях одновременного рассмотрения нескольких случайных величин $\beta_1, \beta_2, \ldots \beta_m$ необходимо принять, что задан случайный вектор $\vec{\beta} : \Omega \to \mathbb{R}^m$, компонентами которого указанные случайные величины и являются.

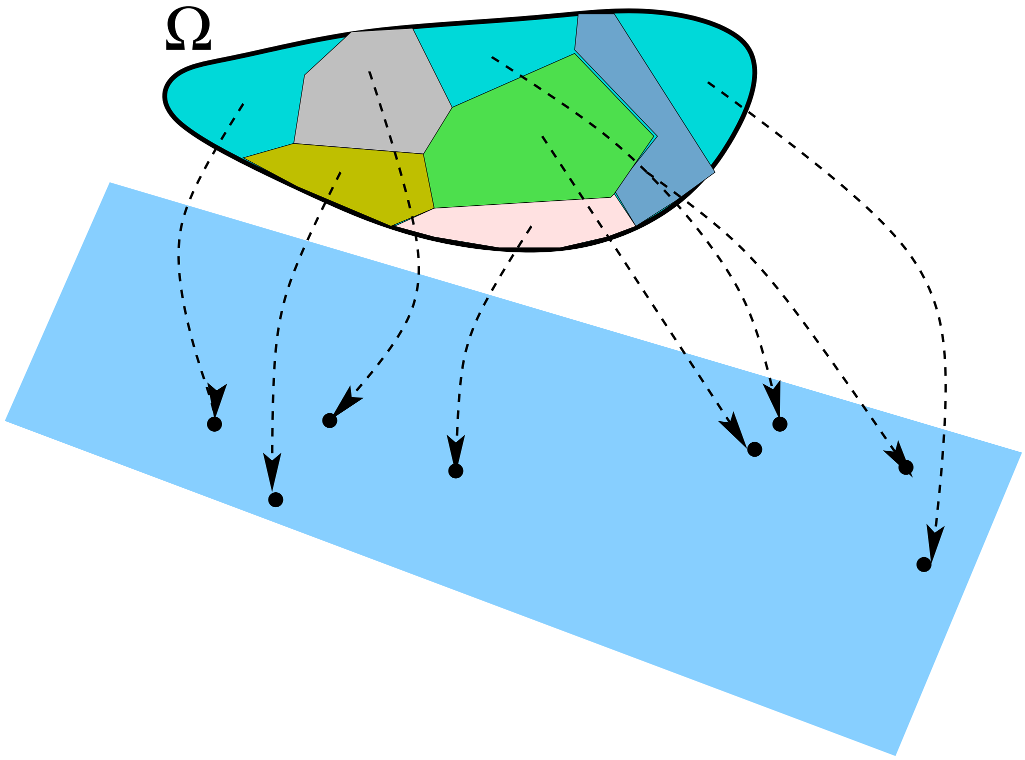

Рис. 5: Двумерный случайный вектор $\xi : \Omega \to \mathbb{R}^2$

Последовательность независимых повторений измерений понятным образом представляет большой практический интерес. Стоит самостоятельно продумать детали конструкции: в модели при независимом повторении одинаковых измеарений проекции (измерения при повторениях $i, j,\ i \neq j$) независимы (восстановите все формальные детали и проверьте). В англоязычной литературе для таких случайных величин имеется специальная аббревиатура i.i.d., что означает independent and identically distributed. Впредь мы будем употреблять это сокращение.

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 и, значит, график *непостоянной* функции g на плоскости $\mathbb{R}^2$ позволяет выбрать множества $U, V \subset \mathbb{R}$ такие, что проекции графика на оси с этими множествами пересекаются, но сам график не пересекается с $U \times V \subset \mathbb{R}^2$. Проверим теперь связь функциональной зависимости и формального определения для случайных величин с помощью выбраных $U, V \subset \mathbb{R}$. По крайней мере в классическом случае очевидно, что $P(\beta_1 \in U) \neq 0$ и $P(\beta_2 \in V) \neq 0$ в то время как $(\beta_1 \in U) \cap (\beta_2 \in V) = \emptyset$. Отсюда функциональная зависимость в привычном смысле обеспечивает формально понятую *НЕ независимость* в смысле теории вероятностей. Однако НЕ независимость (то есть зависимость) в смысле теории вероятностей определение отвечает более широкому смыслу, нежели функциональная зависимость.

Векторные случайные величины и независимость

Набор вероятностей $p_{a_1,a_2,\dots a_n}$ отвечающий всем возможным различным значениям $\{a_1, a_2, \dots a_n\}$ векторнозначной случайной величины $\vec{\xi} = \{\xi_1, \xi_2 \dots \xi_n\}$ называется распределением (иногда – *совместным распределением компонент* вектора). В частности, набор вероятностей для случайной величины (= одномерной векторной случайной величины) – также распределение. Из совместного распределения $\vec{\xi} = \{\xi_1, \xi_2 \dots \xi_n\}$ несложно найти распределения компонент (они часто называются *маргинальными распределениями*), легче всего проследить это на примере двумерной случайной величины, распределение которой может быть записано матрицей.

- Упражнение: для приведенного в разделе 2.3 дискретного случайного вектора (α, β) зависимы ли его компоненты?

Если компоненты случайного вектора независимы в совокупности,то, в частности, для любых наборов значений компонент соответствующие события $\{\xi_k = a_k\}$ независимы, т.е $p_{a_1,a_2,\dots a_n} = p_{a_1} \cdot p_{a_2} \dots p_{a_n}$. Обратное также справедливо

- Как обосновать последнее утверждение?

Как обсуждалось выше, простая с.в. η, принимающая с ненулевыми вероятностями значения $y_1, y_2, \dots y_N$ связана со «своим» разбиением $\mathfrak{F}_\eta$, состоящим из событий $F_k = \{\omega | \eta(\omega) = y_k\}$ ненулевой вероятности[8].

Измеримость относительно разбиения

Случайная величина ξ называется *измеримой относительно разбиения*[9] $\mathfrak{D}$, если на каждом событии $D_k \in \mathfrak{D}$ разбиения она принимает постоянное значение x_k (для разных событий $D_i \in \mathfrak{D}$ эти значения могут и совпадать), то есть имеет место представление:

$$\xi(\omega) = \sum_k x_k \mathbf{1}_{D_k}(\omega)$$

это определение легко обобщается и на случайные векторы.

Ясно, что если с.в. ξ измерима относительно некоторого разбиения $\mathfrak{D}$, то она измерима и относительно его «измельчения» $\tilde{\mathfrak{D}}$ (измельчение разбиений записывается как $\tilde{\mathfrak{D}} \rhd \mathfrak{D}$).

4 Условные вероятности и байесовский подход

4.1 ◦ Условные вероятности

Понятие независимости играет центральную роль в теории вероятностей: именно это понятие определило то своеобразие, которое выделяет теорию вероятностей в общей теории измеримых пространств с мерой. Понятие условной вероятности было связано с идеей, что каждое подмножество $H \subset \Omega$ можно рассматривать как новую генеральную совокупность исходов, если только отнормировать исходные вероятности $p_\Omega(\omega)$ каждого исхода, разделив их на $P(H)$. Говоря о подмножестве мы желаем подчеркнуть, что все время помним о существовании более широкого пространства элементарных событий. Например, для страховой компании может представлять интерес вопрос о том, как часто молнии причиняют убыток определенной величины (событие A). Возможно, что эта компания обслуживает несколько различных категорий объектов, например индустриальные объекты и жилые объекты в городе и в деревне. Отдельно изучать убытки индустриальных объектов означает изучать событие A в соединении с событием H — «убыток причинен индустриальному объекту». Формула условной вероятности применима очевидным образом. Заметим, однако, что если H совпадает со всем пространством элементарных событий, и $P(A|H)$ сводится к $P(A)$.

Хотя символ $P(A|H)$ сам по себе и удобен, трудно дать его точное словесное выражение. Поэтому обычно пользуются менее формальными описаниями. Так, можно говоритьа о вероятности того, что женщина страдает дальтонизмом, вместо того чтобы говорить об «условной вероятности того, что случайно выбранное лицо страдает дальтонизмом при условии, что это лицо — женщина». Часто слова «при условии » 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 условным вероятностям говорят для ясности о безусловных вероятностях. Строго говоря, эпитет «безусловная» является лишним, и его можно опускать.

Все общие теоремы о вероятностях верны также и для условных вероятностей, если эти условные вероятности берутся при одном и том же условии H. В качестве примера приведем основное соотношение для вероятности появления либо события A, либо события B, либо обоих этих событий:

$$P(A \cup B|H) = P(A|H) + P(B|H) - P(AB|H)$$

Аналогичным образом можно перенести на условные вероятности и все теоремы, например, чтобы обобщить $P(AB) = P(A|B)P(B)$ ее на случай трех событий A, B, C примем сначала за условие $H = BC$, а затем еще раз применим формулу умножения; получим

$$P(ABC) = P(A|BC)P(B|C)P(C)$$

[8] в общем случае рассмотрения значений с вероятностями нуль возникает трудность, которую можно преодолеть, но эти детали мы здесь рассматривать не будем

[9] т.е. полной группы событий в Ω

4.2 △ Урновая схема

Урна содержит b черных и r красных шаров. Из урны наугад извлекается шар. Вынутый шар всегда возвращается обратно, причем в урну добавляют c шаров того же цвета и d шаров противоположного цвета. Затем из урны (содержащей теперь $r+b+c+d$ шаров) снова производится случайное извлечение и повторяется тот же процесс. Здесь c и d — произвольные целые числа, их можно взять и отрицательными, однако в этом случае процесс может окончиться после конечного числа извлечений из-за отсутствия шаров. В частности, положив $c=-1$ и $d=0$, мы приходим к модели случайного выбора без возвращения, который кончается через $r+b$ шагов. Чтобы придать этому описанию точный математический смысл, заметим, чго оно определяет условные вероятности, по которым могут быть вычислены некоторые основные вероятности. Точку пространства элементарных событий, соответствующего n извлечениям, можно представить как последовательность n букв К и Ч. Событие «первый шар черный» (т. е. множество всех последовательностей, начинающихся с Ч) имеет вероятность $\frac{b}{r+b}$. Если первый шар черный, то (условная) вероятность извлечь черный шар при втором испытании равна $\frac{b+c}{r+b+c+d}$. Безусловная вероятность последовательности ЧЧ равна, следовательно,

$$\frac{b}{r+b}\cdot\frac{b+c}{r+b+c+d}$$

Вероятность последовательности ЧЧЧ получается умножением на , на $\frac{b+2c}{r+b+2c+2d}$ и т. д. Ясно, что таким путем можно вычислить все вероятности. (Конечно, в случае отрицательных c и d число n извлечений надо выбирать достаточно малым, чтобы избежать появления отрицательного числа шаров.) Нетрудно проверить по индукции, что сумма вероятностей всех элементарных событий равна единице.

Точные выражения для вероятностей получить нелегко, исключая самый важный и лучше всего изученный частный случай, которым является

Урновая схема схема Пойа

. Эту схему характеризуют значения параметров $d=0$, $c>0$. После каждого испытания число шаров того же цвета, что и извлеченный шар, возрастает, в то время как число шаров противоположного цвета не меняется. Это приводит к эффекту последействия, состоящему в следующем: если мы извлекаем шар какого-то цвета, то вероятность извлечь шар того же цвета при следующем испытании возрастает. Урновая схема Пойа является приближенной моделью таких явлений, как, например, эпидемии, при которых осуществление некоторых событий увеличивает вероятность их повторения. Простота этой схемы связана со следующим очевидным обстоятельством: любая последовательность n. извлечений, приводящая к n_1 черным и n_2 красным шарам, имеет ту же вероятность, что и событие, состоящее в извлечении *сначала* n_1 черных и затем n_2 красных шаров, а именно эта вероятность равна

$$\frac{b(b+c)(b+2c)\dots(b+n_1c-c)r(r+c)\dots(r+n_2c-c)}{(b+r)(b+r+c)\dots(b+r+nc-c)}$$

Наряду со схемой Пойа урновая модель содержит другой интересный частный случай, а именно:

Модель Эренфестов теплового обмена между двумя неизолированными телами

С физической точки зрения здесь мы имеем дело с экспериментом, при котором k молекул распределены по двум резервуарам I и II. Случайным образом выбирается одна молекула и перемещается из своего резервуара в другой. Этот процесс по- повторяется. Каково распределение частиц через n шагов? Для того чтобы свести наш эксперимент к урновой схеме, достаточно назвать шары, содержащиеся в резервуаре I красными, а в резервуаре II чер- черными. Тогда при каждом испытании вынутый шар заменяется на шар противоположного цвета, так что мы имеем $c=-1$, $d=1$. Ясно, что в этом случае процесс может продолжаться неограниченно долго (если отсутствуют красные шары, то автоматически вынимается черный шар и заменяется на красный).

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,

$$P(A) = \frac{1^n + 2^n + \dots N^n}{N^n(N+1)}$$
$$P(AB) = P(B) = \frac{1^{n+1} + 2^{n+1} + \dots N^{n+1}}{N^{n+1}(N+1)}$$

Поскольку событие AB означает, что $n+1$ испытаний дали красные шары.

Искомая вероятность равна $P(B|A) = P(B)/P(A)$ 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 $P(B|A) \simeq \frac{n+1}{n+2}$

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 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.

4.3 △ Байесовский подход

Выше мы вычисляли некоторые условные вероятности, исходя прямо из определения вероятностного пространства Ω. Формула Байеса воспроизводит в общем виде то, что мы делали в частных примерах, и является лишь соотношением между условными вероятностями. В наших примерах мы изначально опирались на семейство событий, в частности, на группу взаимно исключающих друг друга и исчерпывающих все пространство событий $H_1, H_2, \dots$ т. е. на *полную группу событий.*

Если называть события H_k гипотезами, то имеем

1. Формула полной вероятности: для любого события A

$$P(A) = \sum_i P(A|H_i)P(H_i)$$

2. Формула Байеса для полной группы событий $\{H_1, H_2, \dots H_n\}$ и события ненулевой вероятности A:

$$P(H_k|A) = \frac{P(A|H_k)P(H_k}{\sum_i P(A|H_i)P(H_i)}$$

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 $P(H_k|A)$ 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).

Байесовский вывод 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.

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[10]

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

$$P(\theta = t_k|\xi = x_1) = \frac{P(\theta = t_k, \xi = x_1)}{P(\xi = x_1)} = \frac{P(\xi = x_1|\theta = t_k)P(\theta = t_k)}{\sum_i P(\xi = x_1|\theta = t_i)P(\theta = t_i)}$$

[10]Указанную процедуру можно итерировать, предполагается, что от этого ответ о распределении ξ должен становиться точнее.

○ Условная вероятность события относительно разбиения

Напомним, что полной группой событий или разбиением вероятностного пространства $(\Omega, \mathcal{A}, P)$ называется семейство $\mathfrak{D}$ попарно несовместных событий $D_t \in \mathcal{A}$ так чтобы $\bigoplus D_t = \Omega$, ниже всюду предполагается, что семейство $\mathfrak{D}$ конечно и состоит из событий с ненулевыми вероятностями. Случайная величина, принимающая все свои значения с ненулевыми вероятностями[11], естественным образом предоставляет такие разбиения.

Для события $A \in \mathcal{A}$ возникает набор условных вероятностей $P(A|D_i)$, из них построим новую простую *случайную величину* (которую мы специально не станем обозначать греческой буквой, а обозначим иначе!)

$$\mathcal{P}(A|\mathfrak{D})(\omega) = \sum_j P(A|D_j)\mathbf{1}_{D_j}(\omega)$$

которая называется **условной вероятностью события** A **относительно разбиения** $\mathfrak{D}$) . Если разбиение $\mathfrak{F}_\eta$, состоящее из событий вида $F_i = \{\omega|\eta(\omega) = y_i\}$ возникло из простой с.в. η, то случайная величина $\mathcal{P}(A|\mathfrak{F}_\eta)$ называется *условной вероятностью события A относительно простой с.в* η и обозначается $\mathcal{P}(A|\eta)$.

Проверьте[12] свойства условной относительно разбиения вероятности:

- $\mathcal{P}(A|\Omega) = P(A)$
- $E(\mathcal{P}(A|\mathfrak{D})) = P(A)$
- $\mathcal{P}(A + B|\mathfrak{D}) = \mathcal{P}(A|\mathfrak{D}) + \mathcal{P}(B|\mathfrak{D})$

Если задано несколько дискретных случайных величин $\eta_1, \eta_2, \dots \eta_m$, то имеется разбиение на события вида $\left\{\omega \middle| \eta_1(\omega) = y_1, \eta_2(\omega) = y_2, \dots \eta_m(\omega) = y_m\right\}$. Соответственно возникает *условная вероятность события A относительно набора с.в* $\{\eta_k\}$.

- Пусть ξ, η одинаково распределенные и независимые с.в. принимающие значения $\{0, 1\}$, найдите условную вероятность $\mathcal{P}(\xi + \eta = 1|\eta)$.

4.4 □ Групповая независимость событий, лемма Ловаса

Насколько вообще групповая независимость может быть ограничительна? Приведем пример неравенства, указывающего на тонкие свойства групповой независимости *событий*. Выберем параметры $n \geqslant d \geqslant 1$ и параметр p

$$0 < p < \begin{cases} \frac{1}{2} & d = 1 \\ \frac{d^d}{(d+1)^{d+1}} & d \geqslant 2 \end{cases}$$

Лемма Ловаса утверждает, что для набора событий $A_1, A_2, \dots A_n$ таких, что все $P(A_i) < p$ и вдобавок любое событие A_i независимо по крайней мере от $n - d$ остальных, пересечение их дополнений $\bigcap_j \overline{A_j}$ имеет строго ненулевую вероятность и, значит, непусто. Эта лемма в частности объясняет, представимо ли полное событие в виде объединения попарно независимых.

Доказательство Сначала обратим внимание на почти очевидные равенства (проверьте самостоятельно!) для непересекающихся множеств индексов $S = \{j_1, j_2, \dots j_s\}$, $R = \{j_{s+1}, \dots j_n\}$

$$P\left(\prod_{j=1}^{n} \overline{A_j}\right) = P(\overline{A_n}) \cdot P(\overline{A_{n-1}}|\overline{A_n}) \cdot P\left(\overline{A_{n-2}}|\overline{A_{n-1}} \cdot \overline{A_n}\right) \dots P\left(\overline{A_1}|\overline{A_2} \cdot \dots \overline{A_{n-1}} \cdot \overline{A_n}\right)$$

$$P\left(\prod_{i\in S} \overline{A_i} \,\middle|\, \prod_{k\in R} \overline{A_k}\right) = P\left(\overline{A_{j_1}} \,\middle|\, \prod_{i\in\{j_2,\dots j_s\}} \overline{A_i} \cdot \prod_{k\in R} \overline{A_k}\right) \cdot P\left(\overline{A_{j_2}} \,\middle|\, \prod_{i\in\{j_3,\dots j_s\}} \overline{A_i} \cdot \prod_{k\in R} \overline{A_k}\right) \dots P\left(\overline{A_{j_s}} \,\middle|\, \prod_{k\in R} \overline{A_k}\right)$$

Если теперь для любого собственного подмножества I индексов в $1, \dots n$ такого, что $j \notin I$ индукцией по мощности множества $\mathrm{card}(I)$ установим

$$P\left(A_j \middle| \prod_{i\in I} \overline{A_i}\right) \leqslant \frac{1}{d+1}$$

[11] для определенности, в этом разделе договоримся рассматривать только такие с.в.

[12] Внимание: в этом списке не обязательно все свойства верны! В некоторых книгах по теории вероятностей знак + для событий имеет специальный смысл, поэтому всегда проверяйте точный смысл обозначений.

то, поскольку

$$P\left(A_j\Big|\prod_{i\in I}\overline{A_i}\right)=1-P\left(\overline{A_j}\Big|\prod_{i\in I}\overline{A_i}\right)\quad\Rightarrow\quad P\left(\overline{A_j}\Big|\prod_{i\in I}\overline{A_i}\right)\geqslant\frac{d}{d+1}$$

будет доказано, что все члены в указанном выше первом выражении для $P\left(\prod_{j=1}^{n}\overline{A_j}\right)$ ненулевые.

Начало индукции Для пустого подмножества I имеем $P(A_j)<p<\frac{1}{e(d+1)}<\frac{1}{(d+1)}$.

Шаг индукции Выберем для рассматриваемого j множество I_j тех индексов, для которых соответствующие события **зависимы** с A_j, их количество не превосходит числа d, среди них, очевидно, есть и индекс j, поэтому $I_j\neq I$. В силу общего соотношения $P(U|V\cdot W)\cdot P(V|W)=P(U\cdot V|W)$ имеем

$$P\left(A_j\Big|\prod_{i\in I}\overline{A_i}\right)=\frac{P\left(A_j\cdot\prod_{i\in I\cap I_j}\overline{A_i}\;\Big|\prod_{i\in I\cap\overline{I_j}}\overline{A_i}\right)}{P\left(\prod_{i\in I\cap I_j}\overline{A_i}\;\Big|\prod_{i\in I\cap\overline{I_j}}\overline{A_i}\right)}$$

Числитель дроби справа может быть оценен сверху значением p исходя из независимости A_j и всех A_i : $i\in I\cap\overline{I_j}$

При этом в множестве $S=I\cap I_j$ найдется всего $s\leqslant d$ элементов $i_1,\dots i_s$. Тогда для $R=I\cap\overline{I_j}$ с учетом предположения индукции имеем для величины знаменателя:

$$P\left(\prod_{i\in S}\overline{A_i}\;\Big|\prod_{k\in R}\overline{A_k}\right)\geqslant\prod_{l=1}^{s}\left(1-\frac{1}{d+1}\right)\geqslant\left(1-\frac{1}{d+1}\right)^d$$

окончательно

$$\text{при}\quad p\leqslant\frac{d^d}{(d+1)^{d+1}}\quad\text{имеем}\quad P\left(A_j\Big|\prod_{i\in I}\overline{A_i}\right)\leqslant\frac{p}{\left(1-\frac{1}{d+1}\right)^d}\leqslant\frac{1}{d+1}$$

∎

5 Дискретные случайные величины и числовые характеристики

5.1 ○ Случайные величины и математическое ожидание

Напомним: в традиционной вероятностной модели измерений рассматривают числовую функцию $\xi:\Omega\to\mathbb{R}$, то есть *случайную величину.* На каждом элементарном исходе ω_k ей отвечает конкретное числовое значение $\xi(\omega_k)$. Случайная величина принимающая всего два значения (как правило, эти значения выбираются как $\{0,1\}$) называется битом или индикатором. Чаще всего ее обозначают $\mathbf{1}_A$, где событие A отвечает значению индикатора единица.

С помощью индикаторов ранее было сформулировано общее определение дискретной случайной величины,

$$\xi(\omega)=\sum_{\{c_k\in\xi(\Omega),k\in\mathbb{N}\}}\mathbf{1}_{\{\xi=c_k\}}$$

Опять же в приложениях общепринято вместе со случайной величиной рассматривать ее усреднение с учетом вероятности исходов $E(\xi)=\sum\xi(\omega_k)p_k$. В классической теории вероятностей это определение *математического ожидания.* Если Ω – конечное множество, то такое определение вполне корректно, но уже для общей дискретной случайной величины корректность неясна: сумма не обязана быть конечной и соответствующий ряд не обязательно сходится! К тому же в дальнейшем будет неясно откуда берется математическое ожидание, если случайная величина, скажем, дискретна, а вероятностное пространство Ω — нет. Поэтому мы Поэтому для дискретных с.в. вначале положим, что $E\left(\mathbf{1}_{\{\xi=c_k\}}\right)=P(\{\xi=c_k\})$, а для общей дискретной с.в. продоложим по линейности, *если соответствующий ряд сходится абсолютно.* Математическое ожидание тем самым определяется не для всех с.в.

$$E(\xi)=\sum_{\{a_k\in\xi(\Omega),k\in\mathbb{N}\}}E\left(\mathbf{1}_{\{\xi=a_k\}}\right)=\sum_{\{a_k\in\xi(\Omega),k\in\mathbb{N}\}}a_kP(\{\xi=a_k\})$$

Такое определение не ссылается непосредственно на элементы вероятностного пространства. Из него, вообще говоря, не сразу очевидны свойства математического ожидания, которые в классической версии определения были вполне прозрачны:

- Если $\xi \geqslant 0$, то $E(\xi) \geqslant 0$
- Для чисел a, b верно $E(a\xi + b\eta) = aE(\xi) + bE(\eta)$
- Если $\xi \geqslant \eta$, то $E(\xi) \geqslant E(\eta)$
- $|E(\xi)| \leqslant E(|\xi|)$
- Для индикатора ξ $E(\xi) = P(\xi = 1)$

Совершенно аналогично для векторнозначной случайной величины $\xi : \Omega \to Y = \mathbb{R}^n$ вводится понятие ее математического ожидания, которое в этом случае есть вектор, см. Рис. 5.

Физики называют случайные величины[13] *наблюдаемыми* и используют вместо $E(\xi)$ обозначение $\langle \xi \rangle$.

Независимость случайных величин и математическое ожидание

Напомним, что для подмножества $U \subset \mathbb{R}$ его прообраз $\xi^{-1}(U)$ является событием, вероятность которого можно обозначать $P(\xi \in U)$. Случайные величины ξ и η называются попарно независимыми, если для любых подмножеств $U, V \subset \mathbb{R}$ независимы события $\xi \in U$ и $\eta \in V$. Совокупная независимость системы многих случайных величин определяется далее очевидным образом. Независимость случайных величин представляется на первый взгляд довольно сильным условием, которое однако в схеме независимых повторений легко выполнимо.

- В модели независимых повторений одинаковых измерений соответствующие случайные величины (измерения при повторениях i, j, $i \neq j$) независимы — восстановите все формальные детали и проверьте.
- Выразить математическое ожидание произведения двух независимых случайных величин ξ и η через математические ожидания сомножителей.
- Выразить $E\left[(\xi + \eta)^2\right]$ для двух независимых случайных величин ξ и η таких, что $E\xi = E\eta = 0$, через $E\xi^2$ и $E\eta^2$.

5.2 ○ Моменты случайных величин

Математические ожидания степеней ξ^k случайной величины ξ при $k = 0, 1, 2 \ldots$ называются ее *моментами* $\mathfrak{m}_k(\xi)$, моменты для случайной ведичины $\xi - E(\xi)$ называются *центральными* моментами $\overset{0}{\mathfrak{m}}_k(\xi)$ случайной величины ξ. Второй центральный момент чаще называется *дисперсией* и обозначается[14] $D(\xi)$, а квадратный корень из дисперсии — *стандартным уклонением* и очень часто обозначается как $\sigma(\xi)$ или σ_ξ.

- Чему равен момент степени 0?
- Выразить дисперсию случайной величины через ее моменты, не являющиеся центральными. Ответ: $D(\xi) = E(\xi^2) - (E(\xi))^2$
- Выразить дисперсию суммы *независимых* случайных величин через дисперсии слагаемых.

Взаимные моменты случайных величин

Для пары случайных величин ξ и η вводят их взаимные моменты $\mathfrak{m}_{p,q}(\xi, \eta) = E(\xi^p \eta^q)$ и соответственно их центральные аналоги $\overset{0}{\mathfrak{m}}_{p,q}(\xi, \eta) = E((\xi - E(\xi))^p (\eta - E(\eta))^q)$, при этом $\overset{0}{\mathfrak{m}}_{1,1}(\xi, \eta)$ называется ковариацией, а частное от деления ковариации на $\sigma_\xi \sigma_\eta$ — коэффициентом корреляции. Разумеется, определения имеют смысл лишь когда соответствующие ряды сходятся.

- Коэффициент корреляции независимых случайных величин всегда равен одному и тому же числу (укажите какому).
- Выразить дисперсию суммы двух случайных величин через моменты и взаимные моменты слагаемых. Указание: рассмотрите сначала две центрированные (то есть с нулевым мат.ожиданием) случайные величины ξ, η. Для них $D(\xi) = E(\xi^2)$, $D(\xi + \eta) = E\left(\xi^2 + 2\xi\eta + \eta^2\right) = D(\xi) + 2E(\xi\eta) + D(\eta)$. Общий случай следует, так как $\xi - E(\xi)$ – центрированная случайная величина.

[13] но помните: *в квантовой теории вероятности определение случайной величины вовсе не $\Omega \to \mathbb{R}$, а радикально иное, связанное со структурой линейного пространства на Ω!*

[14] В англоязычной литературе чаще встречается обозначение $Var(\xi)$

Упражнение

- Привести пример дискретной случайной величины у которой не определено математическое ожидание.
- Привести пример дискретной случайной величины у которой математическое ожидание определено, а дисперсия не определена.
- Есть ли пример дискретных случайных величин у которых математическое ожидание и дисперсии определены, а коэффициент корреляции не определен?

5.3 △ Условные математические ожидания

Условное относительно события мат.ожидание Если для дискретной с.в. $\xi : \Omega \to \mathbb{R}$ определено ее математическое ожидание $E(\xi)$ и задано событие ненулевой вероятности $A \subset \Omega$, то *условным математическим ожиданием* $E(\xi|A)$ *с.в.* ξ *относительно события* A называется число

$$E(\xi|A) = \frac{E(\xi \cdot \mathbf{1}_A)}{P(A)}$$

Условное относительно разбиения мат.ожидание Для с.в. $\xi(\omega) = \sum_i x_i \mathbf{1}_{D_k}(\omega)$ *условным математическим ожиданием* $\mathcal{E}(\xi|\mathfrak{F})$ *с.в.* ξ *относительно разбиения* $\mathfrak{F}$ (соответственно, когда $\mathfrak{F} = \mathfrak{F}_\eta$ условным математическим ожиданием $\mathcal{E}(\xi|\eta)$ относительно с.в η) называется случайная величина

$$\sum_k x_k \mathcal{P}(D_k|\mathfrak{F})(\omega)$$

В частности, если ξ измерима относительно разбиения $\mathfrak{F}$, то $\mathcal{E}(\xi|\mathfrak{F}) = \xi$

Важное замечание Заметим, что на исходах из события $\{\eta = \text{const}\}$ условное математическое ожидание $\mathcal{E}(\xi|\eta)$ относительно с.в η также принимает постоянные значения — иными словами $\mathcal{E}(\xi|\eta)$ можно рассматривать как композицию случайной величины η и некоторого отображения $g : \mathbb{R} \to \mathbb{R}$, то есть $\mathcal{E}(\xi|\eta) = g(\eta)$

Замечание Определения выше очевидным образом (сделайте это!) обобщаются до условного математического ожидания при условии не одной с.в. η, но целой системы случайных величин $\eta_1, \eta_2 \dots \eta_N$ в том смысле, что соответствующее необходимое разбиение строится при помощи всех разбиений $\mathfrak{F}_{\eta_k}$.

Замечание Альтернативная версия определения условного математического ожидания $\mathcal{E}(\xi|\mathfrak{F})$ такова: сначала для каждого $F_k \in \mathfrak{F}$ определить число $E(\xi|F_k)$ (условное мат.ожидание относительно события $F_k \in \mathfrak{F}$)

$$E(\xi|F_k) = \sum_j x_j P(D_j|F_k)$$

а потом уже положить

$$\mathcal{E}(\xi|\mathfrak{F})(\omega) = \sum_k E(\xi|F_k)\mathbf{1}_{F_k}(\omega)$$

Проверьте следующие простейшие свойства условного математического ожидания:

- Для $\xi = \mathbf{1}_A$ выполнено $\mathcal{E}(\xi|\mathfrak{D}) = \mathcal{P}(A|\mathfrak{D})$.
- Проверьте, что условное математическое ожидание случайной величины η, относительно порожденного ею разбиения $\mathfrak{F}_\eta$ совпадает с этой случайной величиной: $\mathcal{E}(\eta|\mathfrak{F}_\eta) = \eta$, другими словами, $\mathcal{E}(\eta|\eta) = \eta$
- Проверить $\mathcal{E}\left(a \cdot \alpha + b \cdot \beta|\mathfrak{F}\right) = a \cdot \mathcal{E}\left(\alpha|\mathfrak{F}\right) + b \cdot \mathcal{E}\left(\beta|\mathfrak{F}\right)$
- $\mathcal{E}(\text{const}|\mathfrak{D}) = \text{const}$
- Если $\forall i, k$ события $F_i \in \mathfrak{F}$ и $\{\xi = x_k\}$ независимы, то $\mathcal{E}(\xi|\mathfrak{F}) = E(\xi)$
- Если ξ, η независимы, то $\mathcal{E}(\xi|\eta) = E(\xi)$

Вот некоторые менее очевидные результаты о свойствах условного математического ожидания.

Предложение 5.1 $E\left(\mathcal{E}(\xi|\mathfrak{D})\right) = E(\xi)$

Доказательство

$$E\left(\mathcal{E}(\xi|\mathfrak{D})\right) = E\left[\sum_j x_j \mathcal{P}(D_j|\mathfrak{D})\right] = \sum_j x_j E\left[\mathcal{P}(D_j|\mathfrak{D})\right] = \sum_j x_j P(D_j) = E(\xi).$$

Для двух простых случайных величин ξ, η их произведение $\xi\eta$ — также простая с.в. и разбиение, ею порожденное, очевидно таково, что обе с.в. ξ, η измеримы относительно него: действительно, оно состоит из событий вида $D_i F_j$, $D_j \in \mathfrak{D}_\xi$, $F_i \in \mathfrak{F}_\eta$ и $\xi\eta = \sum_{i,j} x_j y_i \mathbf{1}_{D_j F_i}$.

Предложение 5.2 *Пусть с.в. η $\mathfrak{F}$- измерима, тогда $\mathcal{E}(\xi\eta|\mathfrak{F}) = \eta\mathcal{E}(\xi|\mathfrak{F})$*

Доказательство

$$\begin{aligned}
\mathcal{E}(\xi\eta|\mathfrak{F})(\omega) &= \sum_j \sum_i x_j y_i \mathcal{P}(D_j F_i|\mathfrak{F})(\omega) = \sum_j \sum_i x_j y_i \sum_k P(D_j F_i|F_k)\mathbf{1}_{F_k}(\omega) \\
&= \sum_j \sum_i x_j y_i P(D_j F_i|F_i)\mathbf{1}_{F_i}(\omega) = \sum_j \sum_i x_j y_i P(D_j|F_i)\mathbf{1}_{F_i}(\omega) \\
\eta\mathcal{E}(\xi|\mathfrak{F})(\omega) &= \left[\sum_i y_i \mathbf{1}_{F_i}(\omega)\right] \cdot \sum_j x_j \mathcal{P}(D_j|\mathfrak{F})(\omega) = \left[\sum_i y_i \mathbf{1}_{F_i}(\omega)\right] \cdot \sum_j x_j \left[\sum_k P(D_j|F_k)\mathbf{1}_{F_k}(\omega)\right] \\
&= \sum_i \sum_j y_i x_j P(D_j|F_i)\mathbf{1}_{F_i}(\omega);
\end{aligned}$$

∎

Следствие 5.3 *Для всякой функции $f(\eta)$ выполнено $E\left(\xi f(\eta)\right) = E\left[f(\eta)\mathcal{E}(\xi|\eta)\right]$*

Предложение 5.4 *Если дано разбиение $\mathfrak{D}$ и его «измельчение» $\tilde{\mathfrak{D}}$ (что записывается как $\tilde{\mathfrak{D}} \rhd \mathfrak{D}$), то выполняется $\mathcal{E}\left[\mathcal{E}(\xi|\tilde{\mathfrak{D}})\Big|\mathfrak{D}\right] = \mathcal{E}(\xi|\mathfrak{D})$.*

Доказательство несколько громоздко, но несложно и вполне аналогично доказательствам предыдущих Предложений; мы его пропустим.

△ Важное свойство условных математических ожиданий

Напомним, что условное математическое ожидание $\mathcal{E}(\xi|\eta)$ можно рассматривать как композицию $\mathcal{E}(\xi|\eta) = f(\eta)$ для некоторой функции f. В этом разделе мы рассматриваем только простые случайные величины, но следующий результат, на самом деле, справедлив и в гораздо более общем случае.

Теорема 5.5 *Пусть ξ, η, α случайные величины, определенные на одном вероятностном пространстве Ω, причем α представимо композицией $\alpha = g(\eta)$ для некоторой функции $g : \mathbb{R} \to \mathbb{R}$. Тогда минимум выражения*[15] *$E\left[(\alpha - \xi)^2\right]$ достигается в случае, когда $\alpha = \mathcal{E}(\xi|\eta)$.*

Доказательство

$$\begin{aligned}
E\left[(\xi-\alpha)^2\right] =& E\left[\Big(\xi - \mathcal{E}(\xi|\eta) + \mathcal{E}(\xi|\eta) - \alpha\Big)^2\right] \\
=& E\left[(\xi - \mathcal{E}(\xi|\eta))^2\right] + 2E\left[\Big(\xi - \mathcal{E}(\xi|\eta)\Big)\Big(\mathcal{E}(\xi|\eta) - \alpha\Big)\right] + E\left[(\mathcal{E}(\xi|\eta) - \alpha)^2\right]
\end{aligned}$$

Используя Предложение 5.1 и Следствие 5.3, позволяющее выносить η-измеримые случайные величины за знак условного математического ожидания, получим во втором слагаемом

$$E\left[\Big(\xi - \mathcal{E}(\xi|\eta)\Big)\Big(\mathcal{E}(\xi|\eta) - \alpha\Big)\right] = E\left[\mathcal{E}\Big((\xi - \mathcal{E}(\xi|\eta))(\mathcal{E}(\xi|\eta) - \alpha)\Big|\eta\Big)\right] = E\left[\Big(\mathcal{E}(\xi|\eta) - \alpha\Big) \cdot \mathcal{E}\Big(\xi - \mathcal{E}(\xi|\eta)\Big|\eta\Big)\right]$$

[15] по смыслу это квадрат «расстояния между с.в.»

Однако в силу Предложения 5.4 это второе слагаемое равно нулю, потому что

$$\mathcal{E}\Big(\xi - \mathcal{E}(\xi|\eta)\Big|\eta\Big) = \mathcal{E}\Big(\xi|\eta\Big) - \mathcal{E}\Big(\mathcal{E}(\xi|\eta)\Big|\eta\Big) = \mathcal{E}\Big(\xi|\eta\Big) - \mathcal{E}\Big(\xi|\eta\Big) = 0$$

Таким образом, $E\left[\Big(\xi - \alpha\Big)^2\right] \geqslant E\left[\Big(\xi - \mathcal{E}(\xi|\eta)\Big)^2\right]$ ∎

5.4 □ Мартингалы

В теории вероятностей исследуют последовательности случайных величин, причем i.i.d-последовательности являются основным объектом изучения. Также в современной теории важным для изучения объектом являются последовательности, на которые наложено достаточно жесткое условие в терминах условных мат.ожиданий: предыдущий член последовательности совпадает с условным математическим ожиданием последующего. Для информации приведем определение.

Пусть на вероятностном пространстве $(\Omega, \mathcal{A}, P)$ задана последовательность разбиений $\mathfrak{D}_1 \lhd \mathfrak{D}_2 \lhd \ldots \lhd \mathfrak{D}_n$ (т.е.такая, что каждое последующее разбиение является измельчением предыдущего).

Определение Последовательность с.в. $\{\xi_k\}_{1 \leqslant k \leqslant n}$ называется *мартингалом относительно разбиений* $\mathfrak{D}_1 \lhd \mathfrak{D}_2 \lhd \ldots \lhd \mathfrak{D}_n$ (что записывается также как $\{\xi_k, \mathfrak{D}_k\}$), если

1. каждая с.в. ξ_k $\mathfrak{D}_k$-измерима (и, следовательно, $\mathfrak{D}_m$-измерима при $m > k$),
2. $\xi_k = \mathcal{E}(\xi_{k+1}|\mathfrak{D}_k)$ при $1 \leqslant k < n$

В случае, когда когда разбиения порождаются самой последовательностью, то есть $\mathfrak{D}_k = \mathfrak{D}_{\xi_1, \xi_2 \ldots \xi_k}$, то последовательность называется мартингалом без отсылки к разбиениям.

Мартингалы и симметричное случайное блуждание

В одномерном симметричном случайном блуждании (из нуля) последовательность положений S_k точки на оси в момент времени $k \geqslant 1$ образует мартингал. Действительно, любой элемент разбиения $\mathfrak{D}_k = \mathfrak{D}_{\xi_1 \xi_2 \ldots \xi_k}$ характеризуется путем длины k, построенном по независимым шагам-битам η_i. Легко видеть также, что независимость η_{k+1} и S_k влечет $\mathcal{E}(\eta_{k+1}|\mathfrak{D}_k) = E(\eta_{k+1}) = 0$, поэтому

$$\mathcal{E}(S_{k+1}|\mathfrak{D}_k) = \mathcal{E}(S_k + \eta_{k+1}|\mathfrak{D}_k) = \mathcal{E}(S_k|\mathfrak{D}_k) + \mathcal{E}(\eta_{k+1}|\mathfrak{D}_k) = S_k + E(\eta_{k+1}) = S_k$$

5.5 △ Неравенства для характеристик случайных величин

Теорема 5.6 (Неравенство Маркова) *Пусть ξ - неотрицательная случайная величина, у которой есть математическое ожидание. Тогда $\forall \varepsilon > 0$ $P\{\xi \geqslant \varepsilon\} \leqslant \frac{E(\xi)}{\varepsilon}$.*

Доказательство $E(\xi) \geqslant E\left(\xi \cdot I_{(\xi \geqslant \varepsilon)}\right) \geqslant \varepsilon E\left(I_{(\xi \geqslant \varepsilon)}\right) = \varepsilon P\{\xi \geqslant \varepsilon\}$ ∎

Простыми подстановками отсюда получаются еще два утверждения, они называются неравенствами Чебышёва:

Теорема 5.7 *Если соответствующие моменты с.в. определены, то $P\{\xi \geqslant \varepsilon\} \leqslant \frac{E(\xi^2)}{\varepsilon^2}$*

Теорема 5.8 *Если соответствующие моменты с.в. определены, то $P\{|\xi - E(\xi)| \geqslant \varepsilon\} \leqslant \frac{D(\xi)}{\varepsilon^2}$*

Теорема 5.9 (Неравенство Чернова) *Если соответствующие моменты с.в. определены, то* $P\{\xi \geqslant \varepsilon\} \leqslant \inf\limits_{t \geqslant 0} e^{-t\varepsilon} E\left(e^{t\xi}\right)$

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

Большие уклонения при случайном блуждании

Вспомним встречавшийся ранее пример с одномерным дискретным случайным блужданием: рассматривалась сумма последовательности независимых шагов вправо-влево, то есть сумма 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. $S_n = \sum_{i=1}^n \eta_i$, причем $E(e^{t\eta}) = 1-p+pe^t = g(t)$. В силу независимости слагаемых $E(e^{tS_n}) = (g(t))^n$. Отсюда

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

$$\inf_{u>0} e^{-n[\varepsilon u - \ln g(u)]} = e^{-n \sup\limits_{u>0}[\varepsilon u - \ln g(u)]} = e^{-nr(\varepsilon)}$$

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

□ Неравенство Хёфдинга

Лемма 5.10 *Для с.в. ξ, такой что $a \leqslant \xi \leqslant b$ и $E(\xi) = 0$ и любого $t > 0$ выполнено $E\left(e^{t\xi}\right) \leqslant \exp\left[\frac{1}{8}t^2(b-a)^2\right]$*

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 $E(\xi) = 0$ имеем $E\left(e^{t\xi}\right) \leqslant \frac{b}{b-a}e^{ta} - \frac{a}{b-a}e^{tb} = e^{h\left(\frac{t}{b-a}\right)}$, 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 $E\left(e^{t\xi}\right) \leqslant \exp\left[\frac{1}{8}t^2(b-a)^2\right]$ ■

Теорема 5.11 (Неравенство Хёфдинга) *Пусть случайные величины $\xi_1, \xi_2, \ldots \xi_n$ независимы, $a_i \leqslant \xi_i \leqslant b_i$ и у них одинаковое математическое ожидание $E(\xi_i) = m$, тогда для $\varepsilon > 0$ и их усреднения $\overline{\xi}$ при $\forall t > 0$*

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

В частности, для таких **и вдобавок одинаково распределенных** *с.в. минимум в правой части достигается при $t = \frac{4\varepsilon}{(b-a)^2}$ и потому*

$$P\left\{|\overline{\xi} - m| \geqslant \varepsilon\right\} \leqslant 2\exp\left[-\frac{2n\varepsilon^2}{(b-a)^2}\right]$$

Применительно к схеме Бернулли неравенство Хёфдинга означает, что наблюдаемая в эксперименте частота «успехов» распределена вокруг вероятности «успеха» в одном испытании с экспоненциально быстрым убыванием по n вероятностей возможных отклонений.

Доказательство Без ограничения общности положим $m = 0$, далее запишем неравенство Чернова для $t > 0$, учитывая свойство независимости:

$$P\left\{\overline{\xi} \geqslant \varepsilon\right\} = P\left\{\sum_i \xi_i \geqslant n\varepsilon\right\} \leqslant e^{-nt\varepsilon}E\left[\exp\left(t\sum_i \xi_i\right)\right] = e^{-nt\varepsilon}\prod_i E\left(e^{t\xi_i}\right)$$

По Лемме 5.10 $E\left(e^{t\xi_i}\right) \leqslant \exp\left[\frac{1}{8}t^2(b_i-a_i)^2\right]$, теперь, используя $P\left\{|\overline{\xi}| \geqslant \varepsilon\right\} = P\left\{\overline{\xi} \geqslant \varepsilon\right\} + P\left\{-\overline{\xi} \geqslant \varepsilon\right\}$, получаем требуемое. ■

5.6 △ Закон больших чисел в форме Чебышёва

Теорема 5.12 *Пусть случайные величины* $\xi_1, \xi_2, \ldots \xi_n$ *независимы и их дисперсии меньше некоторой константы* C. *Тогда для любого положительного* ε *при увеличении* $n \to \infty$ *вероятность*

$$P\left(\left|\frac{\xi_1+\xi_2+\ldots+\xi_n}{n} - \frac{E(\xi_1)+E(\xi_2)+\ldots+E(\xi_n)}{n}\right| \geqslant \varepsilon\right)$$

стремится к нулю.

Доказательство По неравенству Чебышёва достаточно установить, что $D\left((\xi_1+\xi_2+\ldots+\xi_n)/n\right) \to 0$. Но поскольку в данном случае дисперсия суммы равна сумме дисперсий имеем:

$$D\left(\frac{\xi_1+\xi_2+\ldots+\xi_n}{n}\right) = \frac{1}{n^2} D\left(\xi_1+\xi_2+\ldots+\xi_n\right) = \frac{1}{n^2}\sum_i D(\xi_i) \leqslant \frac{nC}{n^2} \to 0$$

Далее мы увидим, что эта (довольно абстрактная) формулировка в применении к схеме Бернулли независимых повторений опыта с двумя исходами соответствует представлению о том, что частота «успехов» стремится к вероятности «успеха» в одном испытании. Это тот случай пользы теории, когда прямое доказательство прикладного результата (про «успехи» в повторениях) достаточно громоздко и запутано, а общая теорема короткая и прямая. Наше доказательство взято из книги В.Тутубалин "Теория вероятностей и случайных процессов. Основы математического аппарата и прикладные аспекты", Глава 1, §6, разделы 6.3, 6.4

5.7 △ Техническое: использование производящих функций

Для принимающей *только неотрицательные целые значения дискретной случайной величины* η возможно определить *производящую функциию* $Q_\eta(z)$ *случайной величины* η [16]

$$Q_\eta(z) = P(\eta=0) + P(\eta=1)z + \ldots + P(\eta=k)z^k + \ldots$$

Из анализа известно, что этот степенной ряд заведомо сходится при $|z| \leqslant 1$ поскольку сумма всех вероятностей $P(\eta=k)$ равна 1. Основная идея использования производящей функции возникает при рассмотрении значения ее производной:

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

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

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

$$D(\eta) = \overset{0}{\mathfrak{m}}_k(\eta) = Q''_\eta(1) + Q'_\eta(1) - \left(Q'_\eta(1)\right)^2$$

Пусть ξ и η неотрицательные *независимые* целочисленные случайные величины с распределениями вероятностей $P(\xi=j)=a_j$ и $P(\eta=k)=b_k$. Событие $(\xi=j, \eta=k)$ в силу независимости имеет вероятность a_jb_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 $a_0b_m + a_1b_{m-1} + \ldots + a_mb_0$. 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 $Q_{\xi+\eta}(z) = Q_\xi(z)Q_\eta(z)$

Таким образом, производящие функциии удобны для определения моментов дискретных случайных величин специального вида и *для вывода формул суммы независимых случайных величин.* Таким образом, для дискретных с.в. специального вида кое-что известно и о законе распределения суммы большого числа независимых случайных величин этого же вида: надо перемножить соответствующие производящие функции и это произведение даст производящую функцию суммы. Разложение ее в ряд Маклорена определит соответствующие вероятности в дискретном распределении.

[16] выбор обозначения переменной z не случаен: в будущем будет объяснена возможная роль комплексных чисел в этой конструкции

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

$$\beta(\omega) = \xi_0(\omega) + \ldots + \xi_{\eta(\omega)}(\omega)$$

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

$$P(\beta = n) = \sum_{k=0}^{\infty} P(\eta = k) P(\xi_0 + \ldots + \xi_k = n)$$

Чтобы найти вероятность $P(\xi_0 + \ldots + \xi_k = n)$ достаточно разложить $(Q_\xi(z))^k$ в ряд и взять коэффициент при z^n. Таким образом, производящая функция $Q_\beta(z)$ отвечает выражению

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

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, $Q_\beta(z) = Q_\eta(q + pz)$

6 Дополнение о геометрических вероятностях: равномерные распределения в больших размерностях

Ранее на упражнениях вы уже сталкивались с недискретными геометрическими моделями вероятностных пространств, когда вероятностная мера интерпретировалась как геометрическая характеристика: длина, площадь, объем. Но есть специфика геометрических объектов высоких размерностей: проявляются неожиданные связи между вероятностными и метрическими свойствами, причем они часто противоречат интуиции, основанной на опыте работы с геометрическими объектами малой размерности.

Начнем с тривиального примера: если в n-мерном единичном кубе уменьшить каждое ребро до $(1 - \varepsilon)$, $0 < \varepsilon \ll 1$, то при $n \to \infty$ его (гипер)объем будет $(1 - \varepsilon)^n \to 0$. Аналогичное утверждение справедливо и для многомерного шара — что в общем можно трактовать как свойство многомерных объемов концентрироваться вблизи границ областей. В частности, при случайном выборе точки вероятность получить близкую к границе точку больше.

Это наблюдение достаточно важно для понимания некоторых статистических свойств систем из многих «равноправных» объектов. Соответствующая область знания обычно именуется статистической механикой, в средней школе соответствующие результаты излагают на примере кинетической теории идеального газа, и говорят о стационарных состояниях, обладающих свойством сохранять во времени макроскопические характеристики газа (вроде температуры, энтропии или давления). Стационарные состояния являются примером проявления именно свойства концентрации, то есть «статистического постоянства» значений макроскопической характеристики, которая достаточно симметрична по отношению к входящим в нее переменным и потому практически неизменна во времени.

В этом разделе рассматриваем свойства равномерных распределений на несложных многомерных объектах; свойство концентрации формулируем в терминах вероятностей попадания в подобласти геометрических объектов.

6.1 □ 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:

$$P\Big\{\Big|\sum_{k=1}^{n} \xi_k \cdot \frac{1}{\sqrt{n}}\Big| \geqslant r\sqrt{n}\Big\} = P\Big\{\Big|\sum_{k=1}^{n} \frac{\xi_k}{n}\Big| \geqslant r\Big\} \leqslant 2e^{-2nr^2} \tag{3}$$

Иными словами, ортогональная проекция многомерной случайной величины на главную диагональ (с единичным направляющим вектором $\left(\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{n}}, \dots \frac{1}{\sqrt{n}}\right)$) 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 $\left(\pm\frac{1}{\sqrt{n}}, \pm\frac{1}{\sqrt{n}}, \dots \frac{1}{\sqrt{n}}\right)$).

6.2 □ Концентрация на сфере

Координатные функции в многомерном шаре не являются независимыми, поэтому рассуждение из предыдущего раздела напрямую применить не получится. Несколько более сложная конструкция, тем не менее, дает сходный результат о концентрации проекций точек сферы (и, значит, точек шара) на радиус.

В евклидовом пространстве $\mathbb{R}^n$ рассмотрим гиперсферу $S^{n-1}(r)$ радиуса r с центром в начале координат. Введем в $\mathbb{R}^n$ сферические координаты, отсчитывая угол ψ от положительного направления первой координатной оси, которую мы будем обозначать как ось x. Значит, $x = r\cos\psi$,

$$dx = -r\sin\theta d\psi \qquad \sin\psi = \frac{\sqrt{r^2 - x^2}}{r} = \sqrt{1 - \frac{x^2}{r^2}}$$

Если обозначить площадь $n-2$-мерной сферы радиуса ρ через $\sigma_{n-2}(\rho)$, $\sigma_{n-2}(1) = \sigma_{n-2}$, то площадь элементарного сферического слоя, отвечающего промежутку углов $(\theta, \psi + d\psi)$ дается формулой

$$\sigma_{n-2}(r\sin\psi)\cdot rd\psi = \sigma_{n-2}\cdot r^{n-1}\sin^{n-2}\psi d\psi = \sigma_{n-2}r^{n-2}\sin^{n-3}\psi\ r\sin\psi d\psi = -\sigma_{n-2}r^{n-2}\left[1 - \frac{x^2}{r^2}\right]^{\frac{n-3}{2}} dx$$

Нас будет интересовать (гипер)площадь «сферической шапочки» (она же «круг на сфере») — подмножества точек на сфере, отвечающих $0 \leqslant \psi \leqslant \varepsilon \leqslant \frac{\pi}{2}$. По соображениям простоты последующих выкладок удобно сделать замену переменной угла: $\theta = \frac{\pi}{2} - \psi$, так что (гипер)площадь сферической шапочки , проектирующейся на радиус будет

$$\sigma_{n-2}r^{n-2}\int\limits_{\varepsilon}^{\frac{\pi}{2}}\cos^{n-2}\theta d\theta$$

а доля от (гипер)площади всей сферы в этой шапочке (то есть вероятность множества в проекции на ось !) соответственно равна отношению

$$P([a,b]) = \frac{\int\limits_{\varepsilon}^{\frac{\pi}{2}}\cos^{n-2}\theta d\theta}{2\int\limits_{0}^{\frac{\pi}{2}}\cos^{n-2}\theta d\theta} \tag{4}$$

Для краткости обозначим пока что знаменатель последнего выражения через I_{n-2}, тем самым речь идет о вычислениях при $m \gg 1$ интегралов вида $\int\limits_{\varepsilon}^{\pi/2}\cos^m\theta d\theta$.

Нам еще понадобится при $0 \leqslant t \leqslant \pi/2$ неравенство $\cos t \leqslant e^{-t^2/2}$, его доказательство простое: при $0 \leqslant t \leqslant \pi/2$ для из ряда Маклорена для $\operatorname{tg} t$ очевидно, что $t \leqslant \operatorname{tg} t = \frac{\sin t}{\cos t}$, поэтому производная на указанном интервале отношения двух (равных друг другу в общем максимуме $t = 0$) положительных функций $\frac{\cos t}{e^{-t^2/2}}$ при $t > 0$ строго отрицательна:

$$\frac{t\cos t - \sin t}{e^{-t^2/2}} = \cos t\cdot\frac{t - \operatorname{tg} t}{e^{-t^2/2}} < 0$$

Для указанного интеграла имеем:

$$\int_{\varepsilon}^{\frac{\pi}{2}} \cos^m(\theta)d\theta = \frac{1}{\sqrt{m}} \int_{\varepsilon\sqrt{m}}^{\frac{\pi}{2}\sqrt{m}} \cos^m(s/\sqrt{m})ds$$
$$\leqslant \frac{1}{\sqrt{m}} \int_{\varepsilon\sqrt{m}}^{\frac{\pi}{2}\sqrt{m}} e^{-s^2/2}ds \leqslant \frac{1}{\sqrt{m}} \cdot e^{-\varepsilon^2 m/2} \int_0^{(\frac{\pi}{2}-\varepsilon)\sqrt{m}} e^{-s^2/2}ds$$
$$\leqslant \frac{1}{\sqrt{m}} \cdot e^{-\varepsilon^2 m/2} \sqrt{\frac{\pi}{2}} = \sqrt{\frac{\pi}{2m}} \cdot e^{-\varepsilon^2 m/2}$$

Тем самым вероятность $P(x \geqslant \varepsilon)$ попасть в проекцию шапочки сферы S^{n-1} на ось не превосходит быстро убывающую при $n \to \infty$ функцию

$$\frac{1}{\sqrt{n-2}I_{n-2}} \cdot \sqrt{\frac{\pi}{8}} \cdot e^{-\frac{1}{2}\varepsilon^2(n-2)}$$

Попутно это же вычисление при $\varepsilon = 0$ позволяет написать (и без того уже нам известные см. задачи выше) соотношения между величинами I_m. Во-первых, $\sqrt{m}I_m \leqslant \sqrt{\frac{\pi}{2}}$, кроме того, $\sqrt{m}I_m \geqslant \sqrt{m-2}I_{m-2}$. Поэтому

$$\sqrt{m}I_m \geqslant \min\left(I_1, \sqrt{2}I_2\right) = \min\left(1, \frac{\pi}{4}\sqrt{2}\right) = 1$$

Для шапочек на сфере S^{n+1} за счет переиндексации формула выглядит чуть проще: вероятностная мера «сферической шапочки» на сфере S^{n+1} ограничена сверху величиной $\sqrt{\frac{\pi}{8}}e^{-\frac{n}{2}\varepsilon^2}$, также *можно показать*, что $\lim_{n\to\infty} \sqrt{n}I_n = \sqrt{\frac{\pi}{2}}$.

В частности вычисление показало, что подавляющая часть площади многомерной сферы сосредоточена в малой окрестности ее (любого) экватора. Далее выбранные наугад два единичных вектора в пространстве $\mathbb{R}^n$ большой размерности n с очень большой вероятностью окажутся почти ортогональными (например, вероятность того, что их скалярное произведение сколь-нибудь заметно отклонится от нуля, быстро убывает с ростом величины отклонения). Действительно, все направления в пространстве $\mathbb{R}^n$ считаются равноправными, векторы пары случайны и независимы. Если один из них выбран, то направление второго с большой вероятностью окажется в окрестности экватора единичнгой сферы, сопряженного направлению первого вектора.

7 Основные дискретные модели

В естествознании имеется базовый набор примеров дискретных вероятностных пространств и связанных с ними случайных величин. Значение этих общепринятых примеров черезвычайно 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. Схема рассмотрения включает производящую функцию распределения, о связи которой с моментами распределения надо сказать несколько слов.

7.1 Индикатор

Случайная величина принимает всего два значения:

$\mathbf{1}_{\{\omega\|\omega\in A\}}$	0	1
	q	p

Модель

Распределение возникает рассмотрении опыта с двумя исходами (однократное бросание неровной монеты), которые мы назовем герб и решка или успех и неудача. Мы уже использовали эту с.в. при явных описание пространства элементарных исходов и событий.

Производящая функция

Явное выражение характеристической функции индикатора со значениями 0,1: $Q(z) = (q + pz)$.

Основные характеристики

Непосредственным вычислением устанавливаем, что $E(\eta) = p$ и $D(\eta) = pq$

7.2 Распределение Бернулли (биномиальное распределение)

Случайную величину традиционно принято обозначать S_n, где $n > 0$ целочисленный параметр.

S_n	0	$\ldots$	k	$\ldots$	n
	$\binom{n}{0}q^n$	$\ldots$	$\binom{n}{k}p^k q^{n-k}$	$\ldots$	$\binom{n}{n}p^n$

Модель

Распределение Бернулли возникает при рассмотрении n-кратных независимых повторений опыта с двумя исходами. Пространство элементарных исходов можно отождествить с последовательностями длины n из двух букв Г и Р (герб и решка). Вероятность элементарного исхода назначается по правилу независимости в модели повторения. Из комбинаторных формул ясно, что всего таких последовательностей 2^n, вероятность элементарного исхода в котором в точности k букв Г равна $p^k q^{n-k}$, а всего таких исходов, входящих в событие « произошло ровно k успехов» будет $\binom{n}{k}$

Случайная величина

Случайная величина сопоставляет каждому элементарному исходу число успехов в нем. Понятно, что так определенная случайная величина S_n является суммой n независимых индикаторов успеха, то есть $S_n = \eta_1 + \eta_2 + \ldots + \eta_n$

Производящая функция

Биномиальное распределение совпадает с суммой независимых индикаторов, поэтому производящая функция биномиального распределения равна $(q + pz)^n$.

Основные характеристики

Воспользуемся свойствами математического ожидания и дисперсии, чтобы убедиться в том, что $E(S_n) = np$ (аддитивность математического ожидания), $D(S_n) = npq$ (дисперсия суммы независимых индикаторов равна сумме их дисперсий).

7.3 Геометрическое распределение

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

и

α	0	1	$\ldots$	k	$\ldots$
	p	qp	$\ldots$	$q^k p$	$\ldots$

Модель

Геометрическое распределение возникает при рассмотрение независимых повторений эксперимента с двумя исходами до наступления первого успеха. 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. Надо только проверить, что это действительно распределение, в самом деле:

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

Случайная величина

Случайная величина α принимает на исходе РРРРРР…РГ (k решек подряд) значение равное k, считая тем самым время (в повторениях) до первого успеха.

Производящая функция

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

Основные характеристики

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

- Найдите самостоятельно дисперсию геометрического распределения.

7.4 Распределение Паскаля

Это распределение является обобщением предыдущего и определено для натуральных значений параметра r:

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

Проверка свойств и заодно обобщенная формула Бинома Ньютона

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

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

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

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

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: РРРГРРГРГ...РГ, общее число букв Г в последовательности равно r. Вероятности элементарных событий выбираются по правилу независимых повторений. Нетрудно заметить, что всего таких последовательностей фиксированной длины $r+k$ существует в точности $\binom{r+k-1}{r-1} = \binom{r+k-1}{k}$.

Случайная величина

Как и для геометрического распределения случайная величина на исходе, кодируемом последовательностью длины $r+k$ принимает значение $r+k-1$, показывая тем самым число неудач, предшествовавших первому успеху. С другой стороны, эти 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: $E(\xi) = rq/p$. Дисперсия суммы независимых случайных величин равна сумме дисперсий слагаемых, выразите ее в явном виде самостоятельно.

7.5 Формула Пуассона

Рассмотрим теперь случай перехода к пределу в биномиальном распределении при одновременном увеличении $n \to \infty$ и уменьшении $p(n)$: $p(n) = \lambda/n + o(1/n)$. В этом случае $P\left(S(n) = m\right) \to \frac{\lambda^m}{m!}\exp(-\lambda)$. Действительно:

$$P\left(S(n)=m\right) = \binom{n}{m} p^m(n)\left(1-p(n)\right)^{n-m} = \frac{n(n-1)\ldots(n-m+1)}{m!}\left(\frac{\lambda}{n}+o(1/n)\right)^m \left(1-\frac{\lambda}{n}-o(1/n)\right)^{n-m}$$

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

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

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.

7.6 Распределение Пуассона,его производящая функция и характеристики

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.

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 из формулы $p_0 + p_1 + \ldots + p_{k-1}$, где $p_i = \frac{(mp)^j}{j!} e^{-mp}$. 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.

7.7 △ Схема Бернулли и теоремы теории вероятностей

Рассматривается схема n независимых повторений эксперимента с двумя исходами, то есть соответствующее дискретное вероятностное пространство $\Omega, \mathcal{A}, P)$, с элементами $\omega = (a_1, a_2, \ldots a_n)$, где каждое $a_i = 0, 1$ в зависимости от исхода i-го эксперимента, примем, что вероятность 1 равна p, вероятность 0 равна $q = 1 - p$, вероятность $P(\omega) = p^{\sum a_i} q^{n - \sum a_i}$ Эту же конструкцию можно определить как набор из n i.i.d случайных величин-индикаторов $\xi_i = \mathbf{1}_{\{\xi_i > 0\}}$. Ясно, что частичные суммы $S_m = \xi_1 + \ldots + \xi_m$ дают набор зависимых и *по-разному* распределенных с.в. $\begin{array}{c|c|c|c|c|c} S_m & 0 & \ldots & k & \ldots & m \\ \hline & \binom{m}{0} q^m & \ldots & \binom{m}{k} p^k q^{m-k} & \ldots & \binom{m}{m} p^m \end{array}$ причем $E\left(\frac{S_k}{k}\right) = p$. Последние дроби уместно называть *частотами* в серии повторений Бернулли, разумеется частота — также случайная величина. Не приходится рассчитывать на то, что с увеличением n величины отклонения частот от параметра p (от вероятности «успеха» при *каждом* испытании) будут уменьшаться, например:

$$P\left(\frac{S_n}{n} = 1\right) = P(\xi_1 = 1) P(\xi_2 = 1) \ldots P(\xi_n = 1) = p^n$$

Однако можно зафиксировать величину ε, характеризующую отклонение, и оценить зависимость (от n) вероятности $P\left(|\frac{S_n}{n} - p| \geqslant \varepsilon\right)$ того, что отклонение будет не меньше ε, Как показывает неравенство Чебышева

$$P\left(|\frac{S_n}{n} - p| \geqslant \varepsilon\right) \leqslant \frac{D\left(\frac{S_n}{n}\right)}{\varepsilon^2} = \frac{pq}{n\varepsilon^2} = \frac{\text{const}}{n}$$

указанная вероятность действительно мала.

Хотелось бы по-простому сказать, что предел этой вероятности стремится к нулю с возрастанием n, но пока что у нас нет такого объекта, в котором бы рассматривалась бесконечная последовательность повторений: во всяком случае пока непонятно, как можно вычислять вероятности на пространстве бесконечного числа повторений.

△ Версия Закона Больших Чисел для схемы Бернулли

Строгая формулировка результата, известного как *Закон Больших Чисел Бернулли* пока что выглядит так:

Пусть $\Omega^{[n]}, \mathcal{A}^{[n]}, P^{[n]}$ — последовательность схем Бернулли таких как выше:

$$\Omega^{[n]} = \left\{\omega^{[n]} : (a_1^{[n]}, a_2^{[n]}, \ldots a_n^{[n]})\right\}, \quad a_i^{[n]} = 0, 1$$
$$P^{[n]}\left(\omega^{[n]}\right) = p^{\sum a_i^{[n]}} q^{n - \sum a_i^{[n]}}$$
$$S_k^{[n]}(\omega^{[n]}) = \xi_1^{[n]}(\omega^{[n]}) + \ldots + \xi_k^{[n]}(\omega^{[n]})$$

Тогда

$$P^{[n]}\Big(\omega^{[n]}:\left|\frac{S_n^{[n]}(\omega^{[n]})}{n}-p\right|\geqslant\varepsilon\Big)\underset{n\to\infty}{\longrightarrow}0$$

Чуть позже в разделе 9.6 мы рассмотрим так называемую конструкцию Радемахера для построения вероятностного пространства бесконечного числа повторений и формулировка ЗБЧ упростится, а пока что укажем, что идея и трудность фиквентистского подхода к теории вероятностей может быть понята на этом примере. А именно, в основе теории лежит рассмотрение бесконечных последовательностей, конечные фрагменты которых подчинены неравенствам указанного вида. Фриквентисты говорят, что по конкретной такой последовательности можно ввести понятие *шанса* «успеха» в одном эксперименте, численно этот шанс отвечает величине p. Неравенство тем самым является исходным пунктом для введения понятия вероятности, относимого однако к конкретным сериям повторений, а не к единичному эксперименту, аналогично, понятие независимости относится к последовательностям повторений. Неформально говоря, закон больших чисел для повторений Бернулли является теоремой, а у фриквентистов — аксиомой.

△ Энтропия как мера неопределенности

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

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

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 $P(\omega)\sim exp\,[n-H]$, 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 $p_1, p_2, \dots p_k$, 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.

△ 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 (мы ранее экспериментально воспроизводили этот эффект, называемый законом арксинуса), 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.

В чем реальный смысл закона больших чисел и какова его эмпирическая интерпретация. Пусть производится большое число, скажем, N, *серий* экспериментов, каждая из которых состоит из «независимых испытаний с вероятностью интересующего нас события-успеха равной p». Пусть $S_n^{[i]}/n$ — частота появления успеха в i-й серии и N_ε — число серий, в которых частоты отклоняются от p меньше чем на ε (N_ε равно числу тех i, вё которых частоты отклоняются менее, чем на ε) Из ЗБЧ заключаем, что отношение N_ε/N приближает вероятность, скажем,

$$P^{[1]}\Big(\Big|\frac{S_n^{[1]}}{n}-p\Big|\leqslant\varepsilon\Big)$$

однако насколько именно хорошо это приближение не очень понятно без использования меры на сериях. Позже в разделе 13.3 мы вернемся к подобным формулировкам теорем о *предельном поведении в схеме серий.*

Типичный вопрос для статистики состоит в оценке необходимой длины в серии, чтобы

$$P\Big(\Big|\frac{S_n}{n}-p\Big|\leqslant\varepsilon\Big)\geqslant 1-\alpha$$

и из ЗБЧ вытекает такая оценка: $n\geqslant(4\varepsilon^2\alpha)^{-1}$. Назовем серию повторений $(n.\varepsilon)$-*типичной* если соответствующее неравенство выполнено, ЗБЧ утверждает что при достаточно большом n вероятность множества типичных реализаций близка к единице. а вот сколько таких реализаций и каковы их вероятности — это уже зависит от параметра p и выражается функцией, которая называеся *энтропия*, см. предыдущий раздел.

7.8 ЗБЧ и оценки вероятностей в схеме Бернулли

В распределении Бернулли значение для $P\left(S(n)=m\right)$ при больших значениях n,m может быть получено исходя из асимптотической аппроксимации факториалов по формуле Стирлинга (см. **"Напоминание"**). При этом соответствующее вычисление имеет существенно разную точность; чтобы понять, где эта аппроксимация работает эффективно проведем следующую аналогию с ЗБЧ, утверждающего, что с точки зрения вероятности среднее значение индикаторов с ростом числа их независимых повторений n слабоотличимо от математического ожидания одного индикатора.

Неравенство Чебышёва давало оценку в терминах дисперсии для вероятности отклонений с.в. от своего среднего.

$$P\left(\Big|\frac{S_n}{n}-p\Big|\geqslant\varepsilon\right)\leqslant\frac{pq}{n\varepsilon}$$

По аналогии, поскольку $E\left(\frac{S_n}{n}-p\right)=0$, проследим в терминах вероятностей за отклонениями с.в. $\left|\frac{S_n}{n}-p\right|$ от нулевого значения. Поскольку

$$E\left[\Big(\frac{S_n}{n}-p\Big)^2\right]=\frac{pq}{n}\qquad\text{и значит}\qquad\Big|\frac{S_n}{n}-p\Big|\quad\text{удобно измерять в терминах}\quad\sqrt{\frac{pq}{n}}$$

Нас интересуют вероятности

$$P\left(\Big|\frac{S_n}{n}-p\Big|\leqslant\varepsilon\sqrt{\frac{pq}{n}}\right)=P\left(\Big|\frac{S_n-E(S_n)}{\sqrt{D(S_n)}}\Big|\leqslant\varepsilon\right)=\sum_{\left\{k:\,\frac{k-np}{\sqrt{npq}}\leqslant\varepsilon\right\}}\binom{n}{k}p^kq^{n-k}$$

Оценки этих вероятностей в схеме Бернулли допускают значительные обобщения для других случайных величин, соответствующие результаты известны как версии Центральной Предельной Теоремы.

△ Версия Центральной Предельной Теоремы

Изначальная формулировка представляет практический интерес, поскольку модель схемы повторений опыта с двумя исходами очень распространена и мы разберем на простейшем примере этапы доказательства. Избавиться от громоздких формул помогает небольшое изменение: будем рассматривать i.i.d. последовательность случайных величин $\eta_1,\eta_2,\ldots\eta_M$ с параметром $P(\eta_i=-1)=P(\eta_i=1)=\frac{1}{2}$, измененное по сравнению с S_k обозначение $\mathrm{S}_k=\eta_1+\ldots+\eta_k$ будет напоминать, что значения случайной величины S_k уже иные. Для вероятностей этих значений имеет место точная формула, посмотрим на ее аппроксимацию при помощи формулы Стирлинга $n!\sim n^ne^{-n}\sqrt{2\pi n}$ (см. в **"Напоминание"** идею доказательства формулы Стирлинга).

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

Упрощая,

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

Нам нужны два элементарных утверждения из математического анализа:

1. Если $c_j \to 0$, $a_j \to \infty$ и $a \cdot c_j \to \lambda = \mathrm{const}$, 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:

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

Это последнее утверждение называется (версией) локальной предельной теоремы.
Следующий шаг — аппроксимировать вероятность

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

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:

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

Подинтегральное выражение называется *плотностью стандартного гауссовского распределения*, а аппроксимационная формула известна как формула де Муавра.

△ Формулы Муавра-Лапласа

Формула Муавра была доказана для специального случая равных вероятностей в однократном испытании. Приведем теперь ее формулировки для i.i.d. последовательности $\{\xi_k\}$ испытаний с двумя исходами $\begin{array}{c|c|c} \xi_k & 0 & 1 \\ \hline & q & p \end{array}$ и $S_k = \xi_1+\ldots+\xi_k$: Локальная предельная теорема для распределения Бернулли: для $x \in \mathbb{R}$ таких, что $x = o(npq)^{1/6}$ и при этом $np + x\sqrt{npq}$ равно неотрицательному целому k, вероятность в схеме Бернулли

$$P\left(S_n = np + x\sqrt{npq}\right)\sim\frac{1}{\sqrt{2\pi npq}}e^{-x^2/2}$$

Интегральная формула Муавра-Лапласа:

$$P\left(a<\frac{S_n-np}{\sqrt{npq}}\leqslant b\right)\sim\frac{1}{\sqrt{2\pi}}\int_a^b e^{-x^2/2}dx$$

□ **Версия доказательства теоремы Вейерштрасса**

Теорема Вейерштрасса утверждает возможность равномерного приближения на отрезке непрерывной функции $f(p)$ полиномами. Пусть аргумент $p \in [0,1]$. Полиномом Бернштейна называется

$$B_n(p) = \sum_{k=0}^{n} f\left(k/n\right) \binom{n}{k} p^k (1-p)^{n-k}$$

Для i.i.d. последовательности бернуллиевских случайных величин $\xi_1, \xi_2, \dots \xi_n$ с параметром $P(\xi_i = 1) = p$ мат.ожидание $E\left(f(\mathrm{S}_n/n)\right) = B_n(p)$. Из равномерной непрерывности $\forall \varepsilon > 0\ \exists \delta > 0 : |x-y| \leqslant \delta \Longrightarrow |f(x)-f(y)| \leqslant \varepsilon$ и компактности отрезка $|f(p)| \leqslant M < \infty$ Применение ЗБЧ дает:

$$\begin{aligned}
|f(p) - B_n(p)| \quad &= \quad \left| \sum_{k=0}^{n} \Big[f(p) - f(k/n) \Big] \binom{n}{k} p^k (1-p)^{n-k} \right| \\
&\leqslant \sum_{\{|k/n-p| \leqslant \delta\}} \Big| f(p) - f(k/n) \Big| \binom{n}{k} p^k (1-p)^{n-k} + \sum_{\{|k/n-p| > \delta\}} \Big| f(p) - f(k/n) \Big| \binom{n}{k} p^k (1-p)^{n-k} \\
&\leqslant \varepsilon + 2M \sum_{\{|k/n-p| > \delta\}} \binom{n}{k} p^k (1-p)^{n-k} \leqslant \varepsilon + \frac{2M}{4n\delta^2} = \varepsilon + \frac{M}{2n\delta^2}
\end{aligned}$$

Таким образом, полиномы Бернштейна в аргументах $p \in [0,1]$ приближают функцию $f(p)$:

$$\lim_{n\to\infty} \max_{0 \leqslant p \leqslant 1} |f(p) - B_n(p)| = 0$$

8 Аксиоматический подход А.Н.Колмогорова и нестандартные теории вероятностей

Достаточно много содержательных примеров в традиционной дискретной теории вероятностей составили основу для естествознания. Недискретный случай (как и вообще рассмотрение математических примеров с использованием бесконечностей) требует значительных технических усилий, поэтому вначале основные понятия теории стандартно рассматриваются на интуитивно очевидных традиционных вероятностных моделях. Большинство ученых (но не все!) согласны с тем, что в теории должны рассматриваться семейства событий $\mathcal{A}$ и каждому событию из $\mathcal{A}$ по некоторому правилу сопоставлена его вероятность $0 \leqslant P(A) \leqslant 1$, называемая также вероятностной мерой. Однако структура таких семейств и свойства вероятностной меры остаются предметом обсуждения, несмотря на то, что почти столетие назад сформулирована аксиоматика, которая связывает понятия общей теории меры и структуры вероятностной модели — это так называемые аксиомы Колмогорова и соответствующая версия науки называется колмогоровской теорией вероятностей.

8.1 ○ Аксиомы колмогоровской теории вероятностей

В рамках аксиоматики Теории Вероятностей, созданной А.Н.Колмогоровым, при построении вероятностной модели следует рассматривать σ-алгебру событий $\mathcal{B}$ и σ-аддитивную вероятностную меру P на $\mathcal{B}$. Разумеется, в ранее рассматривавшихся дискретных случаях условие σ-аддитивности иногда неявно использовалось при рассмотрении, например событий из бесконечного числа исходов, но мы никак не задерживали на этом внимание полагаясь на «интуитивную очевидность» дискретных моделей. В недискретном случае требование σ-алгебры и σ-аддитивности формулируется явно — образуя систему аксиом. **Любое рассуждение в колмогоровской теории вероятностей подразумевает, что должна быть задана вероятностная модель** $(\Omega, \mathcal{B}, P)$ — **набор из пространства элементарных исходов, σ-алгебры событий $\mathcal{B}$ и заданной на событиях вероятностной σ-аддитивной меры P.** Эта тройка обычно называется структурой вероятностного пространства.

В конкретных моделях вероятностных пространств обычно понятно, из чего состоит множество Ω и какие именно объекты следует включить в алгебру событий Упомянутые в аксиомах понятия σ-алгебры событий $\mathcal{B}$ и σ-аддитивной вероятностной меры P на $\mathcal{B}$ в принципе знакомы из курса общей теории меры. По традиции в курсах теории вероятностей эти понятия пересказывают в большей или меньшей степени привлечения подробностей. Но в нашем курсе мы достаточно кратко обсудим лишь некоторые специфические свойства[17].

[17] в частности: как превратить алгебру $\mathcal{A}$ в σ-алгебру $\mathcal{B}$ и как продолжить введенную на алгебре $\mathcal{A}$ конечно-аддитивную меру до σ-аддитивной

Несложно проверить, что вероятностные пространства, удовлетворяющие аксиомам Колмогорова, существуют: основной здесь пример доставляет $(\Omega, \mathcal{B}, P)$, где $\Omega = (0,1) \subset \mathbb{R}$, $\mathcal{B}$ – σ-алгебра борелевских подмножеств, P мера Лебега на этой σ-алгебре. Небольшими модификациями из этого примера получается, например, модель бесконечной схемы Бернулли в разделе 9.6, а также примеры i.i.d последовательностей индикаторов в разделе 11.1.

Хотя в этом курсе мы стоим на позициях аксиоматического подхода А.Н.Колмогорова, но стоит хотя бы поверхностно ознакомиться с тем, как вероятность интерпретируется в иных подходах.

8.2 □ 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 рассматриваемого 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.

8.3 □ 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 $P(A)$, 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 объяснии принципиальной разницы с традиционной теорией вероятностей. Сумма событий $U \oplus V$ в квантовой теории вероятностей определяется как линейная оболочка объединения соответствующих подпространств, произведение $U \cdot V$ 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

$$A \cdot \left(\bigoplus_i B_i\right) = \bigoplus_i (A \cdot B_i) \qquad \text{и} \qquad A + \prod_i B_i = \prod_i (A + B_i)$$

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.

9 Конструкция вероятностных мер

9.1 ○ О связи с теорией меры

В курсе по общей теории меры вводятся базовые для Теории Вероятностей понятия алгебры и σ-алгебры, аддитивной меры и σ-аддитивной меры на алгебре, которую единственным образом можно продолжить на σ-алгебру, порожденную этой алгеброй. Формализм вероятностей, в рамках которого изучаются бесконечные модели, основан на этих понятиях, поэтому обычно эту часть теории меры включают в учебники теории вероятностей. Для матфака ВШЭ с одной стороны можно предполагать у слушателей знакомство с языком σ-алгебр, с другой — специфичные вероятностные задачи в вводном курсе используют достаточно ограниченную часть общей теории, придавая не слишком большое значение примерам патологического характера; в центре рассмотрения находятся связи теории меры и независимость событий.

В этом разделе основным объектом рассмотрения будут случайные величины и связанные с ними σ-алгебры, аналогичные конструкции для многомерных случайных векторов также имеют место и вполне аналогичны одномерному случаю, но изложение несколько более громоздко и мы его представим лишь тезисно.

Пусть $\xi = \xi(\omega)$ случайная величина — измеримая функция из вероятностного пространства $(\Omega, \mathcal{A}, P)$ в множество вещественных чисел $\mathbb{R}$, оснащенное борелевской σ-алгеброй $\mathcal{B}$. Несложно проверить, что множества A из $\mathcal{A}$ $A = \{\omega : \xi(\omega) \in B, B \in \mathcal{B}\}$ образуют σ-алгебру, которую обозначим $\sigma(\xi)$, подобная же конструкция имеет место для семейств (необязательно конечных) случайных величин, в частности, определена σ-алгебра $\sigma(\xi, \xi_2, \dots)$ для последовательности случайных величин.

Для борелевской функции $\phi : \mathbb{R} \to \mathbb{R}$ композиция $\phi \circ \xi$ является $\sigma(\xi)$-измеримой случайной величиной, обратное утверждение также справедливо.

Предложение 9.1 *Пусть η $\sigma(\xi)$-измеримая случайная величина, тогда найдется борелевская функция $\phi : \mathbb{R} \to \mathbb{R}$ такая, что $\eta = \phi \circ \xi$.*

Доказательство Пусть $A \in \sigma(\xi)$ проверим, что $\mathbf{1}_A$ представим в виде композиции. Действительно, если $A \in \sigma(\xi)$, то найдется $B \in \mathcal{B}(\mathbb{R})$ такое, что $A = \{\omega : \xi(\omega) \in B\}$ и поэтому $\mathbf{1}_A = \mathbf{1}_B \circ \xi$. Аналогично, любая простая $\sigma(\xi)$-измеримая случайная величина представима в виде композиции. Произвольную $\sigma(\xi)$-измеримую случайную величину η можно представить как предел последовательности $\sigma(\xi)$-измеримых простых, для каждого члена последовательности найдется борелевская функция ϕ_n и $\phi_n(\xi(\omega)) \to \eta(\omega)$, осталось разбраться с пределами последовательности $\phi_n(x)$ при разных $x \in \mathbb{R}$. Рассмотрим множество тех $x \in \mathbb{R}$, где предел существует, оно является борелевским (упражнение!), положим в точках этого множества функцию $\phi(x)$ совпадающей с соответствующим пределом, а в прочих точках x положим ее равной нулю. Теперь $\eta(\omega) = \lim_n \phi_n(\xi(\omega)) = \phi(\xi(\omega))$ ∎

9.2 ○ От вероятностной меры на прямой $\mathbb{R}$ к функции распределения

По вероятностной мере P на борелевской σ-алгебре подмножеств прямой $\mathbb{R}$ [18] определяем функцию $F_P(x) = P((-\infty, x])$, свойства достаточно очевидны из требований к непрерывности вероятностной меры:

1. $F_P(x)$ неубывающая
2. Пределы $F_P(x)$ при стремлении x к $\pm\infty$ существуют и равны соответственно $F(-\infty) = 0$, $F(+\infty) = 1$.
3. $F_P(x)$ имеет в каждой точке пределы слева и непрерывна справа

Перечисленные свойства это как раз свойства *функции распределения* на прямой $\mathbb{R}$, осталось лишь привести пример соответствующей случайной величины (поскольку функция распределения все-таки должна отвечать какой-то случайной величине). Но это как раз очень просто: надо взять тавтологическое отображение $\xi(\omega) = \omega$, $\omega \in \mathbb{R}$.

[18] Тем самым по структуре вероятностного пространства $(\mathbb{R}, \mathcal{B}, P)$

9.3 ∘ От функции распределения к вероятностной мере на $\mathbb{R}$

Если задана функция распределения $F(x)$, то на борелевской алгебре $\mathcal{B}$ существует и единственна вероятностная мера P_F со свойством $P_F\left((a,b]\right) = F(b) - F(a)$.

Указанная формула определяет конечно-аддитивную меру на порожденной полуинтервалами *алгебре* $\mathcal{A}$, ограниченное множество тогда представляется как объединение конечного числа полуинтервалов, но в принципе можно его выразить и через счетное их объединение и тогда можно сравнить его меру с формальным вычисленым ответом через сумму ряда. Если два значения всегда совпадают, то соответствующая мера называется σ-аддитивной и для нее справедлива общая теорема Каратеодори из теории меры. Эта теорема утверждает, что можно продолжить меру P_F с алгебры $\mathcal{A}$ на минимальную σ-алгебру $\mathcal{B}$ ею порожденную, причем продолжение единственно. Мы не станем доказывать здесь теорему Каратеодори, ограничимся, изложением идеи доказательства в рассматриваемом случае вероятностной меры на алгебре $\mathcal{A}$.

Пусть B множество из $\mathcal{B}$, рассматриваем (счетные) покрытия множествами A_k из $\mathcal{A}$ и точную нижнюю грань меры таких покрытий (поскольку мера P на $\mathcal{A}$ σ-аддитивная) – эта точная нижняя грань называется *внешней* мерой $P^*(B)$. Переходя к дополнению множества $\bar{B}$, рассматриваем также *внутреннюю* меру $P_*(B) = 1 - P^*(\bar{B})$, семейство тех множеств B для которых $P_*(B) = P^*(B)$ образует σ-алгебру $\mathcal{C} \subset \mathcal{B}$, которая содержит $\mathcal{A}$ и совпадает с $\mathcal{B}$ в силу минимальности $\mathcal{B}$.

Из этого рассуждения, между прочим, вытекает,

Замечание Для любого $\varepsilon > 0$ и любого B множества из $\mathcal{B}$ найдется множество A в $\mathcal{A}$ такое, что $P\left(A \triangle B\right) < \varepsilon$.

Осталось проверить условие применимости теоремы Каратеодори, то есть надо убедиться в σ-аддитивности меры на алгебре $\mathcal{A}$, то есть проверить предельные выражения для счетного объединения полуинтервалов. Существует ряд эквивалентных формулировок свойства σ-аддитивности и в нашем случае рассуждение легче всего получить исходя из двойственного свойства пересечений в алгебре $\mathcal{A}$, а именно что для последовательности множеств $A_n \in \mathcal{A}$ с $A_n \downarrow \emptyset$ выполняется $\lim P_F(A_n) = 0$. Сначала рассмотрим случай, когда все множества $\ldots \supseteq A_n \supseteq A_{n+1} \supseteq \ldots$ в последовательности ограничены фиксированным отрезком $[-C, C]$, тогда, поскольку каждое из A_n состоит из конечного числа интервалов, по свойству правой непрерывности функции F для произвольного $\varepsilon > 0$ в последовательности можно выбрать *замкнутые* подмножества $B_n \subset A_n$ мера $P_F(B_n)$ которого отличается от $P_F(A_n)$ менее, чем на $2^{-n-1}\varepsilon$. Поскольку $A_n \downarrow \emptyset$ и $B_n \subset A_n$, то $\bigcap_{n=1}^{\infty} B_n = \emptyset$, а это означает, что мы имеем систему открытых множеств (дополнений к B_n) покрывающих компакт $[-C, C]$. У этого покрытия есть конечное подпокрытие, что означает, что для некоторого номера m выполнено $\bigcap_{m=1}^{\infty} B_m = \emptyset$. Отсюда несложно[19] оценить меру пересечения $\bigcap_{m=1}^{\infty} A_m$ — она окажется не превосходящей ε, поскольку множества в последовательности $A_n \in \mathcal{A}$ вложены друг в друга, то и для всех n $P_F(A_n) < \varepsilon$. Общий случай сводится к рассмотренному выбором подходящего значения C такого, что $P_F\left([-C, C]\right) > 1 - \varepsilon/2$.

9.4 △ Классификация случайных величин

В математическом анализе — интересующиеся отсылаются к книге И.Натансон "Теория функций вещественной переменной"Глава VIII §1, стр 191 — не слишком сложно установить, что монотонные функции обязательно должны в точке разрыва иметь и предел справа и предел слева, когда эти пределы совпадают – функция называется непрерывной. В соответствии с этим, если функция распределения непрерывна, то и соответствующая случайная величина (и вероятностная мера на $\mathbb{R}$) называется непрерывной. В Главе 6 книги А.Колмогоров и С.Фомин ”Элементы теории функций и функционального анализа” объяснена следующая классификация всех функций распределения: любая функция распределения $F_\xi(x)$ представима в виде линейной комбинации с неотрицательными коэффициентами $a_1 + a_2 + a_3 = 1$ трех функций распределения: кусочно-постоянной, которая тем самым является функцией распределения $F_1(x)$ для дискретной случайной величины; непрерывной функции распределения $F_2(x)$, допускающей представление в виде интеграла

$$F_2(x) = \int\limits_{-\infty}^{x} f(t)dt$$

— соответствующая вероятностная мера на $\mathbb{R}$ называется при этом абсолютно непрерывной, а соответствующая функция $f(x)$ – называется плотностью распределения); функции распределения $F_3(x)$, которая также непрерывна и даже имеет нулевую производную в точках некоторого борелевского множества W, причем мера Бореля (то есть та, которая на борелевской σ- алгебре порождена длиной полуинтервалов) дополнения $\mathbb{R} \backslash W$ равна нулю.

[19] Проведите это вычисление самостоятельно!

Это третье слагаемое и мера ему отвечающая называется (непрерывной) сингулярной мерой. Соответствующая сингулярной мере случайная величина также называется сингулярной.

Сингулярные меры и распределения в вводных курсах теории вероятностей обычно подробно не рассматривают, хотя они вполне могут возникать при изучении свойств последовательностей случайных величин. Для общего представления о сингулярной функции распределения вспомним об известной из курса функционального анализа «канторовой лестнице» (в англоязычной литературе эта конструкция известна под названием Devil's case). Канторово множества, как известно, есть дополнение в отрезке $[0,1]$ к объединению интервалов: центрального $(1/3, 2/3)$, двух центральных $(1/9)$ и $(7/9, 8/9)$ в оставшихся левой и правой частях и т п. Нетрудно видеть, что совокупная длина всех интервалов равна единице:

$$\frac{1}{3}+\frac{2}{3^2}+\frac{2^2}{3^3}+\ldots=\frac{1}{3}\cdot\left(1+\frac{2}{3}+\frac{2^2}{3^2}+\ldots\right)=\frac{1}{3}\cdot\frac{1}{1-\frac{2}{3}}=1$$

Канторова лестница — монотонное отображение отрезка $[0,1]\to[0,1]$, принимающее на интервале $(1/3, 2/3)$ постоянное значение $\frac{1}{2}$, на $(1/9)$ и $(7/9, 8/9)$ соответственно $\frac{1}{4}$ и $\frac{3}{4}$ и.т.д; на оставшиеся аргументы оно продолжается условием непрерывности. Дополняя это отображение нулем и единицей вне отрезка $[0,1]$ получим непрерывную монотонную неубывающую функцию $[0,1]\to\mathbb{R}$, которая и задает сингулярную меру на прямой, а значит и непрерывную случайную величину ν, которая также называется сингулярной. График функции распределения $F_\nu(x)$ показан на графике Рис.6.

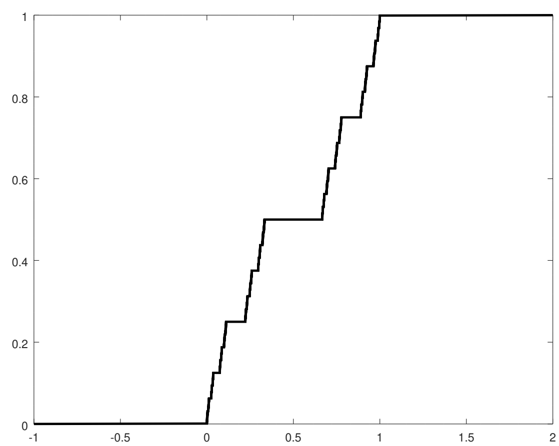

Рис. 6: Функция распределения $F_\nu(x)$ непрерывной сингулярной случайная величины ν.

9.5 △От случайных величин к случайным векторам

Речь здесь пойдет о функции распределения $F_{\vec{\xi}}(x)$ —- аналоге изложенной выше конструкции.

Вероятностное пространство $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), P)$

Обозначим

$$F_n(x_1,x_2,\ldots,x_n)=P\left((-\infty,x_1]\times\ldots(-\infty,x_n]\right)$$
$$\Delta_{a_ib_i}F_n(x_1,x_2,\ldots,x_n)=F_n(x_1,x_2,\ldots,x_{i-1},b_i,x_{i+1},\ldots,x_n)-F_n(x_1,x_2,\ldots,x_{i-1},a_i,x_{i+1},\ldots,x_n)$$

Итерация разностных операторов дает

$$\Delta_{a_1b_2}\ldots\Delta_{a_nb_n}F_n(x_1,x_2,\ldots,x_n)=P\left((a_1,b_1]\times\ldots(a_n,b_n]\right)\geqslant 0$$

По свойству непрерывности вероятности $F_n(\vec{x})=F_n(x_1,x_2,\ldots,x_n)$ имеет в каждой точке пределы слева и непрерывна справа, кроме того $\forall k\ \lim_{x_k\to-\infty}F_n(x_1,x_2,\ldots,x_n)=0$, а также $F_n(+\infty,+\infty,\ldots,+\infty)=1$. Любая функция с такими свойствами называется функцией распределения.

Справедлив аналог рассмотренного выше свойства, что по функции распределения можно единственным образом ввести вероятностную меру P_F такую, что

$$\Delta_{a_1b_2}\ldots\Delta_{a_nb_n}F_n(x_1,x_2,\ldots,x_n)=P_F\left((a_1,b_1]\times\ldots(a_n,b_n]\right)\geqslant 0$$

△ Последовательности независимых случайных величин. Закон «нуля и единицы»

Последовательность $\xi_1, \xi_2, \xi_3 \dots$ независимых случайных величин (необязательно одинаково распределенных) определяет события, которые относятся к поведению случайных величин «в бесконечности», а именно принадлежат $\mathcal{K} = \bigcap_{m=1}^{\infty} \mathcal{D}_m^{\infty}$ пересечению σ-алгебр $\mathcal{D}_m^{\infty} = \sigma(\xi_m, \xi_{m+1,\dots})$ (легко видеть, что пересечение[20] вложенных σ-алгебр является σ-алгеброй), которые отвечают подпоследовательностям вида $\xi_m, \xi_{m+1}, \dots$. Оказывается, что требование независимости случайных величин в последовательности влечет очень сильное ограничение на вероятности событий из σ-алгебры $\mathcal{K}$. При условии независимости всякое событие A и $\mathcal{K}$ не зависит от значений случайных величин $\xi_1, \xi_2, \dots \xi_m$ при любом конечном числе m, а определяется лишь поведением «бесконечно далеких» значений последовательности $\xi_1, \xi_2, \xi_3 \dots$, теорема Колмогорова утверждает, что вероятность такого события A из $\mathcal{K}$ удовлетворяет $P(A) = P(A)P(A)$, то есть его вероятность равна нулю или единице. В частности любая $\mathcal{K}$-измеримая случайная величина должна быть неотличима от константы.

Доказательство Если $A \in \mathcal{K}$, то это событие лежит в минимальной σ-алгебре $\sigma\{\xi_1, \xi_2 \dots\}$ всей последовательности, можно считать, что $\sigma\{\xi_1, \xi_2 \dots\}$ порождена событиями из объединения $\bigcup_n \sigma\{\xi_1, \xi_2 \dots \xi_n\}$. Поэтому найдутся такие события $A_n \in \sigma\{\xi_1, \xi_2 \dots \xi_n\}$ такие, что $P(A_n \triangle A) \to 0$ при $n \to \infty$. Отсюда ясно, что $P(A_n) \to P(A)$, а также $P(A_n \cap A) \to P(A)$. При этом события A_n и A независимы, так что при $n \to \infty$ одновременно

$$P(A_n \cap A) \to P(A) \quad \text{и} \quad P(A \cap A_n) = P(A)P(A_n) \to P(A)P(A)$$

∎

Пример

Рассмотрим последовательность независимых *одинаково распределенных* случайных величин $\xi_1, \xi_2, \xi_3 \dots$ таких, что каждая принимает значения ± 1 с равными вероятностями. Что можно сказать о вероятности сходимости (к конечным значениям) ряда $\sum_n^{\infty} \frac{\xi_n}{n}$?

Несложно видеть, что событие $A = \left\{\omega : \sum_n^{\infty} \frac{\xi_n}{n} \text{сходится}\right\}$ определяется именно поведением «бесконечно далеких» членов исходной последовательности $\{\xi_n\}$ и потому вероятность события A равна либо нулю, либо единице.

9.6 ○ Конструкция Радемахера для «бесконечной схемы Бернулли». Пример σ-алгебры

Рассмотрим независимые повторения эксперимента с двумя событиями $\{0\}$ и $\{1\}$. Богатый запас практически значимых утверждений теории вероятностей может быть проинтерпретирован на сравнительно несложном примере (конечного или бесконечного) независимого повторения эксперимента с двумя результатами-событиями $\{0\}$ и $\{1\}$, как и прежде этот эксперимент традиционно называют *бросанием несимметричной монеты.* Но в этом разделе мы более не ограничиваемся конечным числом повторений. Ранее пояснялось, что множество Ω исходов в случае *независимого повторения* должно быть произведением множеств исходов в единичном эксперименте.Мы уже рассматривали случай конечного n- кратного повторения эксперимента с двумя исходами: алгебра событий в этом случае состояла из любых подмножеств в $\Omega = \prod_1^n \Upsilon_i$. Естественно рассмотреть аналогичную конструкцию В бесконечном случае, здесь будут важные отличия, это даст понимание зачем нужна концепция σ-алгебры.

В соответствии с теоретико-множественным пониманием произведения элементами ω бесконечного произведения $\Omega = \prod_1^{\infty} \Upsilon_i$ являются бесконечные последовательности элементов из Υ, то есть последовательности из нулей и единиц, осталось определить σ-алгебру событий. Разумеется, Ω и $\varnothing$ следует считать событиями. В соответствии с топологическим подходом к событиям в $\prod_1^{\infty} \Upsilon_i$ следует причислить прообразы событий $\{0\}$ и $\{1\}$ при проекциях $\pi_k : \prod_1^{\infty} \Upsilon_i \to \Upsilon_k$, а также их всевозможные *конечные* пересечения, объединения и дополнения. Таким образом возникает алгебра $\mathcal{A}$ событий на Ω: система подмножеств, замкнутая относительно конечных объединений,пересечений и дополнений, содержащая кроме того пустое множество $\varnothing$ и дополнение к нему Ω. Каждый элемент в Ω отождествлен с последовательностью из нулей и единиц $\omega_1\omega_2\omega_3 \dots$ Следующие важные вопросы демонстрируют некоторую ограниченность алгебры событий $\mathcal{A}$, порожденной прообразами:

- Пусть $\omega \in \Omega$, Входит ли событие $\{\omega\}$ в $\mathcal{A}$?
- Является ли $\mathcal{A}$ σ-алгеброй?

[20]Здесь важно не забывать, что для бесконечного объединения вложенных σ-алгебр аналогичное утверждение неверно!

- Верно ли, что пересечение двух σ-алгебр является σ-алгеброй?

Таким образом для того, чтобы включить в рассмотрение одноточечные события $\{\omega\}$ необходим переход от алгебры событий $\mathcal{A}$ к порождаемой ею σ-алгебре, которая имеет прямое отношение к а σ-алгебре борелевских подмножеств на отрезке $[0, 1]$.

Действительно, геометрический смысл устройства множества Ω можно усмотреть опять-таки из отождествления с последовательностями из нулей и единиц, каждой такую последовательности, вообще говоря, можно сопоставить ряд — «двоичное разложение действительного числа» $\sum_{k=1}^{\infty} \frac{\omega_k}{2^k}$ из отрезка $[0, 1]$, тем самым получая некоторое наглядное представление о строении Ω.

- Упражнение: верно ли, что при таком соответствии множество Ω совпадает с $[0, 1]$? Ответ: нет.

Таким образом Ω совпадает с объединением отрезка $[0, 1]$ и счетного множестве точек, а прообразы событий при проекциях порождают борелевскую σ-алгебру на отрезке.

Осталось понять, как устроена вероятностная мера: в общем случае независимых повторений устройство ее нетривиально и содержит сингулярности. Но когда в единичном эксперименте исходы $\{0\}$ и $\{1\}$ равновероятны, эта вероятностная мера в точности совпадает с борелевской мерой на отрезке, счетное множество при этом имеет меру ноль. Для этого случая легко предъявить и последовательность i.i.d. случайных величин-индикаторов $\varepsilon_k = \mathbf{1}_{\{\varepsilon_k=1\}}$, определенных для k-го повторения эксперимента. Более того, члены последовательности в этом случае могут быть представлены (с точностью до значений на событии меры ноль) своими графиками!

- Упражнение: нарисуйте график с.в. совпадающей с произведением $\varepsilon_1\varepsilon_3$.
- Упражнение: Зависимы ли случайные величины $\varepsilon_1\varepsilon_3$ и $\varepsilon_1\varepsilon_4$?
- Упражнение: нарисуйте функцию распределения $F_{\varepsilon_1\varepsilon_3}$?

$\triangle$ Усиленный Закон Больших Чисел в схеме Бернулли

Для модели с бесконечным повторением эксперимента с двумя исходами имеем i.i.d. последовательность случайных величин $\{\xi_k\}$ таких, что $\begin{array}{c|c|c} \xi_k & 0 & 1 \\ \hline & q & p \end{array}$, а также последовательность их сумм $S_k = \xi_1 + \ldots + \xi_k$ – т.е. числе успехов в первых k испытаниях. Наглядное представление о вероятности основывается на интуитивном представлении о правильности следующего утверждения: «S_n в пределе по n дает значение p». В теории это никак не может быть верным для *каждой* последовательности испытаний, однако можно доказать, что утверждение выполняется с вероятностью единица; поэтому случаи, когда оно неверно, являются исключением, которым можно пренебречь. Упоминавшаяся в разделе 8.2 фриквентистская версия теории вероятностей как раз использует этот принцип при определении самого понятия вероятности. Сформулируем теперь (версию) Усиленного закона Больших Чисел для указанной ситуации (в принципе, это следствие так называемой Леммы Бореля-Кантелли, которая в наш курс не вошла).

Теорема 9.2 *Для любого $\varepsilon > 0$ вероятность того, что осуществится только конечное число событий*

$$\left|\frac{S_n}{n} - p\right| > \varepsilon$$

равна 1.

Несмотря на сходство этой формулировки с законом «нуля и единицы» утверждение к нему не сводится, поскольку последовательность $\{S_m\}$ не является i.i.d.

Лемма 9.3 *Пусть $A_1, A_2, \ldots$ —бесконечная последовательность - событий в указанной схеме бесконечных повторений, каждое событие из которых определено с помощью только конечного числа испытаний и $P(A_k) = a_k$. Если ряд $\sum_k a_k$ сходится, то с вероятностью единица произойдет*[21] *лишь конечное число событий A_k*

Доказательство Начнем с доказательства леммы. Для произвольного $\delta > 0$ выберем r таким, чтобы $a_{r+1} + a_{r+2} + \ldots < \delta$, это возможно из-за сходимости ряда. Лемма утверждает, что вероятность произойти $N > 1$ событиям из $A_{r+1} \ldots$ меньше δ. Это легко видеть из формулы включений и исключений поскольку

$$P(A_{r+1} \cup \ldots \cup A_{N)} \leqslant a_{r+1} + a_{r+2} + \ldots < \delta$$

[21] Переформулируйте неформальное выражение «произойдет» более строго

Переходим к доказательству самой теоремы. Обозначим для $a > 2$ через A_k событие

$$\left|\frac{S_k - kp}{\sqrt{kpq}}\right| \geqslant \sqrt{2a \ln k}$$

его вероятность можно асимптотически точно оценить по формуле Муавра-Лапласа (или с помощью формулы оценки вероятности больших отклонений при случайном блуждании, см. 5.5) в окажется, что $P(A_k) \sim \frac{1}{k^a}$, то есть имеем сходящийся ряд из вероятностей и условие Леммы выполнено. ∎

□ Особенности использования языка σ-алгебр

Описание бесконечных подмножеств и их семейств связано с аксиоматикой теории множеств, как известно, согласия в том, что следует считать «общеобязательной» системой аксиом среди математиков нет и не предвидится. В первой половине XX века имело место увлечение патологическими примерами бесконечных множеств, которые в зависимости от системы аксиом могли бы считаться «существующими», и достаточно быстро была отмечена роль аксиомы выбора в построении подобных примеров. Однако уже стандартный математический анализ без аксиомы выбора выглядит в высшей степени странно: например, придется говорить о подмножествах отрезка в которых не существует счетного плотного подмножества.

В теории вероятностей, конструкции борелевских множеств, σ-алгебр и соответствующих мер включают интуитивно неочевидные для прикладных интерпретаций утверждения также как правило связанные с аксиомой выбора, мы приведем здесь (только для информации, но можно воспринимать и как задачи для самостоятельного решения) некоторые из них.

1. Ранее (см. 9.6) было предложено описание вероятностного пространства $\{0,1\}^\infty$ для бесконечных повторений равновероятных исходов, с точностью до меры ноль совпадающее с отрезком $[0,1]$. В континууме элементарных исходов можно ввести отношение эквивалентности, отождествляющее последовательности различающиеся не более, чем в конечном множестве позиций. Выбирая по эдементу в каждом классе эквивалентности по одному элементу мы получаем пример неборелевского множества (пример неконструктивный, так как опирается на аксиому выбора).

2. σ-алгебра называется *счетно-порожденной*, если она порождается некоторым счетным классом подмножеств. Например σ-алгебра $\mathcal{B}$ борелевских подмножеств в $(0,1]$ является счетно-порожденной полуинтервалами с концами в рациональных точках. Однако, например, σ-алгебра, состоящая из счетных подмножеств из $\mathcal{B}$ конечно-порожденной уже не является.

3. Объединение вложенных друг в друга σ-алгебр является алгеброй, но *вовсе не обязано* быть σ-алгеброй.

4. Объединение *строго* вложенных друг в друга алгебр заведомо не является σ-алгеброй.

10 Случайные величины в общем случае

Напомним, что за отдельными примерами, мы выносим за рамки курса сингулярные случайные величины и остаемся в рамках семейства, содержащего абсолютно непрерывные с.в. и смешанные с.в. —- последнее означает, что в функция распределения $F_\xi(x)$ представима в виде линейной комбинации с неотрицательными коэффициентами $a_1 + a_2 = 1$ двух функций распределения: кусочно-постоянной, которая тем самым является функцией распределения дискретной случайной величины и непрерывной функции распределения $F_2(x)$, допускающей представление в виде интеграла

$$F_2(x) = \int\limits_{-\infty}^{x} f(t)dt$$

Ранее были изложены основные понятия и конструкции для дискретных случайных величин: в условиях аксиоматики Колмогорова случайными величинами называются измеримые по Борелю числовые функции на Ω. Рассмотренные ранее интуитивно ясные понятия в недискретном варианте иногда требуют уточнений, впрочем в большинстве случаев доказательства утверждений могут быть использованы без изменений. Базовая идея здесь опирается на рассмотрения последовательных аппроксимаций общей случайной величины [22]. В некотором виде эта идея использовалась уже в определении математического ожидания дискретной случайной величины с бесконечным числом значений.

[22] то есть измеримой функции $\xi : \Omega \to \mathbb{R}$) с помощью дискретных с.в., принимающих конечное число значений. Такие случайные величины называются *простыми*

10.1 ∘ Математическое ожидание

Для принимающих конечное множество значений дискретных случайных величин введенное ранее понятие математического ожидания есть не что иное как интеграл в смысле Лебега для измеримой функции $\Omega \to \mathbb{R}$, В общем случае математическое ожидание случайной величины вводится именно как интеграл по Лебегу $\int\limits_{\Omega} \xi dP$ от измеримой по Борелю функции $\xi(\omega)$:

шаг 1 Пусть сначала $\xi(\omega)$ неотрицательна. Интегралом Лебега от неотрицательной случайной величины $\xi(\omega)$, или в нашем случае математическим ожиданием, называется, предел[23] $\lim_n E\xi_n$ математических ожиданий $E\xi_n$ простых случайных величин ξ_n из аппроксимирующей последовательности $\xi_n \uparrow \xi$ (т.е. все $\xi_n \leqslant \xi$ и $\lim_n \xi_n(\omega) = \xi(\omega)$). Корректность определения, то есть отсутствие зависимости значения предела от выбора аппроксимирующей последовательности доказывается в теории меры[24].

шаг 2 Для с.в. ξ имеем две неотрицательные с.в. $\xi^+ = \max(\xi, 0)$ и $\xi^- = \min(\xi, 0)$, если у обеих определено математическое ожидание, то полагаем $E\xi = E\xi^+ - E\xi^-$

Свойства математического ожидания вытекают из свойств интеграла Лебега, в частности, указанные в разделе 5.1 свойства мат.ожиданий для простых случайных величин, сохраняются, приведем некоторые, ранее пропущенные (или сформулированные в виде задач):

1. Математическое ожидание $E\xi\eta$ произведения независимых случайных величин ξ η равно произведению $E\xi E\eta$, если сомножители определены.

2. Если математическое ожидание $E\xi$ определено , то для любого события $A \subset \Omega$ определено мат.ожидание $E\xi \mathbf{1}_A$.

3. Пусть $\xi \geqslant 0$ и $E\xi = 0$, тогда $\xi = 0$ всюду, кроме может быть события нулевой вероятности.

Мы обсуждали, что с.в. с функцией распределения $F_\xi(x)$ можно понимать как тавтологическое отображение $\xi : \mathbb{R} \to \mathbb{R}$, где вещественная прямая превращена в вероятностное пространство с борелевской σ-алгеброй и мерой, определенной при помощи $F_\xi(x)$. Это объясняет обозначение $E\xi = \int_{\mathbb{R}} x dF_\xi$ для мат.ожидания. Для абсолюно непрерывных с.в определена **плотность** $f_\xi(x)$ **распределения** и $F_\xi(b) - F_\xi(a) = \int_{(a,b]} f_\xi(x)dx$, что позволяет в этом случае *вычислять* $E\xi = \int_{\mathbb{R}} x dF_\xi$ как несобственный интеграл Римана.

$$E(\xi) = \int\limits_{-\infty}^{\infty} x f_\xi(x) dx$$

Заметим, что последняя формула легко объясняет, почему моменты с.в. определены не всегда: действительно, не все несобственные интегралы сходятся.

В дальнейшем встретится следующее свойство[25] математического ожидания: для борелевской функции $g : \mathbb{R} \to \mathbb{R}$ и случайной величины ξ с плотностью $f_\xi(x)$ математическое ожидание случайной величины $g(\xi)$ равно

$$E\left(g(\xi)\right) = \int\limits_{\mathbb{R}} x dF_{g(\xi)}(x) = \int\limits_{\mathbb{R}} g(x) dF_\xi$$

Для простых случайных величин это вполне очевидно, предельный переход дает удобную формулу для абсолютно непрерывных случайных величин:

$$E\left(g(\xi)\right) = \int\limits_{\mathbb{R}} g(x) f_\xi(x) dx$$

[23]В общем определении интеграла Лебега допускают значение $+\infty$ для предела, в теории вероятностей это приводит к тому, что надо рассматривать бесконечные математические ожидания. *Для простоты в нашем курсе* участвуют только конечные мат.ожидания, что в дискретном случае опиралось на требование абсолютной сходимости ряда $\sum a_i p_i$.

[24]явно такую последовательность можно выбрать, например, из индикаторов так:

$$\xi_n(\omega) = \sum_{k=1}^{n2^n} \frac{k-1}{2^n} \mathbf{1}_{\{\frac{k-1}{2^n} \leqslant \xi(\omega) < \frac{k}{2^n}\}}(\omega) + n\mathbf{1}_{\{\xi(\omega) \geqslant n\}}(\omega)$$

[25]Так называемaz «формула ленивого статистика»

Пример: почему усредняют измерения?

Вернемся к уже рассмотренному нами неравенству Чебышёва и докажем его еще раз для имеющей $D\xi$ плотность распределения $f_\xi(x)$ случайной величины ξ, воспользовавшись приведенным выше правилом

$$E\left(g(\xi)\right) = \int\limits_{\mathbb{R}} g(x) f_\xi(x) dx$$

Действительно:

$$E\left((\xi - E(\xi))^2\right) = \int\limits_{\mathbb{R}} (x - E(\xi))^2 f_\xi(x) dx \geqslant \int\limits_{|x-E(\xi)|\geqslant\varepsilon} (x - E(\xi))^2 f_\xi(x) dx \geqslant \int\limits_{|x-E(\xi)|\geqslant\varepsilon} \varepsilon^2 f_\xi(x) dx \geqslant \varepsilon^2 \int\limits_{|x-E(\xi)|\geqslant\varepsilon} f_\xi(x) dx = \varepsilon^2 P\left(|x - E(\xi)| \geqslant \varepsilon\right)$$

или как обычно $P\left\{|\xi - E(\xi)| \geqslant \varepsilon\right\} \leqslant \frac{D(\xi)}{\varepsilon^2}$. Если $\xi = (\eta_1 + \eta_2 + \dots \eta_n)/n$ где все η_k независимы между собой, то по ЗБЧ в форме Чебышева получается $P\left\{|\xi - E(\xi)| \geqslant \varepsilon\right\} \leqslant \frac{D(\eta)}{n\varepsilon^2}$ и, значит, при больших n интеграл

$$\int_{E(\xi)-\varepsilon}^{E(\xi)+\varepsilon} f_\xi(x) dx \to 1$$

Последнее означает, что весь носитель функции плотности $f_\xi(x)$ с ростом n сосредотачивается в сколь угодно малом интервале вокруг $E(\xi)$, то есть усреднение измерений η_k дает случайную величину, мало отличающуюся от константы $E(\xi)$. Неформально выражаясь, «усреднение гораздо более стабильно, чем отдельные измерения»

Пример: «правило трех сигм»

. Если в неравенстве Чебышёва положить $\varepsilon = 3\sigma = 3\sqrt{D(\xi)}$ 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.

10.2 △ Неравенства

Для начала упомянем[26] теорему о неравенствах, относящуюся к свойствам интеграла Лебега последовательностей функций, но только сформулируем ее в терминах случайных величин и их математических ожиданий.

Теорема 10.1 (О мажорируемой сходимости) *Пусть с.в. $\eta, \xi_1, \xi_2, \dots$ таковы, что $|\xi_n| < \eta$, $E\eta$ существует, последовательность $\{\xi_n(\omega)\}$ сходится к с.в. ξ при всех $\omega \in \Omega$, кроме может быть множества меры ноль*[27]*. Тогда $E\xi$ существует, причем $E\xi_n \to E\xi$ и $E\left|\xi_n - \xi\right| \to 0$*

Следствие 10.2 *Пусть с.в. $\eta, \xi_1, \xi_2, \dots$ таковы, что $|\xi_n| < \eta$, при некотором $m > 0$ $E\eta^m$ существует, последовательность $\{\xi_n(\omega)\}$ сходится к с.в. ξ при всех $\omega \in \Omega$, кроме может быть множества меры ноль Тогда $E\xi^m$ существует , причем $E\xi_n \to E\xi$ и $E\left|\xi_n - \xi\right|^m \to 0$*

Продолжим аналогии с неравенствами из теории интеграла Лебега.

Неравенство Коши-Буняковского

Пусть случайные величины ξ и η таковы, что $0 < E\xi^2 < \infty$ и $0 < E\eta^2 < \infty$. Тогда $E|\xi\eta| < \infty$ и

$$\Big(E|\xi\eta|\Big)^2 \leqslant E\xi^2 E\eta^2$$

Доказательство Перейдем к с.в. $\tilde{\xi} = \frac{\xi}{\sqrt{Ex\xi^2}}$ и $\tilde{\eta} = \frac{\eta}{\sqrt{Ex\eta^2}}$ имеем в силу обычного неравенства $2|xy| \leqslant x^2 + y^2$ для $x, y \in \mathbb{R}$

$$2E|\tilde{\xi}\tilde{\eta}| \leqslant E\tilde{\xi}^2 + E\tilde{\eta}^2 = 2$$
$$E|\xi\eta| \leqslant 1$$

Неравенство Коши-Буняковского когда $E\xi^2 = 0$ или $Ex\xi^2 = 0$, тривиально в силу свойства 3, приведенного выше.

[26]Напомним, что существование мат.ожидания в нашем курсе означает, что оно $< \infty$
[27]Это так называемая *сходимость почти всюду*, см далее 11.1

Неравенство Иенсена

Пусть $g = g(x)$ — выпуклая вниз измеримая по Борелю числовая функция и $E\xi < \infty$, тогда $g\left(E\xi\right) \leqslant E\left[g(\xi)\right]$

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 $x_0 = E\xi\ x = \xi$ имеем

$$g(\xi) \geqslant g(E\xi) + (\xi - E\xi)\lambda(E\xi)$$

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

Полезное следствие неравенства Иенсена: для $0 < s < t$ выполнено

$$\left(E|\xi|^s\right)^{1/s} \leqslant \left(E|\xi|^t\right)^{1/t}$$

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

Наконец заметим, что неравенства Маркова, Чебышева, Чернова и Хёфдинга, появившиеся в связи с дискретными с.в. также верны, причем их доказательства даже не требуют коррекции.

10.3 □ Условное математическое ожидание в общем случае

Ранее для простых с.в. была изложена конструкция условного математического ожидания. Существование условного математического ожидания $\mathcal{E}(\eta|\xi)$ для пары общих случайных величин обычно объясняют при помощи теоремы Радона-Никодима (также теоремы существования) в теории меры. В полной общности это понятие требует рассмотрения случайных величин со значениями в $\pm\infty$ и поэтому мы его лишь коснемся для общей информации.

Пусть $\Omega, \mathcal{B}, P$ — вероятностное пространство и задана σ-подалгебра $\mathcal{A}$ в $\mathcal{B}$ такая, что некоторая случайная величина ξ не обязательно измерима относительно $\mathcal{A}$. В теории вероятностей рассматривают в основном случай, когда $\mathcal{A}$ порождается прообразами (борелевских множеств $\mathbb{R}$) одной или нескольких случайных величин.

шаг 1 Условным математическим ожиданием $\mathcal{E}(\xi|\mathcal{A})$ неотрицательной случайной величины ξ относительно $\mathcal{A}$ называется (вообще говоря расширенная, то есть принимающая и значения в бесконечности) измеримая относительно $\mathcal{A}$, случайная величина такая, что $\forall A \in \mathcal{A}$ совпадают интегралы по мере P:

$$\int_A \xi = \int_A \mathcal{E}(\xi|\mathcal{A})$$

Разумеется $\mathcal{E}(\xi|\mathcal{A})$ определена лишь с точностью до множества нулевой меры.

шаг 2 Для произвольной с.в. ξ условное математическое ожидание $\mathcal{E}(\xi|\mathcal{A})$ считается определенным, если

$$\min\left(\mathcal{E}(\xi^-|\mathcal{A}), \mathcal{E}(\xi^+|\mathcal{A})\right) < \infty$$

и тогда $\mathcal{E}(\xi|\mathcal{A}) = \mathcal{E}(\xi^+|\mathcal{A}) - \mathcal{E}(\xi^-|\mathcal{A})$.

Вместо того, чтобы выводить здесь из одной теоремы существования (теоремы Радона-Никодима) теорему существования условного мат.ожидания $\mathcal{E}(\xi|\mathcal{A})$, для частного случая двух абсолютно непрерывных случайных величин рассмотрим более-менее явную конструкцию, повторяющую в основном знакомые построения для дискретных случайных величин. Итак, для абсолютно непрерывного случайного вектора (ξ, η) *т.е. пары непрерывных случайных величин* с двумерной плотностью распределения $f_{(\xi,\eta)}(x, y)$ сначала определим для аргументов $\{y|f_\eta(y) \neq 0\}$ число $E(\xi|\eta = y)$ формулой

$$E(\xi|\eta = y) = \frac{1}{f_\eta(y)} \int\limits_{-\infty}^{+\infty} x f_{\{\xi,\eta\}}(x, y)dx$$

Далее для каждого $\omega \in \Omega$ положим

$$\mathcal{E}\left(\xi\middle|\eta\right)(\omega) = \begin{cases} 0, & f_\eta(\eta(\omega)) = 0 \\ E(\xi|\eta = \eta(\omega)), & f_\eta(\eta(\omega)) \neq 0 \end{cases}$$

Если распределения простых случайных величин и векторов интерпретировать с использованием понятия обобщенной плотности (то есть δ-функций), то легко видеть, что эти построения для абсолютно непрерывного случая следуют приведенной ранее дискретной конструкции.

10.4 ○ Важные абсолютно непрерывные случайные величины

Минимально-необходимый список абсолютно непрерывных случайных величин (задаваемых с помощью их плотностей) предлагается для запоминания.

1. **Равномерные случайные величины и векторы**

   (a) **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:

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

   Это вероятностная модель для встроенного в язык C и `openssl` компьютерного генератора `rand` псевдослучайных чисел.

   (b) Ранее рассмотренные традиционные модели геометрических вероятностей естественно называть **равномерно распределенными в $G \subset \mathbb{R}^m$ случайными векторами**

2. **Гауссовы случайные величины и векторы**

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

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

   Неформальный смысл вытекает из формулы Муавра-Лапласа: гауссова случайная величина – предел *нормализованных* сумм очень большого числа независимых индикаторов. Если $m = 0$ и $\sigma = 1$, то используют термин *нормальный закон распределения*. Традиционно гауссову плотность с параметрами обозначают $\mathcal{N}(m, \sigma^2)$

   (b) **Гауссовы случайные векторы** определяют при помощи большого набора параметров, которые отвечают математическим ожиданиям и ковариациям компонент: n-мерное распределение задается своей плотностью, зависящей от произвольных параметров $m_1, m_2, \dots m_n$ и симметрической, положительно определенной матрицей $\mathbf{\Lambda}$ порядка n.

   Например, вот как гауссовское распределение устроено в трехмерном случае:

   $$\mathfrak{g}(x_1, x_2, x_3) = \sqrt{\frac{\det \mathbf{\Lambda}}{(2\pi)^3}} \exp\left[-\frac{1}{2}\sum_{i,j=1}^{3} \Lambda^{ij}(x_i - m_i)(x_j - m_j)\right],$$

   Прямым вычислением устанавливается вероятностный смысл параметров: матрица $\mathbf{\Lambda}$ является обращением матрицы, составленной из всевозможных попарных ковариаций компонент вектора, а набор $m_1, m_2, \dots m_n$ образует вектор математических ожиданий компонент. Если определить «$\mathbf{\Lambda}$-скалярное произведение векторов» $(\mathbf{x}, \mathbf{y})_{\mathbf{\Lambda}} = (\mathbf{\Lambda x}, \mathbf{y}) = \sum_{i,j=1}^{3} \Lambda^{ij} x_i y_j$ и соответствующую $\mathbf{\Lambda}$-«длину» $|\mathbf{x}|_{\mathbf{\Lambda}} = \sqrt{(\mathbf{x}, \mathbf{x})_{\mathbf{\Lambda}}}$, то формула становится более напоминающей формулу одномерного гауссовского распределения:

   $$\mathfrak{g}(\mathbf{x}) = \sqrt{\frac{\det \mathbf{\Lambda}}{(2\pi)^3}} \exp\left[-\frac{1}{2}\Big(\mathbf{\Lambda}(\mathbf{x} - \mathbf{m}), \mathbf{x} - \mathbf{m}\Big)\right] = \sqrt{\frac{\det \mathbf{\Lambda}}{(2\pi)^3}} e^{-\frac{1}{2}|\mathbf{x}-\mathbf{m}|^2_{\mathbf{\Lambda}}}$$

   Определение плотности многомерного гауссова вектора в размерности n:

   $$\boxed{\sqrt{\frac{\det \mathbf{\Lambda}}{(2\pi)^n}} e^{-\frac{1}{2}|\mathbf{x}-\mathbf{m}|^2_{\mathbf{\Lambda}}}}$$

Если матрица ковариаций диагональная, то и $\mathbf{\Lambda}$ диагональная и, например в трехмерном случае, на диагонали стоят три числа $\frac{1}{\sigma_1^2}, \frac{1}{\sigma_3^2}, \frac{1}{\sigma_3^2}$, от этого все еще дополнительно упрощается:

$$\mathfrak{g}(\mathbf{x}) = \frac{1}{(2\pi)^{3/2}\sigma_1\sigma_2\sigma_3} e^{-\frac{1}{2}\left(\frac{x_1-m_1}{\sigma_1}\right)^2 - \frac{1}{2}\left(\frac{x_2-m_2}{\sigma_2}\right)^2 - \frac{1}{2}\left(\frac{x_3-m_3}{\sigma_3}\right)^2} = \prod_j \frac{1}{\sqrt{2\pi}\sigma_j} e^{-\frac{1}{2}\left(\frac{x_j-m_j}{\sigma_j}\right)^2}$$

Эти рассуждения объясняют следующее важное утверждение, которое (хотя бы в трехмерном случае) требуется уметь доказывать самостоятельно:

(c) если компоненты многомерного гауссовского вектора некоррелированы, то они независимы. Верно и обратное утверждение.

Большинство практиков-нематематиков не склонны различать отличия в понятиях некоррелированности и независимости, в конечном счете это часть наивного убеждения в том, что «в реальной жизни встречаются только гауссовы распределения».

3. **Показательная случайная величина** с параметром $\lambda > 0$ имеет плотность:

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

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.

4. **Логнормальная с.в.** Если экспонента случайной величины гауссова $\mathcal{N}(m, \sigma^2)$, то сама случайная величина называется логнормальной. Элементарное упражнение дает явную формулу для ее плотности:

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

5. **Семейства**

(a) **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

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

равен 1, а потому эта функция задает плотность распределения, называющуюся *гамма-плотностью с параметром r и масштабным параметром 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 10.3 *Семейство гамма-плотностей замкнуто относительно операции свертки:*

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

Доказательство Действительно, вычисление свертки дает выражение

$$f_{[b,r_1]} * f_{[b,r_2]}(x) = \frac{b^{r_1+r_2}}{\Gamma(r_1)\Gamma(r_2)} e^{-x} \int_0^{\infty} (x-y)^{r_1-1} y^{r_2-1} dy$$

Подставляя $y = xt$ видим, что отличие свертки от $f_{[b,r_1+r_2]}$ заключено лишь в постоянном множителе. Однако свертка плотностей по смыслу также должна быть плотностью, поэтому этот множитель должен быть равен единице. $\blacksquare$

(b) **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 10.3 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.

(c) **Распределения Коши.** Для двух независимых 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) таково:

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

Важно, что свертка плотности Коши с параметром u и плотности Коши с параметром v явно вычисляется несложным интегрированием и в ответе получается свертка плотности 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!

11 Последовательности и сходимость

11.1 ◦ Виды сходимости для последовательностей случайных величин

Сходимость числовых последовательностей $\xi_n(\omega)$ для каждого аргумента $\omega \in \Omega$ считается черезвычайно ограничительным требованием и потому в теории вероятностей употребляются следующие основные виды сходимости последовательности случайных величин $\{\xi_n\}$ к случайной величине ξ: *сходимость по вероятности, сходимость почти всюду (иначе называется «сходимость с вероятностью единица»), сходимость в среднем порядка m, сходимость по распределению.*

Определение Последовательность называется сходящейся по вероятности (используется обозначение $\xi_n \overset{P}{\to} \xi$), если для любого $\varepsilon > 0$ $P\{|\xi_n - \xi| > \varepsilon\} \to 0$ при $n \to \infty$

Этот вид сходимости появлялся в связи с законом больших чисел в схеме Бернулли, а именно было показано, что

$$P\left(\left|\frac{S_n}{n} - p\right| > \varepsilon\right) \to 0, \quad n \to \infty$$

Достаточно легко придумать пример сходящейся по вероятности последовательности $\{\xi_k\}$, чтобы ни для какого $\omega \in \Omega$ не существовал бы при $n \to \infty$ предел $\xi_n(\omega)$. Для такого примера удобно вероятностное пространство $(\Omega, \mathcal{B}, P)$ с борелевской σ-алгеброй $\mathcal{B}$ интервала $(0,1) = \Omega$ и с обычной мерой (Лебега) P.

Определение Последовательность называется сходящейся почти всюду (обозначение $\xi_n \overset{\text{п.в}}{\to} \xi$), если множество $W \subset \Omega$ тех исходов ω, в которых последовательность значений $\xi_n(\omega)$ стремится к $\xi(\omega)$, во-первых, измеримо, а во-вторых, имеет вероятность 1. Иначе говоря, множество исходов $\omega \in \Omega$, для которых последовательность значений $\xi_n(\omega)$ не сходится к пределу, равному $\xi(\omega)$, имеет нулевую вероятностную меру.

Этот вид сходимости ослабляет обычное в математическом анализе понятие поточечной сходимости последовательности значений функций тем, что разрешает значениям на «маленьком» множестве аргументов вести себя как попало. В каждом из остальных же аргументов $\hat{\omega}$ числовая последовательность $\xi_n(\hat{\omega})$ должна сходиться к $\xi(\hat{\omega})$. Объединяя условие существования предела последовательности с условием малости множества, на котором такой предел не существует, получаем критерий сходимости почти всюду

Предложение 11.1 *Для выполнения свойства $\xi_n \xrightarrow{п.в} \xi$ необходимо и достаточно, чтобы $\forall \varepsilon > 0$*

$$P\left(\sup_{k \geqslant n} |\xi_k - \xi| \geqslant \varepsilon\right) \to 0, \quad n \to \infty$$

Отсюда, между прочим, получается, что из сходимости почти всюду $\xi_n \xrightarrow{\text{п.в}} \xi$ вытекает сходимость по вероятности $\xi_n \xrightarrow{P} \xi$ (но не наоборот!).

Определение Последовательность называется сходящейся в среднем порядка m (обозначение $\xi_n \xrightarrow{L^m} \xi$), если $E|\xi_n - \xi|^m \to 0$ при $n \to \infty$

Если выбрать $m = 2$ и вспомнить интегральную формулу для вычисления второго момента, то легко видеть, что определение сходимости в среднем в точности воспроизводит определения сходимости в известном из функционального анализа гильбертовом пространстве L^2. Действительно, «расстояние» между функциями $f, g \in L^2$ в гильбертовом пространстве вводилось через скалярное произведение функций, то есть через $\int fg$ — нетрудно видеть аналогию с рассмотрениеи случайных величин с конечным вторым моментом ξ, η как функций на Ω и скалярного произведения для них вида $E(\xi\eta)$. На самом деле несложно проверить, что из сходимости в среднем порядка $m > 0$ вытекает сходимость по вероятности $\xi_n \xrightarrow{P} \xi$, в виду сказанного выше, эта проверка по крайней мере для случая $m = 2$— упражнение по функциональному анализу, поэтому мы ограничимся простой ссылкой на этот факт.

Таким образом, среди первых трех определений самым слабым требованием оказалось условие сходимости по вероятности. Тем не менее,

Определение A: последовательность называется слабо сходящейся или сходящейся по распределению (обозначение $\xi_n \Rightarrow \xi$), если при $n \to \infty$ имеет место сходимость функций распределения $F_{\xi_n}(x)$ к $F_\xi(x)$ для всех тех аргументов x, где $F_\xi(x)$ непрерывна.

Заметьте, что последнее определение использует только закон распределения и даже не предполагает, что случайные величины ξ_n определены на одном и том же вероятностном пространстве Ω. Альтернативное (но эквивалентное, что еще надо доказать!) определение сходимости по распределению таково

Определение B: последовательность называется слабо сходящейся или сходящейся по распределению (обозначение $\xi_n \Rightarrow \xi$), если для любой ограниченной и непрерывной функции $g(x)$ выполнено $E\left[g(\xi_n)\right] \to E\left[g(\xi)\right]$

Идея доказательства эквивалентности двух определений лежит в изучении свойств *почти всюду* сходящихся последовательностей функций вещественного аргумента.

Итак, задана последовательность функций распределения F_n, сходящаяся к функции[28] F во всех точках ее непрерывности.

Лемма 11.2 *Если последовательность с.в. $\{\eta_n\}$ сходится в смысле определения A к случайной величине η_∞, то найдутся с.в. $\{\xi_n\}$ с такими же функциями распределения $F_k = F_{\eta_k}$ как и у η_k, но сходящиеся* **почти всюду** *к некоторой ξ_∞.*

Доказательство Выберем структуру вероятностного пространства $(\Omega, \mathcal{B}, P)$ с борелевской σ-алгеброй $\mathcal{B}$ интервала $(0,1) = \Omega$ с обычной мерой (Лебега) P. Положим $\xi_n(x) = \sup\{y : F_n(y) < x\}$, непосредственно проверяется, что $F_{\xi_n}(x) = F_n$ при $1 \leqslant n \leqslant \infty$, так что точках естественно обозначать $\xi_*(x) = F_*^{-1}(x)$, $x \in (0,1)$ ($*$ – индекс n или отсутствие индекса). Пусть $a_x = \sup\{y : F(y) < x\}$ и $b_x = \inf\{y : F(y) > x\}$, $\Omega_0 = \{z : a_x = b_x\}$, вообще говоря, разные интервалы (a_x, b_x) не могут пересекаться, поэтому их счетное или конечное множество и потому множество точек разрыва $\Omega \setminus \Omega_0$ у функции F не более, чем счетно

При $x \in \Omega_0$ имеем $F(y) < x, \forall y < F^{-1}(x)$ и $F(z) > x, \forall z > F^{-1}(x)$. Для доказательства сходимости $F_n^{-1}(x) \to F^{-1}(x)$ при $\forall x \in \Omega_0$ осталось доказать

$$\liminf_{n\to\infty} F_n^{-1}(x) \geqslant F^{-1}(x)$$
$$\limsup_{n\to\infty} F_n^{-1}(x) \leqslant F^{-1}(x)$$

[28]которая в своих точках разрыва может и не быть функцией распределения

Первое неравенство: Пусть $y < F^{-1}(x)$ — точка непрерывности для F. Так как $x \in \Omega_0$, то $F(y) < x$ и для достаточно больших n выполняется $F_n(y) < x$, то есть $F_n^{-1}(x) \geqslant y$.

Второе неравенство: Пусть $y > F^{-1}(x)$ — точка непрерывности для F. Так как $x \in \Omega_0$, то $F(y) > x$ и для достаточно больших n выполняется $F_n(y) > x$, то есть $F_n^{-1}(x) \leqslant y$.

Лемма доказана. ∎

Предложение 11.3 *Определения А и В равносильны.*

из А выведем В Выберем сходящиеся п.в. случайные величинв ξ_n как в Лемме11.2. Поскольку функция $g()$ непрерывна, то композиции функций $g(\xi_n)$ п.в. сходятся к $g(\xi_\infty)$; по свойству интеграла Лебега $E(\xi_n) \to E(\xi_\infty)$

из В выведем А Выберем непрерывную функцию

$$g_{x,\varepsilon}(y) = \begin{cases} 1 & y \leqslant x \\ 0 & y \geqslant x+\varepsilon \\ \text{линейно убывает} \quad x \leqslant y \leqslant x+\varepsilon \end{cases}$$

Тогда для всех x $\limsup_{n\to\infty} P(\xi_n \leqslant x) \leqslant P(\xi_i nfty \leqslant x)$, что получается переходом к пределу по $\varepsilon \to 0$ в цепочке неравенств

$$\limsup_{n\to\infty} P(\xi_n \leqslant x) \leqslant \limsup_{n\to\infty} E\left(g_{x,\varepsilon}(\xi_n)\right) = E\left(g_{x,\varepsilon}(\xi_\infty)\right) \leqslant P(\xi_\infty \leqslant x+\varepsilon)$$

Теперь переход к пределу по $\varepsilon \to 0$ в сходной цепочке неравенств

$$\liminf_{n\to\infty} P(\xi_n \leqslant x) \geqslant \liminf_{n\to\infty} E\left(g_{x-\varepsilon,\varepsilon}(\xi_n)\right) = E\left(g_{x-\varepsilon,\varepsilon}(\xi_\infty)\right) \geqslant P(\xi_\infty \leqslant x-\varepsilon)$$

дает $\limsup_{n\to\infty} P(\xi_n \leqslant x) \geqslant P(\xi_\infty < x) = P(\xi_\infty \leqslant x)$ в точке x, где $F(x)$ непрерывна.

11.2 □ Сходимость по вероятности влечет сходимость по распределению

Проверим условие В для сходящейся по вероятности последовательности $\xi_n \xrightarrow{P} \xi$. Пусть непрерывная функция $f : \mathbb{R} \to \mathbb{R}$ ограничена константой C, тогда на отрезке $-N \leqslant x \leqslant N$ $\forall \varepsilon > 0\ \exists \delta : |x-y| < \delta \Longrightarrow |f(x) - f(y)| < \varepsilon/2$ При этом N можно взять настолько большим, что $P(|\xi| > N) < \frac{1}{4}\varepsilon/C$.

$$\begin{aligned} E\,|f(\xi_n) - f(\xi)| &= E\left(|f(\xi_n) - f(\xi)| \cdot \mathbf{1}_{\{|\xi_n-\xi|>\delta\}}\right) + E\left(|f(\xi_n) - f(\xi)| \cdot \mathbf{1}_{\{|\xi_n-\xi|\leqslant\delta\}}\right) = \\ &= E\left(|f(\xi_n) - f(\xi)| \cdot \mathbf{1}_{\{|\xi_n-\xi|>\delta\}}\right) + \\ &\quad + E\left(|f(\xi_n) - f(\xi)| \cdot \mathbf{1}_{\{|\xi_n-\xi|\leqslant\delta\}} \cdot \mathbf{1}_{\{|\xi|<N\}}\right) + E\left(|f(\xi_n) - f(\xi)| \cdot \mathbf{1}_{\{|\xi_n-\xi|\leqslant\delta\}} \cdot \mathbf{1}_{\{|\xi|\geqslant N\}}\right) \end{aligned}$$

Поэтому $E\,|f(\xi_n) - f(\xi)| \leqslant 2 \cdot P(|\xi_n - \xi| > \delta) + \varepsilon/2 + \varepsilon/2$

Но $P(|\xi_n - \xi| > \delta) \to 0$ по условию сходимости по вероятности, а ε можно выбирать произвольно малым. Тем самым для любой непрерывной и ограниченной f выполнены условия В и потому последовательность ξ_n сходится по распределению к ξ.

12 Аппарат характеристических функций

12.1 ◦ Определение и зачем нужны характеристические функции.

В этом разделе нам встретятся комплекснозначные случайные величины, то есть принимающие значения не в $\mathbb{R}$, а $\mathbb{C}$. Поскольку мы уже имели дело со случайными векторами, то в принципе можно понимать такую случайную величину $\zeta : \Omega \to \mathbb{C}$, как обычный двумерный случайный вектор с компонентами вещественной и мнимой части от ζ.

Пусть задана случайная величина ξ со значениями в $\mathbb{R}$, для действительного параметра t рассмотрим комплекснозначную случайную величину $\exp(it\xi)$, ее математическое ожидание $E\left(\exp(it\xi)\right)$ как функция параметра t называется **характеристической функцией случайной величины** ξ и обозначается обычно $\varphi_\xi(t)$.

В случае принимающей только неотрицательные целые значения дискретной случайной величины η ранее встретилась конструкция, известная под названием *производящей функциии $Q_\eta(z)$ случайной величины η*, выбором обозначения z для аргумента роль комплексных экспонент удалось затушевать. Сейчас мы вернемся к производящим функциям, записав их аргумент как $z = \exp(it)$:

$$Q_\eta(e^{it}) = e^{it0}P(\eta = 0) + e^{it1}P(\eta = 1) + \ldots + e^{itk}P(\eta = k) + \ldots = E\left(\exp(it\eta)\right)$$

В случае абсолютно непрерывного распределения случайной величины ξ характеристическая функция $\varphi_\xi(t)$ связана с плотностью $f_\xi(t)$ следующей формулой (которая встречалась при обсуждении свойства замен переменной в интеграле Лебега для математического ожидания)

$$\varphi_\xi(t) = E\left(\exp(it\xi)\right) = \int_{-\infty}^{-\infty} e^{itx} f_\xi(x)dx$$

Если ограничиться лишь частными случаями распределений: дискретными или абсолютно непрерывными имеющими все моменты, то в обоих случаях можно опознать преобразование Фурье:

в случае дискретного распределения формулу (обратного) преобразования ряда Фурье в функцию параметра z, лежащего на окружности,

в случае абсолютно непрерывного распределения формулу (обратного) интегрального преобразования Фурье функции $f_\xi(x)$ в функцию параметра $t \in \mathbb{R}$

В принципе, для описания свойств именно таких распределений и их характеристических функций можно было бы просто сослаться на несколько (весьма нетривиальных) теорем из анализа об общих свойствах преобразования Фурье, в том числе преобразования Фурье обобщенных функций[29]. Но уже для смешаных с.в. рассуждения усложнятся 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 весьма сжатым изложением с некоторыми пояснениями. Начнем с пояснений: почему вообще использование характеристических функций оказалось столь значимым для теории вероятностей.

Во-первых, между функциями распределения и характеристическими функциями существует взаимно однозначное соответствие. Поэтому изучение свойств функций распределения можно проводить, изучая соответствующие характеристические функции — это так называемая теорема обращения.

Теорема 12.1 (Теорема обращения) *Пусть $F_\xi(x)$ функция распределения (необязательно непрерывная), а $\varphi_\xi(t)$ соответствующая характеристическая функция. Тогда (чтобы не ограничиваться только абсолютно непрерывными распределениями) во всех точках $a < b$ непрерывности F_ξ выполнено:*

$$F_\xi(b) - F_\xi(a) = \lim_{\varepsilon \to 0} \frac{1}{2\pi} \int_{-\varepsilon}^{+\varepsilon} \frac{e^{-ita} - e^{-itb}}{it} \varphi_\xi(t)dt$$

Если же вдобавок $\int_{-\infty}^{+\infty} |\varphi_\xi(t)|dt < \infty$, то $F_\xi(x)$ имеет плотность $f_\xi(x)$ причем

$$f_\xi(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{itx} \varphi_\xi(t)dt$$

Теорема по сути утверждает обратимость преобразования Фурье и относится к функциональному анализу. Чтобы не терять логику изложения доказательство теоремы перенесено в следующий подраздел.

- Контрольный вопрос:] верно ли, что любая функция $\varphi(t)$ является характеристической для некоторой случайной величины ξ, то есть $\varphi(t) = \varphi_\xi(t)$?

[29] поскольку распределениям дискретных с.в. можно сопоставить обобщенные плотности, состоящие из линейных комбинаций δ-функций

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: $\varphi_{S_n}(t) = \prod_1^n \varphi_{\xi_k}(t)$.

В третьих, важнейшим является то обстоятельство, что слабая сходимость $F_n \stackrel{w}{\to} F$ функций распределения[30] *эквивалентна* поточечной сходимости $\varphi_n \to \varphi$ соответствующих характеристических функций. Это — основной инструмент для доказательства теорем о сходимости по распределению последовательностей случайных величин.

Теорема 12.2 (Теорема непрерывности) *Пусть дана последовательность функций $F_n = F_n(x)$, $x \in \mathbb{R}$ распределения и $\{\varphi_n\}$ — соответствующая последовательность характеристических функций,*

$$\varphi_n(t) = \int_{-\infty}^{\infty} e^{itx} dF_n(x), \quad t \in \mathbb{R} \quad \textit{тогда}$$

1. *Если $F_n \stackrel{w}{\to} F$, где $F = F(x)$ — некоторая функция распределения, то $\varphi_n(t) \to \varphi(t)$, где $\varphi(t)$ — характеристическая функция для $F = F(x)$.*
2. *Если при каждом $t \in \mathbb{R}$ существует предел $\lim_n \varphi_n(t) = \varphi(t)$ непрерывный в точке $t = 0$, то он является характеристической функцией некоторого распределения $F = F(x)$ и $F_n \stackrel{w}{\to} F$*

Доказательство первого утверждения теоремы 12.2 следует из определения слабой сходимости функций, примененного к вещественной и мнимой частям e^{itx} Доказательство второго утверждения достаточно громоздкое и использует не вполне тривиальные результаты функционального анализа, интересующихся отсылаем к первому тому учебника А.Ширяева, Глава III «Сходимость вероятностных мер», §3.

Доказательство теоремы 12.1

Предварительно взглянем на последнее утверждение теоремы. Как отмечено выше если распределение имеет плотность $f_\xi(x)$, то характеристическая функция — преобразование Фурье от интегрируемой функции, поэтому плотность представима как преобразование Фурье от (интегрируемой) функции $\varphi_\xi(t)$:

$$f_\xi(x) = \frac{1}{2\pi} \int\limits_{-\infty}^{\infty} e^{-itx} \varphi_\xi(t) dt$$

Интегрируя левую и правую части и применяя теорему Фубини получаем

$$F_\xi(b) - F_\xi(a) = \int\limits_a^b f(x)dx = \frac{1}{2\pi} \int\limits_a^b \left[\int\limits_{-\infty}^{\infty} e^{-itx} \varphi_\xi(t) dt \right] dx$$

В общем же случае рассуждение в целом похожее, при использовании новых обозначений $\Phi, \Psi_c(x)$:

$$\Phi_c \stackrel{\text{def}}{=} \frac{1}{2\pi} \int_{-c}^{c} \frac{e^{-ita} - e^{-itb}}{it} \varphi_\xi(t) dt = \frac{1}{2\pi} \int_{-c}^{c} \frac{e^{-ita} - e^{-itb}}{it} \left[\int\limits_{-\infty}^{\infty} e^{itx} dF_\xi \right] dt$$
$$= \frac{1}{2\pi} \int\limits_{-\infty}^{\infty} \left[\int\limits_{-c}^{c} \frac{e^{-ita} - e^{-itb}}{it} e^{itx} dt \right] dF_\xi = \int\limits_{-\infty}^{\infty} \Psi_c(x) dF_\xi$$

[30] Напомним, что последовательность с.в. называется *слабо сходящейся или сходящейся по распределению* (обозначение $\xi_n \Rightarrow \xi$), если при $n \to \infty$ имеет место сходимость функций распределения $F_{\xi_n}(x)$ к $F_\xi(x)$ для всех тех аргументов x, где $F_\xi(x)$ непрерывна. Термин «слабая сходимость» *функций* распределения закреплен, согласно традициям функционального анализа, за указанным типом сходимости $F_{\xi_n}(x) \stackrel{w}{\to} F_\xi(x)$.

Опять применяется теорема Фубини, справедливая в силу неравенств:

$$\left|\frac{e^{-ita}-e^{-itb}}{it}e^{itx}dt\right| = \Big|\int\limits_a^b e^{-itx}dx\Big| \leqslant b-a$$

$$\int\limits_{-c}^{c}\int\limits_{-\infty}^{\infty}(b-a)dF_\xi \leqslant 2c(b-a) < \infty$$

функция $g(s,t) = \int\limits_s^t \frac{\sin u}{u}du$ равномерно непрерывна по своим аргументам и стремится к π) при увеличении отрезка интегрирования. Функция $\Psi_c(x)$ может быть представлена через g, так как:

$$\Psi_c(x) = \frac{1}{2\pi}\int\limits_{-c}^{c}\frac{\sin t(x-a)-\sin t(x-b)}{t}dt = \frac{1}{2\pi}\left[\int\limits_{-c(x-a)}^{c(x-a)}\frac{\sin u}{u}du - \int\limits_{-c(x-b)}^{c(x-b)}\frac{\sin u}{u}du\right]$$

Функция $\Psi_c(x)$ ограничена по модулю при всех x, c и

$$\lim_{c\to\infty}(x)\Psi_c(x) = \Psi(x) = \begin{cases} 0 & x < a \text{ или } x - b \\ \frac{1}{2} & x = a \text{ или } x = b \\ 1 & a < x < b \end{cases}$$

В этих обозначениях вероятностной меры μ на $\mathbb{R}$, отвечающей функции F_ξ выполнено:

$$\begin{aligned}\Phi_c = \int_{-\infty}^{\infty}\Psi_c(x)dF_\xi \ \to \ int_{-\infty}^{\infty}\Psi(x)dF_\xi = \mu(a,b) + \frac{\mu(\{a\})}{2} + \frac{\mu(\{b\})}{2} \\ = F_\xi(b-) - F(a) + \frac{1}{2}\left[F_\xi(a) - F_\xi(a-) + F_\xi(b) - F_\xi(a)\right] \\ = \frac{F_\xi(b) - F_\xi(b-)}{2} - \frac{F_\xi(a) - F_\xi(a-)}{2} = F_\xi(b) - F_\xi(a)\end{aligned}$$

для точек непрерывности функции $F_\xi(x)$. Это завершает доказательство первого утверждения теоремы обращения.

Для абсолютно непрерывных $F_\xi(x)$ достаточно подставить в получившееся выражение $F_\xi(x) = \int_{-\infty}^{x} f_\xi(u)du$

Для информации: характеристическая функция случайного вектора

Поскольку функция распределения $F_{\vec{\xi}}(\mathbf{x})$ случайного вектора $(\xi_1, \xi_2, \dots \xi_m)$ зависит от векторного аргумента $\mathbf{x} = (x_1, \dots x_m) \in \mathbb{R}^m$ соответствующая характеристическая функция по определению зависит от векторного аргумента $\mathbf{t} = (t_1, \dots t_m) \in \mathbb{R}^m$ и выражается также через математическон ожидание :

$$\varphi_{\vec{\xi}}(\mathbf{t}) = \int\limits_{\mathbb{R}^m} e^{i(\mathbf{t},\mathbf{x})}dF_{\vec{\xi}}(\mathbf{x})$$

Некоторые примеры характеристических функций

1. Дискретная пуассоновская с.в. Ранее была найдена формула производящей функции, осталось подставить в нее $z = e^{it}$. Ответ: $\exp\left[\lambda(e^{it}-1)\right]$
2. Гауссова с.в.
   (a) Характеристическая функция $\varphi_{\mathcal{N}(0,1)(t)}$ для стандартного гауссова распределения $\mathcal{N}(0,1)$ равна $e^{-t^2/2}$. Доказательство опирается на следующий факт из математического анализа $a > 0$

$$\int_{-\infty}^{+\infty}\exp\left[-ax^2 \pm bx - c\right]dx = \sqrt{\frac{\pi}{a}}\exp\left[\frac{b^2}{a} - c\right]$$

который в свою очередь проверяется вычислением с заменой переменной на полярные координаты. Применим этот факт к интегралу от одномерной плотности стандартного гауссова распределения равной $\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right)$

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}\exp\left[-\frac{1}{2}x^2+itx\right]dx=e^{-t^2/2}$$

Рекомендуется этот факт твердо помнить.

(b) Характеристическая функция $\varphi_{\mathcal{N}(m,\sigma^2)(t)}$ для общего гауссова распределения $\mathcal{N}(m,\sigma^2)$ выводится из соотношения $\eta=\frac{\xi-m}{\sigma}$ случайных величин: стандартной гауссовой η и общей ξ; далее применим первое свойство из следующего раздела 12.2. Ответ: $\varphi_\xi(t)=\exp\left(itm-\frac{t^2\sigma^2}{2}\right)$

3. Для самопроверки умения аналитически вычислять несобственные интегралы вычислите самостоятельно характеристическую функцию стандартного распределения Коши с плотностью $\frac{1}{\pi}\cdot\frac{1}{1+x^2}$. Ответ: $e^{-|t|}$.

12.2 △ Свойства характеристических функций

Индекс с.в. ξ в перечне свойств характеристической функции $\varphi_\xi(t)$ кое-где будем пропускать.

1. Для $a,b\in\mathbb{R}$ $\varphi_{a\xi+b}(t)=e^{ibt}\cdot\varphi_\xi(at)$. Действительно:

$$\varphi_{a\xi+b}(t)=E\left(e^{it(a\xi+b)}\right)=e^{itb}E\left(e^{ita\xi}\right)=e^{itb}\cdot\varphi_\xi(at)$$

2. $|\varphi(t)|\leqslant\varphi(0)=1$ это следует непосрелственно из определения.

3. $\varphi_\xi(t)$ равномерно непрерывна по аргументу $t\in\mathbb{R}$, поскольку

$$|\varphi_\xi(t+h)-\varphi_\xi(t)|=\left|E\left[e^{it\xi}(e^{ih\xi}-1)\right]\right|\leqslant|E(e^{ih\xi}-1)|\underset{h\to 0}{\longrightarrow}0$$

4. $\varphi(t)=\overline{\varphi(-t)}$ это также следует непосрелственно из определения.

5. $\varphi(t)$ вещественнозначная тогда и только тогда, когда соответствующая случайная величина симметрична, то есть $\forall B\in\mathcal{B}(\mathbb{R})$ выполняется $P(\xi\in B)=P(\xi\in -B)$, где $-B=\{-x:x\in B\}$.

   Для доказательства утверждения в одну сторону при вычислении математического ожидания используем аппроксимации исходной симметричной случайной величины также симметричными простыми случайными величинами. Обратно,

$$\varphi_{-\xi}(t)=\varphi_\xi(-t)=\overline{\varphi_\xi(t)}=\varphi_\xi(t)$$

   то есть характеристические функции ξ и $-\xi$ совпадают, а, значит, по теореме обращения совпадают и функции распределения.

6. Если для некоторого $n\geqslant 1$ $E|\xi|^n<\infty$, то $\forall k\leqslant n$ существуют производные $\varphi^{[k]}(t)$, причем

$$E\xi^k=\frac{\varphi^{[k]}(0)}{i^k}\qquad\varphi(t)=\sum_{k=0}^{n}\frac{(it)^k}{k!}E\xi^k+\varepsilon(t)\frac{(it)^k}{k!}$$

   где $\varepsilon(t)\leqslant 3E|\xi|^n$ и $\lim_{t\to n}\varepsilon(t)=0$ В частном случае абсолютно непрерывных распределений это утверждение о возможности дифференциирования под знаком интеграла. Надо только убедиться, что соответствующие интегралы вида

$$\int_{-\infty}^{+\infty}x^k e^{itx}f_\xi(x)dx$$

   сходятся, то есть соответствующие моменты $\forall k\leqslant n$ конечные. Но из условия $E|\xi|^n<\infty$ следует $E|\xi|^k<\infty$ при $\forall k\leqslant n$ — это следствие неравенства для абсолютных моментов $\left(E|\xi|^k\right)^{1/k}\leqslant\left(E|\xi|^n k\right)^{1/n}$; (см. неравенство Йенсена). Поэтому дифференциирование по параметру t под знаком интеграла законно и заодно возникает выражение производных через моменты — это выражение уже встречалось при рассмотрении связи производящей функции дискретной случайной величины. Аналог формулы Тейлора получается комбинацией разложения Тейлора экспоненты и почленным интегрированием. Несколько более техническое, рассуждение для общей функции распределения, мы пропустим.

Замечание

Последнее свойство 6 утверждает в частности, что моменты $E(\xi^k)$ случайной величины ξ можно выразить через производные характеристической функции: $\varphi_\xi^{[k]}(0) = i^k E(\xi^k)$. Возникает естественный вопрос о взаимно однозначном соответствии распределения и его моментов, поскольку, на первый взгляд, знание всех моментов с.в. ξ по идее должно бы обеспечить разложение характеристической функции в ряд Тейлора и, следовательно, по теореме обращения однозначное восстановление функции распределения $F_\xi(x)$. Однако такое утверждение будет неверно[31] например потому, что оценка остаточного члена ε в общем случае сходимость ряда не обеспечит.

13 Важнейшие теоремы теории вероятностей

В этом разделе мы рассмотрим версии Центральной Предельной Теоремы и Закона Больших Чисел, ранее встретившихся при знакомстве с дискретным распределением Бернулли.

13.1 △ Закон Больших Чисел в версии Хинчина

Ранее изученная версия ЗБЧ была доказана для необязательно одинаково распределенных независимых случайных величин, на второй момент которых наложены некоторые ограничения.

В математической статистике важна версия ЗБЧ, в которой требование существования вторых моментов вообще отсутствует:

Теорема 13.1 *Для i.i.d. последовательности $\xi_1, \xi_2, \ldots$ случайных величин с конечными математическими ожиданиями $E(\xi_k) = a$ имеет место сходимость по вероятности:*

$$\frac{S_n}{n} = \frac{\xi_1 + \xi_2 + \ldots + \xi_n}{n} \xrightarrow{P} a$$

Доказательство Как обычно, вычитая из всех случайных величин константу a, сведем все к случаю $a = 0$. Теперь используем характеристические функции: пусть $\varphi_\xi(t) = E\left(\exp(it\xi)\right)$, поскольку первый момент наличествует и равен нулю, то производная хар.функции определена и $\varphi'_\xi(0) = ia = 0$.

Из независимости имеем

$$\varphi_{S_n/n}(t) = \left[\varphi_{\xi_k}\Big(\frac{t}{n}\Big)\right]^n = \left[1 + \varphi'_\xi(0)\frac{t}{n} + o\Big(\frac{t}{n}\Big)\right]^n = \left[1 + o\Big(\frac{t}{n}\Big)\right]^n \ \to\ 1$$

Таким образом хар.функция $\varphi_{S_n/n}(t)$ сходится к хар.функции случайной величины, принимающей значение ноль с вероятностью единица (и, тем самым, доказана сходимость по распределению). Но мы имеем дело с усреднениями с.в., а ранее в 10.1 было замечено, как ведут себя носители меры при усреднениях. Поэтому в этом конкретном случае верна и сходимость по вероятности, действительно, для данного $\varepsilon > 0$ рассмотрим функцию-«шапочку» $\psi(x)$ равную единице в нуле и нулю при аргументах по модулю больших, чем ε. Тогда $E\left(\psi(S_n/n)\right) \ \to\ E\left(\psi(0)\right) = 1$. Но $E\left(\psi(S_n/n)\right) \leqslant 1 - P(|S_n/n| \geqslant \varepsilon)$ поэтому $P(|S_n/n| \geqslant \varepsilon) \ \to\ 0$ ∎

Усиленным ЗБЧ называется утверждение, в котором сходимость по вероятности заменяется сходимостью с вероятностью единица. Частный случай был рассмотрен ранее для схемы Бернулли, вообще же, учитывая варианты с версиями бесконечных значений у случайных величин и мат.ожиданий, существуют несколько разных формулировок УЗБЧ. Для информации отметим здесь, что имеет место похожая на теорему Хинчина следующая теорема Колмогорова:

Теорема 13.2 *Для i.i.d. последовательности $\xi_1, \xi_2, \ldots$ случайных величин с конечными* **абсолютными** *математическими ожиданиями $E(|\xi_k|) = b$ имеет место сходимость почти всюду:*

$$\frac{S_n}{n} = \frac{\xi_1 + \xi_2 + \ldots + \xi_n}{n} \xrightarrow{n.в} E(\xi_1) = a$$

Отметим , что теорема Колмогорова допускает обращение[32] в следующем смысле:

[31] Контрпример обеспечивает, например, 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

[32] Доказательства обеих теорем можно найти, например, в книге А.Ширяев "Вероятность".

Теорема 13.3 *Для i.i.d. последовательности* $\xi_1, \xi_2, \ldots$ *случайных величин таких, что*

$$\frac{S_n}{n} = \frac{\xi_1 + \xi_2 + \ldots + \xi_n}{n} \stackrel{n.в}{\to} C < \infty$$

Тогда $E|\xi_1| < \infty$ *и* $E(\xi_1) = C$

13.2 △ Центральная Предельная Теорема для независимых одинаково распределенных случайных величин

Теорема 13.4 *Для i.i.d. последовательности* $\xi_1, \xi_2, \ldots$ *случайных величин, имеющих математическое ожидание* a *и дисперсию* σ^2*, рассмотрим последовательность сумм вида* $\frac{S_n - an}{\sigma\sqrt{n}}$*. Тогда имеет место сходимость по распределению:*

$$\frac{S_n - an}{\sigma\sqrt{n}} \Rightarrow \mathcal{N}(0,1)$$

Доказательство Преобразуя величины $\xi_k \to \frac{\xi_k - a}{\sigma}$ видим, что в условии можно предполагать $a = 0$ и $\sigma = 1$, а потому

$$\varphi_{\frac{S_n}{\sqrt{n}}}(t) = \varphi_{S_n}\left(\frac{t}{\sqrt{n}}\right) = \left(\varphi_\xi\Big(\frac{t}{\sqrt{n}}\Big)\right)^n = \left(1 - \frac{t^2}{2n} + o\Big(\frac{t^2}{n}\Big)\right)^n \to e^{-t^2/2} = \varphi_{\mathcal{N}(0,1)}(t)$$

Последнее равенство — это классический второй замечательный предел из математического анализа, предыдущие равенства — свойства характеристических функций (см.выше) ∎

13.3 △ Схема серий и теорема Пуассона

Изложенные выше теоремы относились к асимптотическому поведению вероятностей (нормированных и центрированных) сумм $S_n = \xi_1 + \ldots + \xi_n$ последовательностей i.i.d. случайных величин. Рассмотрим более общую модель, называемую *схемой серий случайных величин.*

Будем предполагать, что для каждого $n \geqslant 1$ задана задана треугольная таблица случайных величин

$$\begin{array}{lll}
\xi_{11} & & \\
\xi_{21} & \xi_{22} & \\
\xi_{31} & \xi_{32} & \xi_{33} \\
\ldots & \ldots & \ldots
\end{array}$$

в каждой строчке имеет место независимость (но необязательно одинаковый закон распределения). Положим $S_n = \xi_{n1} + \ldots + \xi_{nn}$

Теорема Пуассона

Пусть задана схема серий дискретных случайных величин и в каждой строке выполнено i.i.d. Кроме того, $P(\xi_{nk} = 1) = p_{nk}$, $P(\xi_{nk} = 0) = q_{nk} = 1 - p_{nk}$, и при $n \to \infty$ выполнено $\max_{1 \leqslant k \leqslant n} p_{nk} \longrightarrow 0$ и $\sum_{k=1}^n p_{nk} \to \lambda > 0$. Тогда для $m = 0, 1, \ldots$

$$P(S_n = m) \longrightarrow \frac{\lambda^m e^{-\lambda}}{m!}$$

Доказательство Характеристическая функция $E\left(e^{it\xi_{nk}}\right) = p_{nk}e^{it} + q_{nk}$ и, значит

$$\varphi_{S_n}(t) = \prod_{k=1}^n \left(p_{nk}e^{it} + q_{nk}\right) = \prod_{k=1}^n \left(1 + p_{nk}(e^{it} - 1)\right) \longrightarrow \exp\left[\lambda(e^{it} - 1)\right]$$

Последнее выражение совпадает с характеристической функцией известной случайной величины (какой?).

13.4 □ Версии Центральной Предельной Теоремы

ЦПТ для разнородных независимых случайных величин

Как уже говорилось, имеется много версий для ЦПТ, отличающихся формулировками условий на последовательность $\{\xi_k\}$ (или схему серий случайных величин $\xi_{i,j}$), а 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, которые в версии для схемы серий будут также и необходимы. Приведем классический вариант условия Линдеберга для последовательности $\xi_1, \xi_2, \ldots$ независимых случайных величин с конечными вторыми моментамив обозначениях для математических ожиданий $m_k = E\xi_k$, и дисперсий $\sigma_k^2 = E\xi_k^2 - m_k^2 = D\xi_k$, $D_n^2 = D(S_n)$:

Классическое условие Линдеберга Для любого $\varepsilon > 0$ и $n \to \infty$

$$\frac{1}{D_n^2} \sum_{k=1}^{n} \int\limits_{\{x:|x-m_k|\geqslant \varepsilon D_n\}} (x-m_k)^2 dF_{\xi_k} \longrightarrow 0$$

При выполнении этого условия справедлива (классическая) Центральная Предельная Теорема:

$$\frac{S_n - E(S_n)}{D_n} \Longrightarrow \mathcal{N}(0,1)$$

Доказательство в принципе 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 $\frac{S_n - E(S_n)}{\sqrt{D(S_n)}}$. Доказательство сходимости не очень короткое.

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 $a_k < \text{const}$ 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.

Схема серий для ЦПТ, необходимость и достаточность условий Линдеберга

Пусть последовательность $\{\xi_k\}$ независимых случайных величин такова, что $E\xi_k = m_k$, $D\xi_k = \sigma_k^2$ $Dn = \sum_1^n \sigma_k^2$, превратим эти данные в таблицу схемы серий, положив $\xi_{nk} = (xi_k - m_k)/D_n$. Сумма в каждой строке обозначается как обычно через $S_n = \xi_{n1} + \ldots + \xi_{nn}$, ясно, что $DS_n = 1$. Вариант **условия Линдеберга**:

$$\forall \varepsilon > 0 \ \lim_{n\to\infty} \sum_{k=1}^{n} E\left[\xi_{nk}^2 \mathbf{1}_{\{|\xi_{nk}|\geqslant\varepsilon\}}\right] = 0$$

Классическая ЦПТ (хоть мы и не стали воспроизводить детали ее доказательства) утверждает, что с.в. S_n сходятся по распределению к $\mathcal{N}(0,1)$. Вместо того, чтобы требований, что $\xi_{nk} = (\xi_k - m_k)/D_n$ можно рассмотреть условия в таблице серий: $E\xi_{nk} = 0$, $DS_n = 1$. Аналогично классической формулировке эти условия (Линдеберга) достаточны, чтобы S_n сходилось бы по распределению к $\mathcal{N}(0,1)$. Замечательным обстоятельством этой версии является то, что в ней эти условия также и *достаточны* .

Пример-тест, когда ЦПТ выполняется, а условие Линдеберга не выполнено.

Пусть дисперсии в последовательности независимых гауссовских величин с нулевым математическим ожиданием $\{\xi_k\}$ при $k \geqslant 2$ растут как $D\xi_k = 2^{k-2}$, $D\xi_1 = 1$. Это напоминает рассуждение предыдущего примера с равномерными распределениями, то есть условие Линдеберга нарушается в схеме

$$\xi_{nk} = \frac{\xi_k}{\sqrt{\sum_{i=1}^{n} D\xi_i}} = \frac{\xi_k}{\sqrt{2 - 2^{k-1}}}$$

С другой стороны, пределом распределений в схеме серий будет $\mathcal{N}(0,1)$, потому, что в каждой строке схемы серий легко найти сумму и она окажется гауссовой случайной величиной $\mathcal{N}(0,1)$, как это может быть?