Краткий конспект лекций по теории вероятностей
(II-й курс, осенний семестр)
лектор С.В. Шапошников

1. Дискретное вероятностное пространство.

Представим себе некий эксперимент, который можно повторить неограниченное число раз, например, бросание монеты или бросание игрального кубика. Пусть непустое множество Ω – множество всех возможных результатов данного эксперимента. Это множество называют *множеством элементарных исходов*. Пусть Ω конечно, т. е.

$$\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}.$$

Всякое подмножество $A \subset \Omega$ называют *событием*. Функцию $P\colon 2^{\Omega} \to [0,1]$, удовлетворяющую следующим свойствам:

(i) $P(\Omega) = 1$,

(ii) $A \cap B = \varnothing \Rightarrow P(A \bigcup B) = P(A) + P(B)$ (правило суммы или аддитивность),

называют *вероятностной мерой*, а значение $P(A)$ *вероятностью* события A. Вероятностная мера P полностью определяется значениями $P(\{\omega\}) = p_{\omega}$. Из определения вероятностной меры следует, что

$$p_{\omega} \geqslant 0 \quad \text{и} \quad \sum_{\omega} p_{\omega} = 1$$

и вероятность произвольного события A вычисляется по формуле

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

Если все элементарные исходы равновозможны, то полагаем $p_{\omega_1} = \ldots = p_{\omega_n} = 1/n$. В этом случае вероятность события A равна отношению количества исходов из A к числу всех исходов в Ω. Отметим, что предположение о равновозможности не всегда корректно. Если эксперимент состоит в бросании правильного кубика, у которого три грани зеленые, одна синяя и две красные, то выпадение зеленого, красного или синего цвета нельзя считать равновозможными исходами. В данном случае выпадение любой из шести граней равновозможно, а зеленый, красный и синий цвета выпадают с вероятностями $1/2$, $1/3$ и $1/6$ соответственно.

Рассмотрим несколько примеров.

Задача о разделе ставки

Два человека играют в некоторую игру, причем у обоих шансы победить одинаковые. Они договорились, что тот, кто первым выиграет 6 партий, получит весь приз. Однако игра остановилась раньше, когда первый выиграл пять партий, а второй выиграл три партии. Как справедливо разделить приз?

Предлагается разделить приз в отношении, в котором относятся вероятности выиграть для каждого из игроков в случае продолжении игры. Ясно, что еще надо сыграть не более трех партий. Пространство исходов этих трех партий состоит из восьми элементов, причем только один из этих исходов означает выигрыш второго игрока. Значит приз надо разделить в отношении 7 к 1.

Вероятностный алгоритм проверки числа на простоту

Пусть дано некоторое натуральное число $N > 1$. Мы хотим проверить является ли это число простым. Можно перебирать все простые делители до $\sqrt{N}$, но это очень долго. Хотелось бы иметь более быстрый способ проверки.

Если N простое число, то по малой теореме Ферма для всякого натурального числа b такого, что НОД$(b, N) = 1$, число $b^{N-1} - 1$ делится на N. Следовательно, если для некоторого b, удовлетворяющего условию НОД$(b, N) = 1$, число $b^{N-1} - 1$ не делится на N, то N не является простым. Это наблюдение используют для построения простейшего теста на простоту. Если $b^{N-1} - 1$ не делится на N, то говорим, что N не проходит тест для основания b.

Пусть основание мы выбираем случайно из множества $\mathbb{Z}_N^*$. Предположим, что существует такое основание, для которого N не проходит тест. Какова вероятность выбрать такое основание?

Предположим, что для $a \in \mathbb{Z}_N^*$ число N не проходит тест. Если N проходит тест для основания b, то для основания ab число N уже тест не проходит. В противном случае $(ab)^{N-1} \equiv 1(\mathrm{mod} N)$ и $(b^{-1})^{N-1} \equiv 1(\mathrm{mod} N)$. Следовательно, $a^{N-1} \equiv (b^{-1})^{N-1}(ab)^{N-1} \equiv 1$, что противоречит предположению. Таким образом, каждому основанию b, для которого N проходит тест, можно сопоставить основание ab, для которого результат теста отрицательный. Значит, оснований, для которых N не проходит тест, не меньше оснований, для которых N проходит тест на простоту. Искомая вероятность не меньше $1/2$. Если независимым образом повторять выбор основания k раз, то вероятность выбрать основание, для которого данное число проходит тест, меньше $1/2^k$.

Отметим, что бывают числа, которые проходят тест для всех оснований b. Это числа Кармайкла, например 561.

Универсальное хеширование

Пусть $K = \{0, 1, 2, \ldots, n-1\}$ – множество «ключей». Отображение

$$h \colon K \to \{0, 1, 2, \ldots, m-1\}$$

называется хеш-функцией. Предполагается, что $m < n$. Одним из важнейших свойств функции h является равномерность, когда доля ключей k с фиксированным значением $h(k)$ должно быть примерно n/m. Это означает, что вероятность коллизии $h(k_1) = h(k_2)$ при $k_1 \neq k_2$ не больше $1/m$. Ясно, что не всегда можно предполагать, что ключи равномерно распределены по таблице. Предположим, что $h(k) = k(\mathrm{mod}\, m)$ и на вход сначала подаются ключи вида $m, 2m, 3m, \ldots$. Ясно, что всем таким ключам присваивается хеш-код 0 и реально никакого равномерного распределения значений не происходит. Оказывается, с этой проблемой можно справиться, если перед началом хеширования случайным образом выбирать функцию h из некоторого набора таких функций.

Зафиксируем простое число $p > n$. Пусть $a \in \{1, 2, \ldots, p-1\}$ и $b \in \{0, 1, 2, \ldots, p-1\}$. Положим

$$h_{a,b}(k) = \big((ak + b)(\mathrm{mod} p)\big)(\mathrm{mod} m).$$

Пара параметров (a, b) выбирается случайным образом из множества

$$\{1, 2, \ldots, p-1\} \times \{0, 1, 2, \ldots, p-1\},$$

причем все элементы этого множества считаем равновероятными. Докажем, что для любых $k_1, k_2 \in \{0, 2, \ldots, n-1\}$, $k_1 \neq k_2$, вероятность коллизии $h_{a,b}(k_1) = h_{a,b}(k_2)$ не превосходит $1/m$.

Заметим, что $ak_1 + b = ak_2 + b(\mathrm{mod} p)$ тогда и только тогда, когда $k_1 = k_2$. Кроме того, для различных k_1 и k_2 по значениям $ak_1 + b(\mathrm{mod} p)$ и $ak_2 + b(\mathrm{mod} p)$ однозначно находятся числа a и b. Пусть $k_1 \neq k_2$. Отображение

$$(a, b) \to (ak_1 + b(\mathrm{mod} p), ak_2 + b(\mathrm{mod} p))$$

является биекцией множества

$$\{1, 2, \ldots, p-1\} \times \{0, 1, 2, \ldots, p-1\}$$

на множество $\{0, 1, 2, \ldots, p-1\} \times \{0, 1, 2, \ldots, p-1\}$, из которого исключены пары $(0, 0)$, $(1, 1)$, …, $(p-1, p-1)$. Остается отметить, что для всякого $t \in \{0, 1, 2, \ldots, p-1\}$ количество чисел $s \in \{0, 1, 2, \ldots, p-1\}$ таких, что $s \neq t$ и $s = t(\mathrm{mod} m)$, не превосходит $(p-1)/m$, т. е. вероятность выбора такой пары (t, s) или, что эквивалентно, выбора пары (a, b), у которой $h_{a,b}(k_1) = h_{a,b}(k_2)$, не превосходит $1/m$.

Колмогоровская сложность

Вычислимое частичное отображение D из множества двоичных слов в себя называется *способом описания*. Таким образом, задано описание D, если задан алгоритм, применяя который к двоичному слову y получаем двоичное слово $x = D(y)$. Слово y называют *описанием* x при способе описания D. Для каждого способа описания D определяется сложность относительно этого описания:

$$KS_D(x) = \min\{l(y) \colon D(y) = x\},$$

где $l(y)$ – длина слова y. Говорят, что способ описания D_1 не хуже способа D_2, если

$$KS_{D_1}(x) \leqslant KS_{D_2}(x) + const$$

для всех слов x. Теорема Колмогорова–Соломонова утверждает, что существует оптимальный способ описания, т.е. способ описания, который не хуже любого другого. Доказательство этого утверждения несложно. Каждому способу описания соответствует некоторая программа, которая реализует соответствующий алгоритм. Можно считать, что программы являются двоичными словами и никакая программа не может быть началом другой. Определим новый способ описания R равенством:

$$R(py) = p(y).$$

Описание R получает на вход двоичное слово py, выделяет в нем программу и применяет эту программу к остатку слова. Пусть есть другой способ описания D и пусть p – соответствующая ему программа. Если y является описание x относительно D, то $x = p(y)$ и py является описание x относительно R, причем длина описания увеличилась на $l(p)$. Следовательно, верно неравенство

$$KS_R(x) \leqslant KS_D(x) + l(p).$$

Фиксируем некоторый оптимальный способ описания D и будем называть *колмогоровской сложностью* $KS(x)$ слова x величину $KS_D(x)$.

Имеет место следующее наблюдение:

для всякого натурального n количество слов x, для которых $KS(x) < n$, меньше 2^n.

Действительно, все такие слова можно получить из описаний y, имеющих длину меньше n, но двоичных слов длины меньше n ровно $2^n - 1 < 2^n$.

Это наблюдение имеет следующую вероятностную интерпретацию. Рассмотрим множество двоичных слов длины n. Вероятность выбрать из них двоичное слово со сложностью меньше $n-k$ не превосходит 2^{-k}. Если считать, что описание слова – результат сжатия, то вероятность того, что случайно выбранное слово длины 100 бит удастся сжать до 90 бит не превосходит 1/1024. Таким образом, большинство слов почти несжимаемы.

Колмогоровская сложность позволяет по новому подойти к понятию случайности. Грубо говоря, двоичное слово случайно, если оно достаточно сложно устроено. Например, мерой случайности можно считать разность $l(x)-KS(x)$. Однако, на конечных двоичных последовательностях (словах) провести грань между случайными и неслучайными словами нельзя. Оказывается, что это можно проделать с бесконечными последовательностями.

Бесконечное множество элементарных исходов

Мы начали разговор с конечного множества исходов Ω. Если множество исходов Ω бесконечно, то ситуация с определением вероятностной меры значительно сложнее. Так на $\Omega = \mathbb{N}$ существует удовлетворяющая свойствам (i) и (ii) функция P, которая равна нулю на всяком конечном множестве и единице на всем N. Для такой функции

$$\sum_k P(\{k\}) = 0 \neq 1 = P(\mathbb{N}).$$

Исправить эту ситуацию можно, если требовать выполнения свойства (ii) не для двух событий, а для любого не более чем счетного набора попарно непересекающихся событий A_n:

(ii)' $P(\cup_n A_n) = \sum_n P(A_n)$ (σ аддитивность).

Правило суммы (ii) называют аддитивностью меры P, а свойство (ii)' называют σ аддитивностью. Кроме того, в случае бесконечного множества Ω возникают проблемы с построением вероятностной меры на множестве всех подмножеств. Пусть $\Omega = [0,1]$. Естественной вероятностной мерой является обычная длина: $P([a,b]) = b-a$, которая обладает важным свойством инвариантности относительно движений. Однако, нельзя определить на множестве всех подмножеств $[0,1]$ такую σ аддитивную вероятностную меру P, что $P([a,b]) = b-a$ и P инвариантна относительно движений. Отметим, что аддитивную меру построить можно в размерности один и размерности два, но нельзя построить в большей размерности.

Таким образом, в случае бесконечного множества элементарных исходов нам придется заменить свойство (ii) на (ii)' и рассматривать в качестве событий не всякое подмножество Ω, а лишь некоторый специальный класс подмножеств. Точные определения мы дадим позднее, а пока ограничимся рассмотрением случая, когда Ω – конечное множество.

Из свойства (ii) по индукции выводится следующее утверждение.

Предложение 1.1. *Для всяких событий $A_1,\ldots A_n$ верно неравенство*

$$P(A_1\cup\ldots\cup A_n) \leqslant \sum_{k=1}^{n} P(A_k).$$

Опять применяя индукцию можно получить *формулу включений и исключений.*

Предложение 1.2. *Для всяких событий $A_1,\ldots A_n$ верно равенство*

$$P(A_1\cup\ldots\cup A_n) = \sum_{k=1}^{n}(-1)^{k-1}\sum_{i_1<\ldots<i_k} P(A_{i_1}\cap\ldots\cap A_{i_k}).$$

Парадокс распределения подарков

N человек принесли подарки друг для друга. Затем эти подарки сложили в мешок и наугад каждый вынул из мешка себе подарок. Какова вероятность того, что конкретный человек вынул подарок, который он принес? Какова вероятность того, что никто не вытащил подарок, который сам принес?

Пространство исходов Ω состоит из всех возможных перестановок чисел $1,2,\ldots,N$, причем все перестановки являются равновозможными. Значит вероятность конкретной перестановки равна $1/N!$. Событие, состоящее в том, что конкретный человек вытащил подарок, который сам принес, состоит из $(N-1)!$ исходов. Следовательно, вероятность такого события равна $1/N$. При больших N эта вероятность стремится к нулю.

Можно было бы думать, что вероятность события: ни один человек не вытащил подарок, который сам принес, стремится к единице, но это ошибочное мнение.

Пусть A_k – событие состоящее в том, что k-й человек вытащил свой подарок. Тогда $A_1 \cup \ldots \cup A_N$ – событие, состоящее в том, что хотя бы один вытащил свой подарок. По формуле включения и исключения

$$P(A_1 \cup \ldots \cup A_n) = \sum_{k=1}^{N} \frac{(-1)^{k-1}}{k!}.$$

Таким образом, вероятность того, что ни один человек не вытащил подарок, который сам принес, равна

$$1 - P(A_1 \cup \ldots \cup A_N) = 1 - 1 + \frac{1}{2!} - \frac{1}{3!} + \ldots$$

и стремится к $1/e$.

Вероятностные рассуждения можно применять в комбинаторных задачах для доказательства существования объекта с заданными свойствами.

Предположим, что заданы события $A_1, \ldots, A_n$. Как проверить, что $P(\cap_k A_k) > 0$?

Пусть B_k – противоположное событие к A_k. Тогда

$$P(\cap_k A_k) = 1 - P(\cup_k B_k) \geqslant 1 - \sum_k P(B_k).$$

Если $\sum_k P(B_k) < 1$, то $P(\cap_k A_k) > 0$. Так как $P(B_k) = 1 - P(A_k)$, то $\sum_k P(B_k) < 1$ тогда и только тогда, когда $\sum_k P(A_k) > n - 1$. Например, если $P(A_k) > 1 - \frac{1}{n}$ для каждого номера k, то $\sum_k P(A_k) > n-1$ и пересечение событий A_k имеет положительную вероятность и, следовательно, непусто.

Проиллюстрируем эти соображения следующим примером.

Конференция

В научном центре работают специалисты по 60 различным разделам компьютерных наук. Известно, что по каждому разделу в центре работает ровно 7 ученых, причем вполне может быть, что один ученый является специалистом сразу по нескольким направлениям. Все ученые должны принять участие в одной (и только одной) из двух конференций, одна из которых проходит в Канаде, а другая в Австралии. Оказывается, что всегда можно так распределить ученых по этим конференциям, что на каждой конференции будут присутствовать специалисты по всем 60 направлениям компьютерных наук.

Будем для каждого ученого выбирать конференцию простым подбрасыванием правильной монеты. Для каждого направления компьютерной мысли рассмотрим событие, состоящее в том, что среди ученых это направления окажутся и те, которые поехали в Канаду, и те, которые поехали в Австралию. Вероятность этого события равна $1 - 2^{-6}$ (нас устроят все исходы кроме двух, когда все отправились на конференцию в одну страну). Остается заметить, что число событий равно 60 и вероятность каждого события больше $1 - 60^{-1}$.

2. Условная вероятность. Формула полной вероятности. Формула Байеса.

Пусть $P(B) > 0$. Условной вероятностью события A при условии B называется число

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

Если фиксировать событие B, то функция $P(\,\cdot\,|B)$ является вероятностной мерой, т. е. удовлетворяет свойствам (i) и (ii). Равенство из определения часто переписывают в виде

$$P(A \cap B) = P(B)P(A|B)$$

и называют правилом произведения.

Теорема 2.1. (Формула полной вероятности) *Пусть* $\Omega = A_1 \cup A_2 \ldots \cup A_n$ *и* $A_i \cap A_j = \varnothing$ *для всех* $i \neq j$. *Предположим, что* $P(A_i) > 0$. *Тогда для всякого события* B *имеет место равенство*

$$P(B) = \sum_i P(B|A_i)P(A_i).$$

Доказательство. Имеем

$$P(B) = \sum_i P(B \cap A_i) = \sum_i P(B|A_i)P(A_i).$$

□

На практике часто возникает ситуация, когда заданы или легко вычисляются именно условные вероятности. Предположим, что задано разбиение Ω на попарно непересекающиеся, непустые события A_i, их вероятности p_i и условные вероятности $P_i(\,\cdot\,) = P(\,\cdot\,|A_i)$. Тогда формула полной вероятности позволяет определить вероятностную меру P, для которой числа p_i и $P_i(B)$ действительно являются вероятностями A_i и условными вероятностями B при условии A_i. Более того, формула полной вероятности гарантирует, что такая вероятностная мера только одна.

Действительно, положим

$$P(B) = \sum_i P_i(B)p_i.$$

Свойства (i) и (ii) очевидно верны. Остается заметить, что $P(A_i) = p_i$ и $P(B \cup A_i) = P_i(B)p_i$.

Коробка с шариками

В коробке лежат m белых и $N-m$ черных шаров. Вынимаем по очереди без возвращения два шара. Какова вероятность того, что вытащили два белых шара?

Решим эту задачу двумя способами.

I. Вероятностное пространство состоит из всех возможных упорядоченных пар чисел от 1 до N. Все пары равновозможны и значит вероятность одной конкретной пары шаров равно $\frac{1}{N(N-1)}$. Количество пар, в которых оба шара белые, равно $m(m-1)$. Значит вероятность того, что вытащили два белых шара, равна $\frac{m(m-1)}{N(N-1)}$.

II. Разобьем вероятностное пространство на два непересекающихся события: A_1 – первый шар белый и A_2 – первый шар черный. Положим $p_1 = \frac{m}{N}$ и $p_2 = \frac{N-m}{N}$. На вероятностном пространстве, состоящем из всех упорядоченных пар чисел от 1 до N, можно определить две вероятностные меры P_1 и P_2. Мера P_1 равна нулю на всех парах, в которых первый шар черный, а все пары, в которых первый шар белый имеют одинаковую вероятность. Мера P_2 равна нулю на всех парах, в которых первый шар белый, а все пары, в которых первый шар черный имеют одинаковую вероятность. Согласно сказанному выше существует вероятностная мера P такая, что $P(A_1) = p_1$, $P(A_2) = p_2$, $P(B|A_1) = P_1(B)$ и $P(B|A_2) = P_2(B)$. Пусть событие B состоит из тех исходов, в которых второй шар белый. Вероятность $P(B|A_1)$ легко вычисляется и равна $\frac{m-1}{N-1}$. По правилу произведения $P(B \cap A_1) = P(B|A_1)P(A_1) = \frac{m(m-1)}{N(N-1)}$.

Обсудим несколько примеров применения формулы полной вероятности.

Задача об экзамене

Программа экзамена содержит N вопросов, а студент выучил только n. В каждом билете ровно один вопрос. На экзамене студенты по очереди подходят и тянут билет. Зависит ли вероятность вытянуть «хороший» билет от места в очереди?

Оказывается, что вероятность не зависит от места в очереди и равна n/N. Пусть перед этим студентом уже вытянуто m билетов. Все пространство исходов разбивается на непересекающиеся события: вытянуто ноль «хороших» билетов, вытянут один «хороший» билет, вытянуты два «хороших билета» и т.д. до m при $m \leqslant n$ или до n при $m > n$. Вероятность вытянуть k «хороших» билетов равна

$$\frac{C_n^k C_{N-n}^{m-k}}{C_N^m}.$$

Если мы знаем, что k ($k < n$) «хороших» билетов уже вытянуто предыдущими m студентами, то вероятность в очередную попытку вытянуть «хороший» билет равна $(n-k)/(N-m)$. По формуле полной вероятности вероятность искомого события равна

$$\sum_k \frac{n-k}{N-m} \frac{C_n^k C_{N-n}^{m-k}}{C_N^m} = \frac{n}{N}.$$

Таким образом, вероятность вытянуть «хороший» билет не зависит от места в очереди.

Задача о сумасшедшей старушке

На посадку в самолет стоят $N \geqslant 2$ пассажиров, среди которых сумасшедшая старушка. Старушка расталкивает всех пассажиров и садится в самолет на произвольное место. Затем пассажиры, когда заходят в самолет, садятся на свое место, если оно свободно, и на произвольное свободное место в противном случае. Какова вероятность того, что последний пассажир сядет на свое место?

Пусть эта вероятность равна P_N. Если $N = 2$, то $P_N = 1/2$. Предположим, что уже для всех $k \leqslant N$ доказано, что $P_k = 1/2$. Докажем равенство $P_{N+1} = 1/2$. Событие B состоит из тех исходов, когда последний пассажир садится на свое место. Событие A_m состоит из тех исходов, когда старушка села на место m-го пассажира. По формуле полной вероятности

$$P_{N+1} = P(B) = \sum_m P(B|A_m)P(A_m).$$

Заметим, что $P(A_m) = 1/(N+1)$ и все кроме двух (когда старушка села на свое место или на место последнего пассажира) вероятности $P(B|A_m) = 1/2$. Следовательно, имеем

$$P_{N+1} = \frac{N-1}{2(N+1)} + \frac{1}{N+1} = \frac{1}{2}.$$

Следующее утверждение позволяет пересчитывать априорные вероятности на основе новых данных.

Теорема 2.2. (формула Байеса) *Пусть* $P(A) > 0$ *и* $P(B) > 0$. *Тогда имеет место равенство*

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

Доказательство. Достаточно заметить, что

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

□

Парадокс Байеса

Пусть имеется тест, используемый для диагностики некоторого заболевания. Известно, что доля больных этим заболеванием равна $0,001$. Если человек болен, то тест дает положительный результат с вероятностью $0,9$. Если человек здоров, то тест дает положительный результат с вероятностью $0,01$. Известно, что тест оказался положительным. Какова вероятность того, что человек на самом деле здоров?

Пусть T_+ и T_- – события состоящие в том, что тест дал положительный результат и тест дал отрицательный результат. Пусть также Z_+ и Z_- – события состоящие в том, что человек здоров и человек болен соответственно. По формуле полной вероятности

$$P(T_+) = P(T_+|Z_+)P(Z_+) + P(T_+|Z_-)P(Z_-) = 0,01 \cdot 0,999 + 0,9 \cdot 0,001 = 0,01089.$$

По формуле Байеса

$$P(Z_+|T_+) = \frac{P(T_+|Z_+)P(Z_+)}{P(T_+)} = \frac{0,01 \cdot 0,999}{0,01089} \geqslant 0,91.$$

Таким образом, при положительном тесте вероятность того, что человек на самом деле здоров, составляет более 90%. Это связано с тем, что точность теста снижается из-за редкости заболевания. В этом случае надо обязательно повторять тест.

Байесовский подход к задаче распознавания образов

Предположим, что существует два класса объектов. Наша задача по некоторому набору параметров X отнести объект к одному из этих классов. Проблема состоит в том, что этих параметров не хватает для однозначной классификации объекта, т. е. такие параметры могут быть и у объекта из первого класса и у объекта из второго класса. Считается, что мы заранее знаем вероятности $P(I)$ и $P(II)$ попадания в каждый класс и условные вероятности $P(X|I)$ и $P(X|II)$ получить именно такой набор параметров X, если известно, что объект относится к I классу или к II классу соответственно. Тогда, получив набор параметров X, с помощью формулы Байеса вычисляются условные вероятности

$$P(I|X) = \frac{P(X|I)P(I)}{P(X|I)P(I) + P(X|II)P(II)} \quad \text{и} \quad P(II|X) = \frac{P(X|II)P(II)}{P(X|I)P(I) + P(X|II)P(II)}.$$

Если оказывается, что $P(I|X) > P(II|X)$, то объект относят к первому классу, а в противоположной ситуации объект относят ко второму классу.

Основная критика такого подхода связана с предположением, что заранее известны распределения вероятностей.

3. Независимые события

С точки зрения вычисления вероятностей независимость события A от события B означает, что вероятность A не меняется от того произошло событие B или нет. Формализовать эту идею позволяет условная вероятность. Событие A не зависит от события B (мы пока предполагаем, что $P(B) > 0$), если

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

что по определению условной вероятности можно переписать в виде $P(A \cap B) = P(A) \cdot P(B)$. Это равенство и принимают в качестве определения независимости.

События A и B называются *независимыми*, если

$$P(A \cap B) = P(A) \cdot P(B).$$

Если события не являются независимыми, то говорят, что они зависимые.

Рассмотрим несколько примеров.

Пример 3.1. Если $A \cap B = \varnothing$ и $P(A) > 0$, $P(B) > 0$, то события A и B зависимые. Действительно, если одно из событий произошло, то другое событие уже произойти не может и значит его вероятность изменилась (стала равна нулю).

Пример 3.2. Если $P(A) = P(B) = 1$, то события независимые (например, $A = B = \Omega$). Действительно,

$$P(A \cap B) = P(A) + P(B) - P(A \cup B) = 1 = P(A) \cdot P(B).$$

Пример 3.3. Выбираем наугад натуральное число от 1 до 100. Событие A состоит из тех чисел, которые делятся на 2, а событие B состоит из тех чисел, которые делятся на 5. Эти события независимые. Действительно, $P(A) = 1/2$, $P(B) = 1/5$ и

$$P(A \cap B) = \frac{1}{10} = \frac{1}{2} \cdot \frac{1}{5} = P(A) \cdot P(B).$$

Пусть теперь выбирается натуральное число от 1 до 101. Тогда уже события A и B окажутся зависимыми:

$$P(A \cap B) = \frac{10}{101} \neq \frac{50}{101} \cdot \frac{20}{101} = P(A) \cdot P(B).$$

Отметим, что из независимости A и B следует независимость $\overline{A} = \Omega \setminus A$ и B:

$$P(\overline{A} \cap B) = P(B) - P(A \cap B) = P(B) - P(A)P(B) = (1 - P(A))P(B) = P(\overline{A})P(B).$$

События $A_1, \ldots, A_n$ называются *независимыми в совокупности*, если

$$P(A_{i_1} \cap A_{i_2} \cap \ldots A_{i_k}) = P(A_{i_1}) \cdot P(A_{i_2}) \cdots P(A_{i_k})$$

для всякого $2 \leqslant k \leqslant n$ и всяких $1 \leqslant i_1 < i_2 < \ldots < i_k \leqslant n$.

Отметим, что независимость в совокупности не совпадает с попарной независимостью.

Это можно проиллюстрировать **парадоксом независимости**.

Два раза бросаем правильную монету. Событие A – при первом бросании выпал герб. Событие B – при втором бросании выпал герб. Событие C – ровно на одном бросании выпал герб. Эти события попарно независимы, но не являются независимыми в совокупности. Действительно, любые два однозначно определяют третье, в частности пересечение A и B исключает C, т. е. $P(A \cap B \cap C) = 0 \neq P(A) \cdot P(B) \cdot P(C)$.

4. Схема Бернулли. Теорема Муавра–Лапласа.

Рассмотрим следующий эксперимент: N раз бросается монета с вероятностью выпадения орла (успеха) p, причем результат одного бросания не влияет на результаты других бросаний. Нас интересует число выпадений орла или число успехов. Пусть $q = 1 - p$.

Можно считать, что множество элементарных исходов состоит из последовательностей 0 и 1 длины N, где 1 соответствует успеху. Каждому исходу с k единицами сопоставляем вероятность $p^k q^{N-k}$. Построенное вероятностное пространство называют *схемой Бернулли*

Так как всего исходов с k единицами C_N^k, то вероятность события $A_{k,N}$, что в последовательности длины N ровно k единиц, равна

$$P(A_{k,N}) = C_N^k p^k q^{N-k}.$$

Ясно, что

$$P(A_{0,N}) + P(A_{1,N}) + \ldots + P(A_{N,N}) = 1.$$

Набор вероятностей $P(A_{0,N}), P(A_{1,N}), \ldots, P(A_{N,N})$ называется *распределением Бернулли*.

Представляет интерес поведение $P(A_{k,N})$ при больших N.

Рассмотрим случай правильной монеты $p = q = 1/2$ и четного числа бросаний $N = 2n$. Положим $p_k = P(A_{n+k,2n})$. Ясно, что $p_{-k} = p_k$. Найдем отношение p_k к p_0 для $k > 0$. Имеем

$$\frac{p_k}{p_0} = \frac{n(n-1)\cdots(n-(k-1))}{(n+1)\cdots(n+k)} = \frac{1 \cdot (1 - \frac{1}{n}) \cdots (1 - \frac{k-1}{n})}{(1 + \frac{1}{n})(1 + \frac{2}{n}) \cdots (1 + \frac{k}{n})}.$$

Заметим, что функция

$$g(x) = \frac{\ln(1+x) - x + x^2/2}{x^3}$$

ограничена в проколотой окрестности нуля и это равенство можно записать так

$$1 + x = e^{x - \frac{x^2}{2} + g(x)x^3}.$$

Следовательно, отношение p_k/p_0 преобразуется к виду

$$\frac{p_k}{p_0} = e^{-\frac{k^2}{n} + \gamma_{k,n}}, \quad \gamma_{k,n} = \frac{k^2}{2n^2} + O(\frac{k^4}{n^3}).$$

По формуле Стирлинга $n! = \sqrt{2\pi n} n^n e^{-n+\alpha_n}$, где $0 < \alpha_n < 1/12n$. Мы можем переписать p_0 в следующем виде:

$$p_0 = \frac{1}{\sqrt{\pi n}} e^{\beta_n}, \quad |\beta_n| \leqslant 1/n.$$

Получаем равенство

$$p_k = \frac{1}{\sqrt{\pi n}} e^{-\frac{k^2}{n} + \gamma_{k,n} + \beta_n}.$$

Предположим, что при каждом n мы выбираем k так, что $|k| \leqslant C\sqrt{n}$, где C не зависит от числа n. Тогда при $N = 2n \to \infty$

$$p_k \sim \frac{2}{\sqrt{N}} \varphi\Big(\frac{2k}{\sqrt{N}}\Big), \quad \varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2},$$

причем погрешность имеет порядок $(1/N)^{3/2}$. Функцию φ называют плотностью *стандартного нормального распределения* и как мы увидим далее φ играет ключевую роль в теории вероятностей.

Если мы хотим оценить вероятность того, что число единиц (число успехов) лежит в некотором интервале, то надо научиться оценивать суммы вида $\sum_{a \leqslant k \leqslant b} p_k$. Если воспользоваться найденной нами асимптотикой, то данная сумма становится похожей на интегральную сумму Римана для функции φ. Рассмотрим отрезок $[0, 2]$. Возьмем его разбиение $x_k = 2k/\sqrt{N}$. Соответствующая сумма Римана имеет вид

$$\Sigma_N = \sum_{0 \leqslant k \leqslant \sqrt{N}} \frac{2}{\sqrt{N}} \varphi\Big(\frac{2k}{\sqrt{N}}\Big)$$

и стремится к интегралу

$$\int_0^2 \varphi(x)\, dx.$$

С другой стороны, k-е слагаемое этой суммы отличается от p_k на величину порядка $(1/N)^{3/2}$ и вся интегральная сумма от соответствующей суммы Римана отличается на величину порядка $1/N$. Таким образом, при $N \to \infty$ имеет место равенство

$$\lim_{N \to \infty} \sum_{0 \leqslant k \leqslant \sqrt{N}} p_k = \int_0^2 \varphi(x)\, dx,$$

т. е. вероятность того, что число единиц лежит в диапазоне от $N/2$ до $N/2 + \sqrt{N}$ (а это значит, что доля успешных бросаний монеты лежит от $1/2$ до $1/2 + 1/\sqrt{N}$) стремится к интегралу от φ по отрезку $[0, 2]$.

Позднее мы докажем, что интеграл по всей числовой прямой от функции φ равен единице. Пусть $a > 1$. Тогда $-\frac{x^2}{2} + x \leqslant -\frac{a^2}{2} + a$ при $x > a$. Имеет место следующее неравенство:

$$\int_a^{+\infty} e^{-x^2/2}\, dx \leqslant e^{-\frac{a^2}{2} + a} \int_a^{\infty} e^{-x}\, dx = e^{-a^2/2}.$$

Следовательно, интеграл от φ по $[0,2]$ не меньше, чем $0,5-(2\pi)^{-1/2}e^{-2}\geqslant 0,45$ и при достаточно больших N вероятность того, что доля успешных бросаний монеты лежит от $1/2$ до $1/2+1/\sqrt{N}$, не меньше $0,45$.

Рассуждения, которые мы провели в частном случае бросания правильной монеты с некоторыми техническими дополнениями можно провести в общем случае. Мы сформулируем общий результат пока без доказательства.

Теорема 4.1. (Теорема Муавра–Лапласа)

(i) *Предположим, что при каждом N число k выбирается так, что для чисел*

$$x_{k,N}=\frac{k-Np}{\sqrt{Npq}}$$

существует константа C, для которой $|x_{k,N}|\leqslant C$ и C не зависит от N. Тогда

$$P(A_{k,N})\sim\frac{1}{\sqrt{Npq}}\varphi(x_{k,N}),\quad N\to+\infty.$$

(ii) *Для любых чисел $a<b$ имеем*

$$\lim_{N\to\infty}P\Big(a\leqslant\frac{k-Np}{\sqrt{Npq}}\leqslant b\Big)=\int_a^b\varphi(x)\,dx.$$

Здесь в левой части написана вероятность того, что число единиц k лежит в диапазоне от $Np+a\sqrt{Npq}$ до $Np+b\sqrt{Npq}$.

Отметим, что в пункте (ii) разница между вероятностью и интегралом оценивается через $\frac{p^2+q^2}{\sqrt{Npq}}$ и эта оценка точна. Следовательно, если p близко к нулю или к единице, то вероятность плохо приближается интегралом от φ.

5. Симметричное случайное блуждание.

Схема Бернулли имеет красивую геометрическую интерпретацию.

По числовой прямой двигается частица, которая каждую секунду перемещается на единицу вправо или на единицу влево, причем выбор обоих направлений равновозможен и не зависит от соответствующего выбора на других шагах. Мы считаем, что в начальный момент времени частица находится в точке $x=0$. Ясно, что траекторию движения частицы за N перемещений можно закодировать последовательностью из 1 или -1 длины N. Набор таких последовательностей – пространство элементарных исходов. Вероятность каждой траектории равна 2^{-N}. Таким образом, с точностью до обозначений мы получили схему Бернулли, описывающую бросание правильной монеты.

При исследовании случайного блуждания, нас интересует вероятность того, что траектория обладает некоторым свойством. Какова вероятность того, что частица не возвращается в начало координат? Какова вероятность, что на N-м шаге частица первый раз вернулась в начало координат?

Для ответа на эти вопросы полезным оказывается следующее простое наблюдение.

Траектории частицы изображаем на координатной плоскости переменных (t,x) в виде ломанных, соединяющих точки с целочисленными координатами t и x. Здесь x – положение частицы, а t – время.

Предложение 5.1. (Принцип отражения) *Пусть $x_0>0$, $x_1>0$ и $t_0<t_1$. Число путей из (t_0,x_0) в (t_1,x_1), которые касаются или пересекают ось времени, равно числу путей из $(t_0,-x_0)$ в (t_1,x_1).*

Доказательство. Устроим биективное соответствие между этими путями. Возьмем путь из (t_0, x_0) в (t_1, x_1), который касается или пересекает ось времени. Пусть $t*$ – первый момент времени, когда $x = 0$. Отразим часть пути, соответствующую отрезку времени $[t_0, t*]$, относительно оси t, а оставшуюся часть оставим без изменений. Получили путь, соединяющий точки $(t_0, -x_0)$ и (t_1, x_1). □

Задача о баллотировке

Какова вероятность того, что частица, которая вышла из нуля и пришла в точку $k > 0$ за N шагов, все время пути находилась в точках с положительными координатами? Обратим внимание, что требуется вычислить условную вероятность, где условием является то, что частица за N шагов пришла в точку k. Следовательно, надо среди таких путей найти долю те, которые проходят только через точки с положительными координатами.

Рассматриваемая задача имеет интересную интерпретацию и называется «теоремой о баллотировке». Если на выборах один кандидат набрал q голосов, а другой r голосов и $r > q$, то какова вероятность того, что победивший кандидат все время выборов был впереди? Предполагается, что голосовавшие не имели предпочтений и отдавали свой голос случайно, а подсчет голосов происходил последовательно.

В первый момент времени частица с вероятностью $1/2$ перемещается в точку $x = 1$ или $x = -1$. Нас устраивает только первый вариант. Затем, нужная нам траектория частицы соединяет точки $(1, 1)$ и (N, k) и не касается и не пересекает ось времени. По принципу отражения мы умеем считать число остальных траекторий, соединяющих $(1, 1)$ и (N, k). Таких траекторий ровно столько, сколько всего траекторий из $(1, -1)$ в (N, k). Несложно посчитать, что таких траекторий $C_{N-1}^{\frac{N+k}{2}}$. Всего траекторий из $(1, 1)$ в (N, k) равно $C_{N-1}^{\frac{N+k}{2}-1}$. Следовательно, число нужных нам траекторий частицы

$$C_{N-1}^{\frac{N+k}{2}-1} - C_{N-1}^{\frac{N+k}{2}} = \frac{k}{N} C_N^{\frac{N+k}{2}}.$$

Здесь $C_N^{\frac{N+k}{2}}$ – количество путей, соединяющих начало координат и точку (N, k). Значит вероятность искомого события равна $\frac{k}{N}$. В условиях задачи о баллотировке соответствующая вероятность равна $\frac{r-q}{r+q}$.

Задача о возвращении в начало координат

Пусть частица вышла из начала координат. Обозначим через u_{2n} вероятность того, что в момент времени $t = 2n$ частица вернулась в точку $x = 0$, а через f_{2n} вероятность того, что это произошло первый раз.

Легко посчитать, что $u_{2n} = C_{2n}^n 2^{-2n}$. Сложнее найти f_{2n}. Частица приходит в точку $x = 0$ в момент времени $2n$ из точек $x = 1$ или $x = -1$ в момент времени $t = 2n - 1$. Число путей в точку $(2n - 1, 1)$ из начала координат таких, что все координаты точек, через которые проходит путь, положительные, равно $\frac{1}{2n-1} C_{2n-1}^n$. Столько же путей в точку $(2n - 1, -1)$ из начала координат таких, что все координаты точек, через которые проходит путь, отрицательные. Следовательно, всего нужных нам путей $\frac{2}{2n-1} C_{2n-1}^n$ и

$$f_{2n} = \frac{2}{2n-1} C_{2n-1}^n 2^{-2n}.$$

Несложно проверить, что $f_{2n} = u_{2n-2} - u_{2n}$ и $f_{2n} = (2n)^{-1} u_{2n-2}$. Формула Стирлинга позволяет найти асимптотику этих вероятностей:

$$u_{2n} \sim \frac{1}{\sqrt{\pi n}}, \quad f_{2n} \sim \frac{1}{2\sqrt{\pi} n^{3/2}}.$$

Вероятность того, что частица за время $2n$ вернулась в начало координат равна

$$f_2 + f_4 + \ldots + f_{2n} = 1 - u_{2n}$$

и с ростом n стремится к единице. Таким образом, частица с вероятностью единица вернется в начало координат. Однако, можно показать, что среднее время возвращения бесконечно.

Броуновское движение

Используем теперь симметричное случайное блуждание для моделирования одномерного диффузионного процесса. В 1826–1827 г английский естествоиспытатель Р. Броун описал хаотическое движение частиц пыльцы в воде. Именно это движение мы и хотим смоделировать.

Будем считать, что частица за время Δt перемещается вправо или влево на Δx, где уже не предполагается, что Δt и Δx равны единице. Пусть в момент времени t частица находится в точке $X(t)$. Хотим узнать распределение значений $X(t)$. Предположим, что $t = N\Delta t$. Если за эти N перемещений частица k раз перемещалась вправо, то

$$X(t) = k\Delta x + (N-k)(-\Delta x) = (2k - N)\Delta x.$$

Из наблюдений известно, что $|\Delta x|^2 = \sigma\Delta t$ для некоторого числа $\sigma > 0$. Тогда

$$X(t) = \frac{k - \frac{N}{2}}{\sqrt{\frac{N}{4}}}\sqrt{t\sigma}.$$

Для моделирования непрерывного движения частицы устремим Δt к нулю. Это равносильно тому, что $N \to \infty$. По теореме Муавра–Лапласа

$$\lim_{N\to+\infty} P\Big(a \leqslant X(t) \leqslant b\Big) = \frac{1}{\sqrt{2\pi}}\int_{\frac{a}{\sqrt{t\sigma}}}^{\frac{b}{\sqrt{t\sigma}}} e^{-x^2/2}\,dx = \int_a^b \frac{1}{\sqrt{2\pi\sigma t}}e^{-\frac{x^2}{2\sigma t}}\,dx.$$

Таким образом, вероятность того, что частица в момент времени t находится в $[a, b]$ вычисляется с помощью плотности $\frac{1}{\sqrt{2\pi\sigma t}}e^{-\frac{x^2}{2\sigma t}}$. Можно проверить, что эта плотность удовлетворяет уравнению теплопроводности

$$\frac{\partial u}{\partial t} = \frac{\sigma}{2}\frac{\partial^2 u}{\partial x^2}.$$

Вывел это уравнения для плотности распределения броуновской частицы в 1905 г А.Эйнштейн.

6. Теорема Пуассона. Распределение Пуассона.

Начнем с нескольких примеров.

Горшочек каши

Всем известна сказка про горшочек каши, который мог сварить сколько угодно каши и даже затопить этой кашей целый город. Предположим, что этот волшебный горшок варит не просто кашу, а с черникой. Более того, горшок хорошо перемешивает кашу и строго соблюдает рецепт и независимо от количества сваренной каши среднее количество ягод черники на половник λ (λ – отношение числа ягод на количество половников каши) остается неизменным. Какова вероятность зачерпнуть половник каши без ягод?

Предположим, что к настоящему моменту горшок сварил кашу с N ягодами черники. Количество половников каши по условию равно N/λ. Вероятность попадания одной ягоды в фиксированный половник равна λ/N. Соответственно, вероятность того, что в этом половнике нет черники, равна

$$\Big(1 - \frac{\lambda}{N}\Big)^N$$

Поскольку горшок варит кашу без остановки, то для получения ответа надо $N \to \infty$. Получаем $e^{-\lambda}$.

Булочка с изюмом

Сколько изюма должны содержать в среднем булочки, для того чтобы вероятность иметь хотя бы одну изюминку в булочке была не меньше $0,99$?

Предположим, что уже изготовлено тесто на некоторое количество булочек. В это тесто добавлено N изюминок так, что отношение числа изюминок к количеству булочек равно λ. Значит количество булочек равно N/λ.

Выделим в тесте кусок, из которого будет изготовлена данная булочка. Вероятность попадания одной изюминки в эту булочку равна λ/N, а вероятность того, что хотя бы одна изюминка попала в булку, равна

$$1-\Big(1-\frac{\lambda}{N}\Big)^N.$$

Поскольку мы рассматриваем серийное производство булочек, то можно предполагать, что $N\to+\infty$, т. е. растет объем теста и количество изюма, но не меняется плотность λ. Как и выше, получаем $\Big(1-\frac{\lambda}{N}\Big)^N\to e^{-\lambda}$. Для решения задачи надо найти λ такое, что $e^{-\lambda}<0,01$. Подходит $\lambda=5$, т. е. плотность изюма должна быть не менее пяти изюминок на булочку.

В рассмотренных выше задачах вероятностная модель строилась следующим образом. Мы рассматривали серии событий, причем N-я серия состоит из N событий. Например, в задаче про горшочек каши такими событиями являются попадание i-й ягоды в половник. В каждой серии все события независимы в совокупности и в N-й серии вероятность каждого события равна p_N, причем число $N\cdot p_N=\lambda$ не зависит от N. Нас интересует вероятность $P(A_{k,N})$ наступления ровно k событий в данной серии из N событий. Поскольку рассматриваемая ситуация представляет собой схему Бернулли, то вероятность $P(A_{k,N})$ вычисляется по формуле $C_N^k p_N^k(1-p_N)^{N-k}$.

Теорема 6.1. (Пуассон) *Пусть $N\cdot p_N=\lambda$ – не зависит от N. Тогда*

$$P(A_{k,N})=C_N^k p_N^k(1-p_N)^{N-k}\to\frac{\lambda^k}{k!}e^{-\lambda},\quad N\to+\infty.$$

Доказательство. Распишем вероятность $P(A_{k,N})$ в следующем виде:

$$P(A_{k,N})=\frac{\lambda^k}{k!}\Big(1-\frac{1}{N}\Big)\cdots\Big(1-\frac{k-1}{N}\Big)\Big(1-\frac{\lambda}{N}\Big)^{-k}\Big(1-\frac{\lambda}{N}\Big)^N.$$

Учитывая, что λ и k не меняются, устремляем $N\to\infty$ и получаем искомое выражение. □

Используя разложение Тейлора для экспоненты, получаем

$$\sum_{k=0}^{\infty}\frac{\lambda^k}{k!}e^{-\lambda}=1.$$

Набор вероятностей $\{\frac{\lambda^k}{k!}e^{-\lambda}\}$ называется *распределением Пуассона*. Если задана последовательность неотрицательных чисел $p_0,p_1,\ldots,p_k,\ldots$ таких, что $\sum_{k=0}^{\infty}p_k=1$, то на множестве $\Omega=\{0,1,2,\ldots\}$ определена вероятностная мера $P(A)=\sum_{k\in A}p_k$. Например, числа $\{\frac{\lambda^k}{k!}e^{-\lambda}\}$ определяют на Ω вероятностную меру P, которую также называют распределением Пуассона.

Пуассоновский процесс

В случайные моменты времени регистрируются некоторые события. Будем отмечать эти моменты времени точками на луче $[0,+\infty)$. Обозначим через $X(t)$ число точек на временном промежутке $(0,t)$. Нас интересует вероятность $P_k(t)$ того, что $X(t)=k$. Будем предполагать, что

(i) вероятность попадания k точек в данный промежуток зависит только от длины этого промежутка, но не зависит от его расположения;

(ii) для любой конечной системы промежутков, которые могут попарно пересекаться лишь концами, попадания точек в каждый из них являются независимыми в совокупности событиями;

(iii) вероятность попадания по крайней мере двух точек в интервал длины δ является о-малым от δ.

На практике многие явления удовлетворяют этим условиям, например вызовы на телефонной станции или радиоактивный распад. Величина $X(t)$ называется пуассоновским процессом. Мы не будем обсуждать существование такого процесса и даже вероятностное пространство, на котором мы работаем в данной модели, а исходя только из свойств (i), (ii), (iii) вычислим вероятности $P_k(t)$.

Рассмотрим сначала $P_0(t)$. Разделим промежуток $[0,t]$ на N промежутков. По свойству (i) вероятность отсутствия событий на каждом промежутке разбиения равна $P_0(t/N)$. По свойству (ii) регистрация события на одном промежутке разбиения не зависит от регистрации событий на других промежутках. Следовательно, мы имеем дело со схемой Бернулли и вероятность отсутствия событий на $[0,t]$ равна $P_0(t) = P_0(t/N)^N$. Положим $P_0(1) = q$. Тогда $P_0\left(\frac{1}{N}\right) = q^{\frac{1}{N}}$ и $P_0\left(\frac{m}{N}\right) = q^{\frac{m}{M}}$. Заметим, что $P_0(t+h) = P_0(t)P_0(h) \leqslant P_0(t)$, т. е. $P_0(t)$ не возрастает. Следовательно, для $\frac{m-1}{N} \leqslant t \leqslant \frac{m}{N}$ выполняются неравенства $q^{\frac{m-1}{N}} \leqslant P_0(t) \leqslant q^{\frac{m}{N}}$. Приближая t последовательностью дробей $\frac{m}{N}$, приходим к равенству $P_0(t) = q^t$. Положим $\lambda = -\ln q > 0$. Тогда $P_0(t) = e^{-\lambda t}$.

Теперь вычислим $P_k(t)$ при $k > 0$. Опять разобьем промежуток $[0,t]$ на N промежутков. Пусть B – событие, состоящее в том, что хотя бы на одном из промежутков зарегистрированы по крайней мере два события. Тогда противоположное событие $\overline{B}$ состоит в том, что в каждом промежутке регистрируется не более одного события. Заметим, что по свойству (iii) вероятность B не превосходит $No(t/N) = o(t)$, что стремится к нулю при $N \to \infty$. Рассмотрим теперь событие $\overline{B}_k$, состоящее в том, что на промежутке $[0,t]$ зарегистрировано ровно k событий и на каждом промежутке разбиения не более одного события. Вероятность отсутствия события на одном промежутке разбиения равна $P_0(t/N) = e^{-\lambda t/N}$ и мы опять находимся в ситуации схемы Бернулли. Следовательно, имеем

$$P(\overline{B}_k) = C_N^k \left(e^{-\lambda t/N}\right)^{N-k} \left(1 - e^{-\lambda t/N}\right)^k.$$

Устремляем здесь $N \to \infty$ и в качестве предела получаем $\frac{(\lambda t)^k}{k!}e^{-\lambda t}$. С учетом сказанного про стремление к нулю $P(B)$ приходим к равенству

$$P_k(t) = \frac{(\lambda t)^k}{k!}e^{-\lambda t},$$

т. е. получаем распределение Пуассона. Число λ называется интенсивностью или параметром процесса $X(t)$.

Аналогичные рассуждения можно провести не только на прямой. Пусть на плоскости случайным образом отмечены точки, например позиции видимых звезд на небосклоне. Будем предполагать, что если множества $V_1, \ldots, V_n$ не накладываются друг на друга, то события, состоящие в попадании некоторого количества точек в V_i, независимы в совокупности. Предположим, что вероятность попадания некоторого числа точек в множество V зависит только от площади $|V|$. Наконец, предположим, что вероятность попадания двух и более точек в V является о-малым от $|V|$. Тогда вероятность $P_k(V)$ попадания k точек в данное множество V можно вычислить в точности повторяя предыдущие рассуждения: $P_k(V) = \frac{(\lambda|V|)^k}{k!}e^{-\lambda|V|}$.

7. Вероятностное пространство. Бесконечное бросание монеты. Парадокс Бертрана.

Пусть множество элементарных исходов Ω теперь может быть бесконечным. С определением вероятностной меры в этом случае возникают некоторые трудности: во первых требуется вместо аддитивности предполагать σ аддитивность, а во вторых во многих содержательных примерах нельзя определить вероятностную меру сразу на всех подмножествах.

Набор множеств, на которых определяется вероятностная мера и которые мы называем событиями, должен удовлетворять некоторым естественным условиям: если A и B события, то к классу событий следует отнести A **и** B, A **или** B, **не** B. Поскольку множество элементарных исходов может быть бесконечным появляется необходимость рассматривать пересечение и объединение бесконечной последовательности событий.

Класс множеств, который содержит $\varnothing$ и Ω, замкнут относительно операций пересечения и объединения, содержит вместе с каждым множеством его дополнение, называется *алгеброй множеств*, или *алгеброй событий*. Если алгебра событий замкнута относительно счетных объединений (или, что равносильно, счетных пересечений), то ее называют *σ алгеброй*.

Например, множество всех подмножеств 2^{Ω}, $\{\varnothing, \Omega\}$, $\{\varnothing, B, \Omega \setminus B, \Omega\}$ являются σ алгебрами. Множество всех конечных объединений попарно непересекающихся промежутков $(a, b] \subset (0, 1]$ является алгеброй, но не является σ алгеброй.

Говорят, что σ алгебра *порождена набором множеств* S и обозначают $\sigma(S)$, если эта σ алгебра является самой маленькой по включению из всех, которые содержат данный набор множеств S.

Важнейшим примером является *борелевская* σ алгебра $\mathcal{B}(\mathbb{R})$ подмножеств прямой $\mathbb{R}$, которая порождена всеми промежутками, т. е. отрезками, интервалами, полуинтервалами. Несложно показать, что в определении не обязательно в качестве порождающего множества брать все промежутки. Можно ограничиться только отрезками или только интервалами или только лучами $(-\infty, c)$. Аналогично определяется $\mathcal{B}([0,1])$ или $\mathcal{B}(\mathbb{R}^n)$ (в многомерном случае надо промежутки заменить на параллелепипеды).

Пусть $\mathcal{A}$ – σ алгебра. Функция $P\colon \mathcal{A} \to [0,1]$ называется *вероятностной мерой*, если $P(\Omega) = 1$ и для всякого не более чем счетного набора попарно непересекающихся событий $\{A_n\} \in \mathcal{A}$ выполняется

$$P(\cup_n A_n) = \sum_n P(A_n).$$

Тройку $(\Omega, \mathcal{A}, P)$ называют *вероятностным пространством.*

Следующие свойства называют непрерывностью меры P.

Предложение 7.1. (i) *Если $A_n \in \mathcal{A}$, $A_n \subset A_{n+1}$ и $A = \cup_n A_n$, то* $\lim_{n\to\infty} P(A_n) = P(A)$.
(ii) *Если $A_n \in \mathcal{A}$, $A_n \supset A_{n+1}$ и $A = \cap_n A_n$, то* $\lim_{n\to\infty} P(A_n) = P(A)$.

Доказательство. Достаточно доказать только (i). Пункт (ii) сводится к (i) переходом от A_n к $\Omega \setminus A_n$. Пусть $C_{n+1} = A_{n+1} \setminus A_n$, $C_1 = A_1$. Тогда $A = \cup_n C_n$ и $A_{n+1} = \cup_{k=1}^n C_k$. По σ аддитивности P

$$P(A) = \sum_n P(C_n) = \lim_{N\to\infty} \sum_{n=1}^N P(C_n) = \lim_{N\to\infty} P(A_{N+1}).$$

□

Важным примером вероятностной меры на $\mathcal{B}(\mathbb{R})$ (или даже на всех подмножествах числовой прямой) является *мера Дирака* в точке a:

$$\delta_a(A) = \begin{cases} 1 & a \in A, \\ 0 & a \notin A \end{cases}$$

С помощью меры Дирака можно построить уже достаточно много разнообразных вероятностных мер. Пусть на прямой отмечены точки $\omega_1, \ldots, \omega_n$ и заданы неотрицательные числа $p_1, \ldots, p_n$ такие, что $p_1 + p_2 + \ldots + p_n = 1$. Положим

$$P(A) = p_1\delta_{\omega_1}(A) + \ldots + p_n\delta_{\omega_n}(A).$$

Это мера сосредоточена в точках ω_i, которые называют атомами.

Замечательный пример доставляет *мера Лебега* λ на $\mathcal{B}([0,1])$. Мера Лебега – обычная длина, т. е. $\lambda([a,b]) = b - a$. Трудным вопросом является существование такой меры. Мы знаем меру промежутка и из-за аддитивности знаем меру конечных объединений попарно непересекающихся промежутков, но этим интересующие нас множества не исчерпываются. Трудное утверждение, что мера Лебега продолжается с конечных объединений промежутков на всю борелевскую σ алгебру мы примем без доказательства.

С помощью отображений можно переносить вероятностную структуру с одного множества на другое. Пусть задано вероятностное пространство $(\Omega, \mathcal{A}, P)$ и определена функция $f\colon E \to \Omega$, тогда на множестве E появляется σ алгебра и вероятностная мера: σ алгебра $\{f^{-1}(A)\colon A \in \mathcal{A}\}$ и вероятностная мера $\mathbb{P}(f^{-1}(A)) = P(A)$.

Бесконечное подбрасывание правильной монеты

Построим вероятностное пространство, которое будет моделировать бесконечное подбрасывание монеты. Напомним, что конечное число бросаний монеты описывается схемой Бернулли. Однако в случае бесконечного числа бросаний ситуация значительно сложнее.

Множество элементарных исходов Ω в этом случае состоит из бесконечных последовательностей 0 и 1 и таким образом является континуальным множеством. Последовательность 0 и 1 задает точку из отрезка $[0,1]$ с помощью двоичной записи, т. е. определена функция $f\colon \Omega \to [0,1]$, сопоставляющая каждой последовательности $\omega = \{\omega_n\}$ число $f(\omega) = \sum_n \omega_n 2^{-n}$. Таким образом, на Ω появляется σ алгебра – прообраз $\mathcal{B}([0,1])$ и вероятностная мера – прообраз λ при отображении f. Этой сигма алгебре принадлежат все множества вида $A_N = \{\omega\colon \omega_1 = \varepsilon_1, \ldots, \omega_N = \varepsilon_N\}$ и $P(A_N) = 2^{-N}$, где $\varepsilon_i = 0$ или 1. Эти утверждения отражают наше интуитивное представление о бесконечном бросании монеты, состоящее в том, что если наше событие касается только первых N бросаний, то вероятность этого события вычисляется исходя из схемы Бернулли для N бросаний и не зависит от результатов последующих бросаний.

Рассмотрим любопытную задачу про бесконечное подбрасывание правильной монеты.

«Penney ante»

Алиса и Боб играют в следующую игру. Бросается правильная монета до тех пор пока не встретится комбинация 110 или 100. Алиса выигрывает, если первой появилась комбинация 110, а Боб в случае, когда первой появилась комбинация 100. Кто будет выигрывать чаще?

Будем считать, что монету продолжают подбрасывать и в случае, когда одна из комбинаций 110 или 100 уже выпала, полагая, что результат дальнейших бросаний не имеет значения. Таким образом, эксперимент выглядит так: случайно выбираем бесконечную последовательность из 0 и 1, а затем смотрим какая из последовательностей 110 или 100 появилась раньше.

Найдем вероятность того, что комбинация 110 появилась первый раз после N-го бросания монеты. Обозначим это событие через R_N. Ясно, что $P(R_1) = P(R_2) = 0$ и $P(R_3) = 1/8$.

Пусть Q_N – событие, состоящее в том, что комбинация 110 не встречается среди первых N бросаний. Тогда с одной стороны

$$P(Q_N) = 1 - P(R_1) - \ldots - P(R_N),$$

а с другой стороны появление последовательности 110 после $N+3$-го бросания означает, что последние три броска фиксированы и на первых N бросаниях нужной последовательности не появилось:

$$P(R_{N+3}) = P(R_{N+3} \cap Q_N) = P(R_{N+3}|Q_N)P(Q_N) = \frac{1}{8}P(Q_N).$$

Объединяя эти два равенства, получаем

$$P(R_{N+3}) = \frac{1 - P(R_1) - \ldots - P(R_N)}{8}.$$

Используя это рекуррентное соотношение можно вычислить вероятность $P(R_N)$, например, $P(R_4) = 1/8$, $P(R_5) = 1/8$, $P(R_6) = 7/64$. Для комбинации 100 рассуждения повторяются дословно и рекуррентное соотношение на вероятность первого появления 100 после N бросаний монеты полностью совпадает с соотношением для $P(R_N)$. Таким образом, эти две комбинации кажутся совершенно равноправными, однако на самом деле игра не является справедливой и одна из комбинаций выигрывает у другой. Найдем вероятность выиграть для Алисы и для Боба.

Пусть A_N – все последовательности, на которых выигрывает Алиса на N-м подбрасывании, а B_N – все последовательности, на которых выигрывает Боб на N-м подбрасывании. Пусть C_N – все последовательности, в которых ни Алиса, ни Боб за первые N бросаний не выиграли. Если после N-го подбрасывания победителей нет и затем на следующих трех бросаниях выпадает 110, то на $N+3$-м подбрасывании выиграла Алиса. Следовательно, имеем

$$P(A_{N+3}) = P(A_{N+3} \cap C_N) = P(A_{N+3}|C_N)P(C_N) = \frac{1}{8}P(C_N).$$

Если после N-го подбрасывания победителей нет и затем на следующих трех бросаниях выпадает 100, то либо на $N+2$-м шаге выиграла Алиса, либо на $N+3$-м шаге выиграл Боб. Следовательно,

$$\frac{1}{8}P(C_N) = \frac{1}{2}P(A_{N+2}) + P(B_{N+3})$$

и

$$P(A_{N+3}) = \frac{1}{2}P(A_{N+2}) + P(B_{N+3}).$$

Заметим, что событие A, состоящее из исходов, когда выиграла Алиса, является объединением непересекающихся событий A_N. Аналогично событие B, когда выиграл Боб, является объединением B_N. По сигма аддитивности

$$P(A) - \frac{1}{8} = \frac{1}{2}P(A) + P(B) - \frac{1}{8}.$$

Следовательно, $P(A) = 2P(B)$. Уже не вычисляя сами вероятности мы можем сделать вывод, что Алиса будет выигрывать вдвое чаще Боба. Покажем, что вероятность того, что никто не выиграл равна нулю. Действительно, это событие является подмножеством множества всех последовательностей, в которых нет 100. Это множество входит множество всех последовательностей таких, что если их разбить на подряд идущие тройки цифр, то среди этих троек нет 100. Вероятность, что если первые $3N$ цифр разбить на N троек, то среди нет 100, равна $(7/8)^N$ и стремиться к нулю при $N \to \infty$.

Итак, $P(A) + P(B) = 1$ и $P(A) = 2/3$, $P(B) = 1/3$. В заключение отметим, что от выбора вероятностного пространства может значительно меняться постановка и соответственно ответ задачи.

Парадокс Бертрана

В круге единичного радиуса проводят случайным образом хорду. Какова вероятность того, что эта хорда длиннее стороны правильного вписанного треугольника?

Задавая различными способами вероятностное пространство в этой задаче можно получить совершенно различные, но абсолютно верные ответы.

Пусть случайное проведение хорды происходит следующим образом: мы отмечаем внутри круга случайным образом точку, а затем проводим через эту точку перпендикулярно диаметру хорду. Вероятностное пространство – это круг единичного радиуса с борелевской σ алгеброй и нормированной мерой Лебега (обычная площадь) в качестве вероятностной меры. Легко понять, что хорда будет длиннее стороны правильного вписанного треугольника тогда и только тогда, когда точка попадет в круг с тем же центром, но радиуса 1/2. Вероятность этого события равна 1/4.

Пусть теперь хорда проводится так: фиксируем на окружности точку и случайно выбираем вторую точку, а затем соединяем ее с первой. Вероятностное пространство – это окружность с σ алгеброй, порожденной дугами, и длиной дуги в качестве вероятностной меры. Ясно, что вероятность провести требуемую хорду равна 1/3.

8. Функция распределения и плотность распределения.

Предположим, что на $\mathcal{B}(\mathbb{R})$ определена вероятностная мера μ. Функция

$$F(t) = \mu((-\infty, t])$$

называется *функцией распределения* меры μ.

Из определения F следует, что $\mu((a, b]) = F(b) - F(a)$.

Предложение 8.1. *Функция F удовлетворяет следующим свойствам:*

(i) $F\colon \mathbb{R} \to [0, 1]$ *не убывает;*

(ii) F *непрерывна справа;*

(iii) $\lim_{t\to-\infty} F(t) = 0$ *и* $\lim_{t\to+\infty} F(t) = 1$.

Доказательство. Обоснуем только пункт (ii). Пусть t_n убывают и стремятся к t. Тогда

$$(-\infty, t] = \bigcap_n (-\infty, t_n].$$

В силу непрерывности μ получаем

$$\lim_{n\to\infty} \mu((-\infty, t_n]) = \mu((-\infty, t]),$$

что означает сходимость $F(t_n) \to F(t)$. □

Теорема 8.1. *Вероятностная мера μ на $\mathcal{B}(\mathbb{R})$ однозначно определяется значениями на промежутках $(a, b]$. Кроме того, если задана функция F, удовлетворяющая свойствам* (i), (ii), (iii), *то существует единственная вероятностная мера μ с функцией распределения F.*

Во многих случаях задавать вероятностную меру на числовой прямой удобно плотностью. Пусть ϱ – неотрицательная и интегрируемая (по Риману) функция на $\mathbb{R}$, причем

$$\int_{-\infty}^{+\infty} \varrho(x)\, dx = 1.$$

Если функция распределения F меры μ задается равенством

$$F(t) = \int_{-\infty}^{t} \varrho(x)\, dx,$$

то говорят, что вероятностная мера μ задана *плотностью* ϱ. В этом случае

$$\mu((a,b]) = F(b) - F(a) = \int_a^b \varrho(x)\,dx.$$

На самом деле можно доказать, что

$$\mu(A) = \int_A \varrho\,dx$$

для всякого множества A, для которого имеет смысл интеграл в правой части, т. е. функция $\mathbb{I}_A\varrho$ интегрируема по Риману, где $\mathbb{I}_A(x) = 1$ при $x \in A$ и $\mathbb{I}_A(x) = 0$ при $x \notin A$.

Приведем два важнейших примера.

(I) **Равномерное распределение**

Равномерным распределением на отрезке $[a,b]$ называют вероятностную меру на числовой прямой, заданную плотностью

$$\varrho(x) = \begin{cases} \frac{1}{b-a} & x \in [a,b] \\ 0 & x \notin [a,b]. \end{cases}$$

Вероятностное пространство в этом случае описывает случайное бросание точки в отрезок $[a,b]$. Вероятность того, что точка попадёт в отрезок $[c,d] \subset [a,b]$ равна $\frac{d-c}{b-a}$.

(II) **Нормальное распределение**

Нормальным распределением с параметрами μ и σ^2 называют вероятностную меру на числовой прямой, заданную плотностью

$$\varrho(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

В случае $\mu = 0$ и $\sigma = 1$ эта плотность появлялась в теореме Муавра–Лапласа.

Рассмотрим теперь вероятностные меры на плоскости. Пусть на $\mathcal{B}(\mathbb{R}^2)$ определена вероятностная мера μ. Функция

$$F(x,y) = \mu((-\infty,x] \times (-\infty,y])$$

называется *функцией распределения* меры μ. Заметим, что

$$\mu((a,b] \times (c,d]) = F(b,d) - F(a,d) - F(b,c) + F(a,c).$$

Предложение 8.2. *Функция F удовлетворяет следующим свойствам:*

(i) $F\colon \mathbb{R}^2 \to [0,1]$ *и* $F(b,d) - F(a,d) - F(b,c) + F(a,c) \geqslant 0$ *для всякого прямоугольника* $(a,b] \times (c,d]$*;*

(ii) F *непрерывна справа по совокупности переменных;*

(iii) $\lim_{(x,y)\to(u,v)} F(x,y) = 0$ *если хотя бы одна из переменных u или v равна $+\infty$;*

(iv) $F(+\infty,+\infty) = 1$.

Доказательство. Доказательство повторяет рассуждения одномерного случая. □

Теорема 8.2. *Вероятностная мера μ на $\mathcal{B}(\mathbb{R}^2)$ однозначно определяется значениями на прямоугольниках $(a,b] \times (c,d]$. Кроме того, для всякой функции F, удовлетворяющей свойствам* (i), (ii), (iii), (iv)*, существует единственная вероятностная мера μ с такой функцией распределения.*

Пусть $\varrho \geqslant 0$ и $\displaystyle\int\int_{\mathbb{R}^2} \varrho(x,y)\,dx\,dy = 1$. Если

$$F(x,y) = \int\int_{(-\infty,x]\times(-\infty,y]} \varrho(x,y)\,dx\,dy,$$

то говорят, что мера μ задана плотностью ϱ. Ясно, что

$$\mu((a,b]\times(c,d]) = \int\int_{(a,b]\times(c,d]} \varrho(x,y)\,dx\,dy.$$

Можно доказать, что

$$\mu(A) = \int\int_A \varrho\,dx\,dy.$$

для всякого множества A, для которого имеет смысл интеграл Римана в правой части.

(I) **Равномерное распределение**

Говорят, что мера μ является *равномерным распределением* на множестве B, имеющем положительную площадь, если μ задана плотностью

$$\varrho(x,y) = \begin{cases} \frac{1}{|B|}, & x\in B, \\ 0, & x\notin B. \end{cases}$$

Аналогичным образом определяется равномерно распределенный вектор с любым конечным числом координат.

(II) **Нормальное распределение**

Мера μ является *нормальным распределением*, если μ задана плотностью вида

$$\varrho(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-q^2}} e^{-\frac{1}{2(1-q^2)}\left(\frac{(x-a)^2}{\sigma_1^2} - 2q\frac{(x-a)(x-b)}{\sigma_1\sigma_2} + \frac{(x-b)^2}{\sigma_2^2}\right)},$$

где $\sigma_1,\sigma_2>0$ и $q\in[0,1)$. Вероятностный смысл коэффициентов $\sigma_{1,2}$ и q мы узнаем позднее.

В многомерной ситуации мера μ является *нормальным распределением*, если ее плотность имеет вид

$$\varrho(x) = Ce^{-q(x)},$$

где $q(x)$ – положительно определенная квадратичная форма, а C нормирующая константа.

9. Кратный интеграл Римана.

Напомним определение кратного интеграла. Сначала определим интеграл от ϱ по $\Pi = [a,b]\times[c,d]$. Разобьем $[a,b]$ на отрезки Δ_i^1 и $[c,d]$ на отрезки Δ_j^2. Тогда прямоугольник Π разбивается на прямоугольники $\Delta_{ij} = \Delta_i\times\Delta_j$. Выбираем в каждом прямоугольнике Δ_{ij} точку Q_{ij} и составляем интегральную сумму

$$\sigma = \sum_{ij} \varrho(Q_{ij})|\Delta_{ij}|.$$

Диаметром D_{ij} прямоугольника Δ_{ij} назовем длину диагонали. Число $D=\max D_{ij}$ называется диаметром разбиения. Если существует предел интегральных сумм при $D\to 0$, то его называют интегралом Римана от ϱ по Π. Интеграл Римана по множеству $A\subset\Pi$ определяется равенством

$$\int\limits_A\int \varrho(x,y)\,dx\,dy = \int\int_\Pi \mathbb{I}_A(x,y)\varrho(x,y)\,dx\,dy.$$

Такой интеграл существует, если например ϱ непрерывна на A и граница множества A составлена из конечного числа кусочков, каждый из которых является графиком непрерывной функции $y=f(x)$ (или $x=f(y)$), где x (или y) принадлежит некоторому отрезку. Такие множества будем называть *допустимыми*.

Нам потребуются два утверждения про кратный интеграл: теорема Фубини и формула замены переменных.

Теорема Фубини утверждает следующее: если функция $\varrho(x,y)$ интегрируема по Π и интегрируема по каждой переменной в отдельности (например, так будет, если ϱ – непрерывная функция), то

$$\int\int\limits_{[a,b]\times[c,d]} \varrho(x,y)\,dx\,dy = \int_a^b \Big(\int_c^d \varrho(x,y)\,dy\Big)\,dx = \int_c^d \Big(\int_a^b \varrho(x,y)\,dx\Big)\,dy$$

Идея доказательства состоит в следующем. Интеграл по Π является пределом интегральных сумм, причем в интегральных суммах можно взять отмеченные точки вида $Q_{ij} = (x_i, y_j)$. Тогда

$$\sum_{i,j} \varrho(x_i,y_j)|\Delta_{ij}| = \sum_i \Big(\sum_j \varrho(x_i,y_j)|\Delta_j|\Big)|\Delta_i|.$$

Левая часть приближается к интегралу по Π, а правая часть приближается к повторному интегралу.

Формула замены переменных

Пусть $\varphi\colon U \to V$, $x = x(u,v)$, $y = y(u,v)$ – диффеоморфизм открытых множеств U и V, т. е. φ – биекция, φ и φ^{-1} непрерывно дифференцируемые функции.

Пусть $A \subset U$ – допустимое множество. Тогда $\varphi(A) \subset V$ – допустимое множество и для всякой интегрируемой по $\varphi(A)$ функции ϱ верно равенство

$$\int\int_{\varphi(A)} \varrho(x,y)\,dx\,dy = \int\int_A \varrho(x(u,v),y(u,v))\Big|J_\varphi(u,v)\Big|\,du\,dv,$$

где J_φ определитель матрицы Якоби

$$J_\varphi(u,v) = \det \begin{pmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{pmatrix}.$$

Идея доказательства состоит в следующем. Пусть

$$\varphi(u,v) = C\begin{pmatrix} u \\ v \end{pmatrix} + b$$

– аффинное невырожденное преобразование. Если Δ – прямоугольник, то $\varphi(\Delta)$ – параллелограмм и площади связаны соотношением: $|\varphi(\Delta)| = |\mathrm{det} C||\Delta|$. Пусть A – прямоугольник, который мы разбили на прямоугольники Δ_{ij}. Тогда параллелограмм $\varphi(A)$ разбивается на параллелограммы $\varphi(\Delta_{ij})$. Выбираем отмеченные точки Q_{ij} в Δ_{ij}. Распишем интегральную сумму

$$\sum_{ij} \varrho(\varphi(Q_{ij}))|\varphi(\Delta_{ij})| = \sum_{i,j} \varrho(\varphi(Q_{ij}))|\mathrm{det} C||\Delta_{ij}|.$$

Левая часть этого равенства приближается к интегралу от ϱ по $\varphi(A)$, а правая приближается к интегралу от $\varrho \cdot |\mathrm{det} C|$ по A. Заметим, что $\mathrm{det} C = J_\varphi$. Общий случай выводится из того, что диффеоморфизм φ в малой окрестности всякой точки (u_0, v_0) хорошо приближается аффинным отображением

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{pmatrix}\begin{pmatrix} u \\ v \end{pmatrix} + \begin{pmatrix} x(u_0,v_0) \\ y(u_0,v_0) \end{pmatrix}.$$

Сформулированная теорема обобщается на случай, когда отображение φ является диффеоморфизмом, если из U выкинуть конечное число точек или гладких кривых (соответственно из V выкинуть их образы). Достаточно потребовать ограниченность J_φ в проколотой окрестности каждой из этих точек или кривых.

Кроме интеграла по ограниченным множествам A нам потребуется интеграл по неограниченным, например по по все плоскости $\mathbb{R}^2$. Исчерпанием множества A называется последовательность допустимых множеств A_n таких, что $A_n \subset A_{n+1}$ и $A = \cup A_n$. Если для всякого исчерпания $\{A_n\}$ множества A существует предел

$$\lim_{n\to\infty} \int\int_{A_n} \varrho(x, y)\, dx\, dy$$

и он не зависит от исчерпания, то этот предел называют интегралом от ϱ по A. Для неотрицательных функций ϱ достаточно проверить существование предела по какому-то одному исчерпанию. Более того, можно показать, что если A допустимое множество (т. е. ограниченное и имеет регулярную границу) и $\{A_n\}$ является исчерпанием A, то интегралы по A_n сходятся к интегралу по A. На интегралы по неограниченным множествам переносятся теорема Фубини и формула замены переменных. Все сказанное здесь для интегралов от функций двух переменных переносится на случай функций любого конечного числа переменных.

Проиллюстрируем наше краткое введение в теорию интеграла функций нескольких переменных следующими примерами.

Интеграл Пуассона

Докажем, что

$$\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx = 1.$$

Имеет место следующая цепочка равенств:

$$\Bigl(\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx\Bigr)^2 = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-x^2/2}\, dx \cdot \int_{-\infty}^{+\infty} e^{-y^2/2}\, dy =$$
$$\frac{1}{2\pi} \int\int_{\mathbb{R}^2} e^{-(x^2+y^2)/2}\, dx\, dy = \frac{1}{2\pi} \int_0^{+\infty} \int_0^{2\pi} re^{-r^2/2}\, dr\, d\varphi = 1.$$

Скорость убывания плотности нормального распределения

Пусть μ является нормальным распределением на $\mathbb{R}^2$, причем $q = 0$ и $\sigma_1 = \sigma_2 = 1$. Плотность распределения μ очень быстро убывает к нулю при $x^2 + y^2 \to \infty$. Оценим величину $\mu(\{x^2 + y^2 \leqslant R^2\})$:

$$\mu(\{x^2 + y^2 \leqslant R^2\}) = \frac{1}{2\pi} \int_0^R \int_0^{2\pi} re^{-r^2/2}\, d\varphi\, dr = 1 - e^{-R^2}.$$

При $R = 2$ эта вероятность больше $0,98$.

10. Случайная величина и ее распределение.

Пусть задано вероятностное пространство $(\Omega, \mathcal{A}, P)$. Функция $\xi\colon \Omega \to \mathbb{R}$ называется *случайной величиной*, если для всякого промежутка I выполнено

$$\xi^{-1}(I) = \{\omega\colon\ \xi(\omega) \in I\} \in \mathcal{A}.$$

Предложение 10.1. *Если ξ случайная величина, то $\xi^{-1}(B) \in \mathcal{A}$ для всякого $B \in \mathcal{B}(\mathbb{R})$.*

Доказательство. Так как

$$\xi^{-1}(A\cap B) = \xi^{-1}(A) \cap \xi^{-1}(B), \quad \xi^{-1}(A\cup B) = \xi^{-1}(A) \cup \xi^{-1}(B), \quad \xi^{-1}(\overline{B}) = \overline{\xi^{-1}(B)},$$

то множество тех элементов $\mathcal{B}(\mathbb{R})$, прообразы которых лежат в $\mathcal{A}$ образуют σ алгебру. По условию эта σ алгебра содержит все промежутки, а значит содержит $\mathcal{B}(\mathbb{R})$. $\square$

Отметим, что если $\mathcal{A} = 2^{\Omega}$, то всякая функция из Ω в $\mathbb{R}$ является случайной величиной.

Функция $f\colon \mathbb{R}^n \to \mathbb{R}$ называется *борелевской*, если $f^{-1}(I) \in \mathcal{B}(\mathbb{R}^n)$ для всякого промежутка I. Несложно показать, что всякая непрерывная функция f является борелевской.

Предложение 10.2. *Пусть $f\colon \mathbb{R}^2 \to \mathbb{R}$ – борелевская функция. Тогда для любых двух случайных величин ξ и η функция $f(\xi(\omega), \eta(\omega))$ является случайной величиной, в частности сумма и произведение случайных величин является случайной величиной.*

Доказательство. По условию $\{(x, y)\colon f(x, y) \in I\} \in \mathcal{B}(\mathbb{R}^2)$. Теперь заметим, что

$$\{\omega\colon (\xi(\omega), \eta(\omega)) \in [a, b] \times [c, d]\} = \xi^{-1}([a, b]) \cap \eta^{-1}([c, d]) \in \mathcal{A}.$$

Как и выше множества из $\mathcal{B}(\mathbb{R}^2)$, прообразы которых при отображении $\omega \mapsto (\xi(\omega), \eta(\omega))$ принадлежат $\mathcal{A}$, образуют σ алгебру и содержат все прямоугольники. Следовательно, эта сигма алгебра совпадает с борелевской и прообраз множества $\{(x, y)\colon f(x, y) \in I\}$, т. е. множество $\{\omega\colon\ f(\xi(\omega), \eta(\omega)) \in I\}$, принадлежит $\mathcal{A}$. $\square$

Предложение 10.3. *Пусть ξ_n – случайные величины и для всякого ω существует предел $\lim_{n\to\infty} \xi_n(\omega) = \xi(\omega)$. Тогда ξ является случайной величиной.*

Доказательство. Рассмотрим множество $\{\omega\colon \xi(\omega) < c\}$. Заметим, что $\xi(\omega) < c$ тогда и только тогда, когда существует такое натуральное число k и номер N, что для всех $n > N$ верно неравенство $\xi_n(\omega) \leqslant c - \frac{1}{k}$. На языке теории множеств эту фразу можно записать так

$$\{\omega\colon \xi(\omega) < c\} = \bigcup_{k,N} \bigcap_{n>N} \{\omega\colon \xi_n(\omega) \leqslant c - \frac{1}{k}\}.$$

Остается заметить, что $\{\omega\colon \xi_n(\omega) \leqslant c - \frac{1}{k}\} \in \mathcal{A}$. $\square$

Таким образом, со случайными величинами можно выполнять арифметические операции и переходить к пределу.

Распределением случайной величины ξ называется вероятностная мера μ_ξ на $\mathcal{B}(\mathbb{R})$, определяемая равенством

$$\mu_\xi(B) = P(\{\omega\colon \xi(\omega) \in B\}) = P(\xi^{-1}(B)).$$

Функция распределения F_ξ вероятностной меры μ_ξ называется *функцией распределения* случайной величины ξ. Таким образом, имеем

$$F_\xi(t) = \mu_\xi((-\infty, t]) = P(\{\omega\colon \xi(\omega) \leqslant t\}).$$

Обратим внимание, что распределение случайной величины – мера на значениях случайной величины, т. е. мера μ_ξ показывает с какой вероятностью принимаются те или иные значения ξ.

Если задана вероятностная мера μ на $\mathcal{B}(\mathbb{R})$, то можно построить вероятностное пространство и случайную величину, для которой эта мера является распределением. Действительно, возьмем $\Omega = \mathbb{R}$, $\mathcal{A} = \mathcal{B}(\mathbb{R})$ и $P = \mu$. Случайная величина $\xi(\omega) = \omega$ имеет распределение μ.

Определения равномерного распределения и нормального распределения, которые мы дали для вероятностных мер, переносятся на случайные величины.

Случайная величина имеет *равномерное распределение на отрезке* $[a, b]$, если мера μ_ξ является равномерным распределением на $[a, b]$. Случайная величина имеет *нормальное распределение*, таковой является мера μ_ξ.

Случайная величина ξ называется дискретной, если ее множество значений конечно или счетно. Если $x_1, \ldots, x_N, \ldots$ – различные значения ξ, то множества $A_i = \xi^{-1}\{x_i\}$ попарно не пересекаются. Пусть $p_i = P(A_i)$. Тогда распределение μ_ξ имеет вид

$$\mu_\xi = p_1\delta_{x_i} + \ldots + p_N\delta_{x_N} + \ldots$$

и полностью определяется значениями x_i и p_i, поэтому распределение дискретной случайной величины представляют в виде таблицы:

x_1	x_2	$\dots$	x_N	$\dots$
p_1	p_2	$\dots$	p_N	$\dots$

Функция распределения в этом случае кусочно постоянная.

Приведем несколько примеров.

1. **Индикатор**

Пусть $A \in \mathcal{A}$. Рассмотрим случайную величину

$$\mathbb{I}_A(\omega) = \begin{cases} 1, & \omega \in A, \\ 0, & \omega \notin A. \end{cases}$$

Соответствующая таблица распределения имеет вид

0	1
q	p

где $p = P(A)$ и $q = 1 - P(A)$. Случайную величину с такой таблицей распределения называют бернуллиевской. Эта случайная величина моделирует однократное бросание монеты с вероятностью орла p.

2. **Схема Бернулли**

Ω – все возможные наборы последовательностей длины N из 0 и 1. Сигма алгебра событий – все подмножества Ω. Вероятностная мера P задана на каждом элементарном исходе следующим правилом: если исход содержит k единиц, то вероятность этого исхода $p^k q^{N-k}$, где $p, q \geqslant 0$ и $p + q = 1$.

Случайная величина $X(\omega)$ – число единиц в исходе ω. Это дискретная случайная величина и таблица ее распределения имеет вид

0	1	$\dots$	k	$\dots$	N
q^N	Npq^{N-1}	$\dots$	$C_N^k p^k q^{N-k}$	$\dots$	p^N

3. **Геометрическое распределение**

Распределение случайной величины X задается следующей таблицей:

1	2	$\dots$	k	$\dots$
p	qp	$\dots$	$q^{k-1}p$	$\dots$

Эта случайная величина моделирует подбрасывание монеты до первого успеха.

11. Совместное распределение случайных величин.

Пусть ξ и η – случайные величины. Рассмотрим отображение $\omega \mapsto (\xi(\omega), \eta(\omega))$. Это отображение определяет на плоскости вероятностную меру μ следующим образом:

$$\mu(B) = P(\{\omega\colon (\xi(\omega), \eta(\omega)) \in B\}).$$

Меру μ называют *совместным распределением* случайных величин ξ и η или распределением случайного вектора (ξ, η). Функцию

$$F(x, y) = \mu((-\infty, x] \times (-\infty, y]) = P(\{\omega\colon \xi(\omega) \leqslant x \,\text{и}\, \eta(\omega) \leqslant y\}).$$

называют *функцией совместного распределения* случайных величин ξ и η или функцией распределения случайного вектора (ξ, η).

Аналогичным образом определяется совместное распределение любого конечного числа случайных величин.

Если известно совместное распределение, то можно найти распределение каждой из величин:

$$\mu_\xi(U) = \mu(U \times \mathbb{R}), \quad \mu_\eta(V) = \mu(\mathbb{R} \times V).$$

Однако, если известны только распределения случайных величин, то найти совместное распределение нельзя.

Рассмотрим пример. Пусть в квадрате $[0,1]\times[0,1]$ случайно выбирается точка (x,y). Случайные величины $\xi(x,y)=x$ и $\eta(x,y)=y$ имеют равномерное распределение на $[0,1]$ и их совместное распределение является равномерным на $[0,1]\times[0,1]$, т. е. вероятность попадания в борелевское множество B равно площади этого множества. Будем теперь выбирать точку (x,y) случайным образом на диагонале квадрата $[0,1]\times[0,1]$, а случайные величины останутся прежними. Для всякого отрезка $[a,b]\subset[0,1]$ вероятность того, что $(x,y)\in[a,b]\times\mathbb{R}$ равна вероятности попасть в отрезок длины $(b-a)\sqrt{2}$ при бросании точки на отрезок длины $\sqrt{2}$, т. е. равна $b-a$. Таким образом, ξ и η опять имеют равномерное распределение на $[0,1]$, но совместное распределение у них совсем другое.

Если известно совместное распределение ξ и η, то можно найти распределение $f(\xi,\eta)$, где f – борелевская функция двух переменных. Действительно, для всякого $B\in\mathcal{B}(\mathbb{R})$

$$\{\omega\colon f(\xi(\omega),\eta(\omega))\in B\}=\{\omega\colon(\xi(\omega),\eta(\omega))\in f^{-1}(B)\}.$$

Если существует интегрируемая (по Риману) и неотрицательная функция $\varrho(x,y)$ такая, что

$$\mu(A)=\int\int_A\varrho(x,y)\,dx\,dy$$

для всякого допустимого (измеримого по Жордану) множества A, то ϱ называется *совместной плотностью распределения*.

Если известна плотность ϱ совместного распределения ξ и η, то можно найти плотности распределения каждой из случайных величин. Например, для случайной величины ξ:

$$\mu_\xi((a,b])=\mu((a,b]\times\mathbb{R})=\int_a^b\Big(\int_{-\infty}^{+\infty}\varrho(x,y)\,dy\Big)\,dx$$

и, следовательно,

$$\varrho_\xi(x)=\int_{-\infty}^{+\infty}\varrho(x,y)\,dy.$$

Если распределение каждой из случайных величин задается плотностью, то совместное распределение может не иметь плотность.

Говорят, что вектор (ξ,η) *равномерно распределен* на множестве B, имеющем положительную площадь, если его мера μ (совместное распределение) является равномерным распределением на B. Вектор (ξ,η) *нормально распределен*, если таковым является его совместное распределение.

Игла Бюффона

Пусть на «коредор» бесконечной длины и единичной ширины на плоскости случайным образом бросается игла единичной длины. Какова вероятность того, что игла пересечет хотя бы одну стенку «коридора»?

Пусть d – расстояние от середины иглы до левой стенки и α – угол между иглой и горизонталью. Будем предполагать, что вектор (d,α) имеет равномерное распределение на множестве $[0,1]\times[-\pi/2,\pi/2]$.

Игла пересекает стенку «коридора» тогда и только тогда, когда

$$(d,\alpha)\in A=\{(d,\alpha)\colon|\alpha|<\frac{\pi}{2},d\in[0,\frac{1}{2}\cos\alpha]\cup[1-\frac{1}{2}\cos\alpha,1]\}.$$

Вероятность нужного нам события равна

$$\int\int_A\frac{1}{\pi}\,dx\,dy=\frac{1}{\pi}\int_{-\pi/2}^{\pi/2}dy\Big(\int_0^{2^{-1}\cos y}1\,dx+\int_{1-2^{-1}\cos y}^1 1\,dx\Big)=\frac{1}{\pi}\int_{-\pi/2}^{\pi/2}\cos y\,dy=\frac{2}{\pi}.$$

Этот результат может быть использован для экспериментального вычисления числа π.

В заключение рассмотрим еще один пример.

Преобразование равномерного распределения в нормальное распределение

Пусть (ξ, η) равномерно распределен на $K = [0,1] \times [0,1]$. Вектор (X, Y), где

$$X = \sqrt{-2\ln\xi}\cos(2\pi\eta), \quad Y = \sqrt{-2\ln\xi}\sin(2\pi\eta),$$

имеет нормальное распределение с плотностью $\frac{1}{2\pi}e^{-(x^2+y^2)/2}$. Найдем сначала плотность распределения вектора $(\sqrt{-2\ln\xi}, \eta)$. Заметим, что при $0 < a < b$ выполнено

$$\sqrt{-2\ln\xi} \in [a,b] \Leftrightarrow \xi \in \xi \in [e^{-b^2/2}, e^{-a^2/2}].$$

Пусть $0 < a < b$ и $0 < c < d < 2\pi$. Имеем

$$P((\sqrt{-2\ln\xi}, 2\pi\eta) \in [a,b] \times [c,d]) =$$
$$\int\int_{[e^{-b^2/2}, e^{-a^2/2}] \times [c/2\pi, d/2\pi]} 1\, dx\, dy = \int\int_{[a,b]\times[c,d]} \frac{1}{2\pi} r e^{-r^2/2}\, dr\, d\varphi.$$

Следовательно, плотность распределения вектора $(\sqrt{-2\ln\xi}, 2\pi\eta)$ имеет вид

$$\varrho_1(r,y) = \begin{cases} \frac{1}{2\pi} r e^{-r^2/2} & (r,\varphi) \in (0,+\infty) \times [0,1], \\ 0 & (r,\varphi) \notin (0,+\infty) \times [0,1]. \end{cases}$$

Для произвольного прямоугольника Π обозначим его образ при полярной замене координат через $\Phi(\Pi)$.

$$P(\{\omega\colon (X(\omega), Y(\omega)) \in \Pi\}) = P((\sqrt{-2\ln\xi}, 2\pi\eta) \in \Phi(\Pi)) =$$
$$\int\int_{\Phi(\Pi)} \frac{1}{2\pi} r e^{-r^2/2}\, dr\, d\varphi = \int\int_{\Pi} \frac{1}{2\pi} e^{-(x^2+y^2)/2}\, dx\, dy.$$

Отметим, что данное преобразование позволяет генерировать нормально распределенный вектор, если уже есть генератор равномерного распределения.

12. Независимые случайные величины.

Случайные величины ξ и η называются *независимыми*, если для всяких промежутков U и V выполняется равенство

$$P(\{\omega\colon \xi(\omega) \in U \text{ и } \eta(\omega) \in V\}) = P(\{\omega\colon \xi(\omega) \in U)\}) \cdot P(\{\omega\colon \eta(\omega) \in V)\}).$$

В терминах совместного распределения это равенство записывается так:

$$\mu(U \times V) = \mu_\xi(U) \cdot \mu_\eta(V).$$

Говорят, что мера μ является произведением мер μ_ξ и μ_η и пишут $\mu = \mu_\xi \otimes \mu_\eta$.

Если есть два вероятностных пространства $(\Omega_1, \mathcal{A}_1, \mu_1)$ и $(\Omega_2, \mathcal{A}_2, \mu_2)$, то можно задать меру $\mu = \mu_1 \otimes \mu_2$ на σ алгебре $\mathcal{A}_1 \otimes \mathcal{A}_2$, порожденной прямоугольниками $A_1 \times A_2$, где $A_1 \in \mathcal{A}_1$ и $A_2 \in \mathcal{A}_2$. Действительно, на каждом таком прямоугольнике определяем меру равенством

$$\mu(A_1 \times A_2) = \mu_1(A_1) \cdot \mu_2(A_2).$$

Затем по аддитивности продолжаем μ на алгебру всех конечных объединений попарно непересекающихся прямоугольников, а затем используем известный, но трудный, факт о существовании единственного продолжения на $\mathcal{A}_1 \otimes \mathcal{A}_2$.

Используя эту конструкцию, можно для любых двух данных вероятностных распределений μ_1 и μ_2 на $\mathcal{B}(\mathbb{R})$ построить вероятностное пространство и две независимые случайные величины ξ и η с такими распределениями. Положим $\Omega = \mathbb{R}_x \times \mathbb{R}_y$, $\mathcal{A} = \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R}) = \mathcal{B}(\mathbb{R}^2)$ и $P = \mu_1 \otimes \mu_2$. Случайные величины $\xi(x,y) = x$ и $\eta(x,y) = y$ имеют требуемые распределения и являются независимыми.

Предложение 12.1. *Независимость ξ и η равносильна тому, что*

$$F(x,y) = F_\xi(x) \cdot F_\eta(y),$$

где F – совместная функция распределения ξ и η.

Доказательство. Достаточно заметить, что совместное распределение однозначно определяется функцией распределения F. Если функция распределения F совпадает с функцией распределения меры $\mu_\xi \otimes \mu_\eta$, то и меры совпадают. □

Предложение 12.2. *Если случайные величины ξ и η независимы, то для всяких борелевских функций f и g случайные величины $f(\xi)$ и $g(\eta)$ независимы.*

Доказательство. Следует из определения независимости. □

Предложение 12.3. *Пусть распределения ξ и η заданы плотностями. Тогда независимость ξ и η равносильна тому, что совместное распределение задано плотностью и эта плотность имеет вид*

$$\varrho(x,y) = \varrho_\xi(x)\varrho_\eta(y).$$

Доказательство. По теореме Фубини

$$\int_a^b \varrho_\xi(x)\,dx \cdot \int_c^d \varrho_\eta(y)\,dy = \int\int\limits_{(a,b]\times(c,d]} \varrho_\xi(x)\varrho_\eta(y)\,dx\,dy = \mu((a,b]\times(c,d]).$$

□

Следствие 12.1. *Предположим, что ξ и η независимы и их распределения ξ и η заданы плотностями. Тогда распределение суммы $\nu = \xi + \eta$ задано плотностью*

$$\varrho_\nu(x) = \int_{-\infty}^{+\infty} \varrho_\xi(t)\varrho_\eta(x-t)\,dt.$$

Доказательство. По определению $F_\nu(t) = P(\{\omega\colon \xi(\omega)+\eta(\omega) < t\})$. С другой стороны, эта вероятность выражается через интеграл

$$\int\int_{x+y<t} \varrho_\xi(x)\varrho_\eta(y)\,dx\,dy.$$

Переходя к новым переменным $u = x+y$, $v = x$, и, применяя теорему Фубини, преобразуем этот интеграл:

$$\int_{-\infty}^{t}\Big(\int_{-\infty}^{+\infty} \varrho_\xi(v)\varrho_\eta(u-v)\,dv\Big)\,du.$$

Следовательно, распределение ν имеет плотность требуемого вида. □

Всякая случайная величина ξ *порождает σ алгебру* $\sigma(\xi) = \{\xi^{-1}(B)\colon B \in \mathcal{B}\}$. Эта σ алгебра является подмножеством σ алгебры событий $\mathcal{A}$, но может с ней и не совпадать. Например, если $\xi = const$, то $\sigma(\xi) = \{\varnothing, \Omega\}$. σ алгебра $\sigma(\xi)$ содержит все события, которые как-то связаны с ξ. Более того, имеет место следующее замечательное, но довольно трудное утверждение: если случайная величина η такова, что $\eta^{-1}(B) \in \sigma(\xi)$ для всякого борелевского множества B, то найдется такая борелевская функция F, что $\eta = F(\xi)$. Таким образом, если события, связанные с величиной η являются частью событий, связанных с ξ, то η выражается через величину ξ.

Определение независимости можно переписать следующим образом: ξ и η независимы тогда и только тогда, когда всякие два события $A_1 \in \sigma(\xi)$ и $A_2 \in \sigma(\eta)$ являются независимыми. Действительно, если $A_1 = \xi^{-1}(B_1)$ и $A_2 = \eta^{-1}(B_2)$, то

$$A_1 \cap A_2 = \{\omega\colon \xi(\omega) \in B_1 \,\text{и}\, \eta(\omega) \in B_2\}$$

и

$$P(A_1 \cap A_2) = P(\{\omega\colon \xi(\omega) \in B_1\}) \cdot P(\{\omega\colon \eta(\omega) \in B_2\}) = P(A_1) \cdot P(A_2).$$

Если для двух σ алгебр верно, что всякие два события, одно из которых принадлежит первой σ алгебре, а второе принадлежит второй σ алгебре, независимы, то говорят, что эти *σ алгебры независимы.*

σ алгебры $\mathcal{A}_1, \ldots, \mathcal{A}_n$ *независимы*, если для всяких $A_i \in \mathcal{A}_i$ выполнено

$$P(A_1 \cap \ldots \cap A_n) = P(A_1) \cdots P(A_n).$$

Это определение равносильно тому, что всякие события взятые по одному из каждой σ алгебры являются независимыми в совокупности.

Случайные величины $\xi_1, \ldots, \xi_n$ *независимы*, если независимы σ алгебры

$$\sigma(\xi_1), \ldots, \sigma(\xi_N),$$

т. е. для всяких борелевских множеств $B_1, \ldots, B_n$ выполнено

$$P(\{\omega\colon \xi_1 \in B_1, \ldots, \xi_n \in B_n\}) = P(\{\omega\colon \xi_1 \in B_1\}) \cdots P(\{\omega\colon \xi_n \in B_n\}).$$

Это равенство достаточно проверять только для интервалов или только для лучей, что сразу дает эквивалентное независимости утверждение, что

$$F(x_1, \ldots, x_n) = F_{\xi_1}(x_1) \cdots F_{\xi_n}(x_n),$$

где $F(x_1, \ldots, x_n) = P(\{\omega\colon \xi_1 \leqslant x_1, \ldots, \xi_n \leqslant x_n\})$ – совместная функция распределения. Кроме того, если распределение каждой случайной величины ξ_k задана плотностью ϱ_{ξ_k}, то независимость равносильна тому, что совместное распределение задано плотностью ϱ и эта плотность имеет вид

$$\varrho(x_1, \ldots, x_n) = \varrho_{\xi_1}(x_1) \cdots \varrho_{\xi_n}(x_n).$$

Распределение Максвелла

Пусть распределение случайного вектора $(\xi_1, \ldots, \xi_n)$ имеет плотность ϱ. Предположим, что выполняются три свойства: (i) плотность ϱ непрерывна и положительна, (ii) сферически симметрична, т. е. зависит только от длины $|x| = \sqrt{x_1^2 + \ldots + x_n^2}$, и (iii) случайные величины ξ_i независимы. Оказывается этих свойств достаточно для определения вида плотности ϱ.

Отметим, что мы ищем плотность с точностью до умножения на ненулевую константу, которая в конце определится исходя из того, что интеграл от плотности должен равняться единице. Поэтому будем пока искать плотность ϱ такую, что $\varrho(0, \ldots, 0) = 1$.

Свойство (iii) влечет равенство

$$\varrho(x_1, \ldots, x_n) = \varrho_1(x_1) \cdots \varrho_n(x_n).$$

Это равенство можно преобразовать к виду

$$\varrho(x_1, \ldots, x_n) = h_1(x_1) \cdots h_n(x_n), \quad h_i(x_i) = \frac{\varrho_i(x_i)}{\varrho_i(0)}, \quad \varrho_1(0) \cdots \varrho_n(0) = \varrho(0, 0, \ldots, 0) = 1.$$

Свойство (ii) означает, что

$$\varrho(x_1, \ldots, x_n) = g(x_1^2 + \ldots + x_n^2).$$

Полагая поочередно все переменные кроме одной равными нулю, выводим равенства

$$h_i(x_i) = g(x_i^2).$$

Следовательно, на положительной полуоси функция g обладает свойством

$$g(t + s) = g(t) \cdot g(s).$$

Так как функция g непрерывна, то последнее свойство вместе с условием $g(0)=1$ определяет вид функции: $g(t)=e^{-\lambda t}$. Здесь $\lambda>0$ так как в итоге надо получить интегрируемую по $\mathbb{R}^n$ функцию. Итак, плотность имеет вид

$$\varrho(x_1,\ldots,x_n)=Ce^{-\lambda(x_1^2+\ldots+x_n^2)}.$$

Таким образом, случайный вектор $(\xi_1,\ldots,\xi_n)$ имеет нормальное распределение. Похожие рассуждения позволили Максвеллу получить явный вид плотности распределения скоростей молекул идеального газа, что фактически стало отправной точкой для развития статистической физики.

$\{\mathcal{A}_i\}$ является последовательностью независимых σ алгебр, если всякий конечный набор элементов этой последовательности является набором независимых σ алгебр.

$\{\xi_i\}$ является последовательностью независимых случайных величин, если всякий конечный набор элементов этой последовательности является набором независимых случайных величин.

Теорема 12.1. *Пусть $\{\mathcal{A}_i\}$ является последовательностью независимых σ алгебр. Предположим, что J_k – попарно непересекающиеся подмножества множества натуральных чисел. Тогда σ алгебры $\mathcal{R}_k$, порожденные множествами из σ алгебр $\mathcal{A}_i$, где $i\in J_k$, независимы.*

Доказательство. Ясно, что доказывать утверждение надо для конечного числа J_k. Более того, достаточно доказать для случая двух J_1 и J_2. Действительно, общий случай доказывается по индукции: если уже доказали для k таких множеств, то переходим к σ алгебре, порожденной множествами из $R_1,\ldots,R_k$, и доказываем ее независимость с R_{k+1}. Аналогично, если мы умеем доказывать независимость R_1 от какой-то одной σ алгебры с индексом из J_2, то на самом деле умеем доказывать, что R_1 и все σ алгебры $\mathcal{A}_i$ с $i\in J_2$ независимы. Значит, для доказательства требуемого утверждения достаточно с точностью до выбора индексов обосновать следующее:

$\mathcal{A}_1$ не зависит от σ алгебры R, порожденной остальными σ алгебрами $\mathcal{A}_i$.

Пусть $A\in\mathcal{A}_1$ и $P(A)>0$. Мы хотим доказать, что для всякого события B из R верно равенство

$$P(A\cap B)=P(A)\cdot P(B)\quad(*).$$

Заметим, что это равенство очевидно выполняется для множеств

$$B=A_{i_1}\cap\ldots\cap A_{i_n},\quad(**)$$

где $A_{i_k}\in\mathcal{A}_{i_k}$ и $i_k>1$. Используя формулу включений и исключений несложно показать, что равенство $(*)$ выполняется для конечных объединений множеств вида $(**)$. Конечные объединения множеств вида $(**)$ образуют алгебру Q, причем эта алгебра содержит все множества из $\mathcal{A}_i$ с $i>1$. Рассмотрим вероятностную меру $P_A(B)=P(B|A)$. На алгебре Q вероятностные меры P и P_A совпадают и, следовательно, совпадают на σ алгебре, порожденной алгеброй Q. □

Результат этой теоремы можно применить к последовательности независимых случайных величин $\{\xi_i\}$. Если разбить эту последовательность на несколько непересекающихся последовательностей $\{x_i^k\}$, то σ алгебры $\sigma(\{\xi_i^k\})$, порожденные каждой из этих последовательностей, будут независимыми.

Теорема 12.2. (Закон 0 и 1 Колмогорова) *Пусть $\{\xi_n\}$ – последовательность независимых случайных величин. Пусть $\mathcal{A}_n$ – σ алгебра порожденная величинами ξ_k с номерами $k\geqslant n$.*

Если $B \in \mathcal{A}_n$ для всякого n, то $P(B) = 0$ или $P(B) = 1$. Если случайная величина η такова, что $\sigma(\eta) \subset \mathcal{A}_n$ для всякого n, то найдется такое числа a, что $\eta = a$ с вероятностью 1.

Доказательство. σ алгебра $\mathcal{A} = \cap \mathcal{A}_n$ независима с последовательностью σ алгебр $\sigma(\xi_n)$. Следовательно, она независима с σ алгеброй $\sigma(\{\xi_n\}_{n=1}^{\infty})$, порожденной этими σ алгебрами. Однако, $\mathcal{A}_n \subset \sigma(\{\xi_n\}_{n=1}^{\infty})$ и, значит, $\mathcal{A}$ независима сама с собой. По определению независимости для каждого события $B \in \mathcal{A}$ выполняется равенство

$$P(B) = P(B)^2,$$

из которого следует, что $P(B) = 0$ или $P(B) = 1$.

Разобьем множество значений ξ на счетное число отрезков $[n, n+1]$. Вероятность попадания значения ξ в каждый из этих отрезков равна 0 или 1. Выбираем отрезок, для которого вероятность равна 1. Теперь делим его пополам и выбираем на каждом шаге деления ту половину, вероятность попадания в которую равна 1. Общая точка всех выбранных отрезков является искомым числом a. $\square$

Следствие 12.2. *Пусть $\{\xi_n\}$ – последовательность независимых случайных величин.*

(i) *Предел $\lim_{n\to\infty} \xi_n$ с вероятностью единица существует или с вероятностью единица не существует, причем в случае существования этот предел является тривиальной случайной величиной.*

(ii) *Последовательности $S_n = \xi_1 + \ldots + \xi_n$ и S_n/n с вероятностью единица сходятся или с вероятностью единица расходятся, причем в случае существования этот предел является тривиальной случайной величиной.*

Доказательство. Для всякого N предел последовательности $\{\xi_n\}_{n=1}^{\infty}$ существует тогда и только тогда, когда существует предел последовательности $\{\xi_n\}_{n=N}^{\infty}$, а последнее событие принадлежит сигма алгебре, порожденной величинами ξ_k с $k \geqslant N$. $\square$

13. Существование случайных величин с заданными распределениями. Генераторы случайных чисел.

Ранее мы показали, что если задано вероятностное распределение μ на $\mathcal{B}(\mathbb{R})$, то можно построить вероятностное пространство и случайную величину с распределением μ. Более того, если заданы два вероятностных распределения μ_1 и μ_2, то можно построить две независимые случайные величины с распределениями μ_1 и μ_2, причем метод построения легко обобщается на любое конечное количество случайных величин. Теперь обсудим построение бесконечной последовательности независимых случайных величин с заданными распределениями.

Начнем с моделирования бесконечного бросания монеты. На вероятностном пространстве $([0,1], \mathcal{B}([0,1]), \lambda)$ определим случайные величины: $\xi_k(\omega)$ – k-й знак в двоичном разложении числа ω. Ясно, что это бернуллиевские случайные величины, принимающие значения 0 или 1 с вероятностью 1/2.

Предложение 13.1. *Случайные величины ξ_k независимы.*

Теперь обсудим простой, но очень важный, способ построения случайной величины ξ с данной функцией распределения F_ξ из случайной величины η, имеющей равномерное распределение.

Напомним, что F_ξ не убывает и непрерывна справа. Положим

$$F_\xi^{-1}(x) = \sup\{t \colon F_\xi(t) < x\}.$$

Если F непрерывна и строго монотонна, то это обычная обратная функция.

Предложение 13.2. *Если случайная величина η имеет равномерное распределение на отрезке $[0,1]$, то случайная величина $\xi = F_\xi^{-1}(\eta)$ имеет функцию распределения F_ξ.*

Доказательство. Заметим, что

$$F_\xi^{-1}(x) \leqslant a \Leftrightarrow x \leqslant F_\xi(a).$$

Следовательно,

$$P(F_\xi^{-1}(\eta) \leqslant a) = P(\eta \leqslant F_\xi(a)) = F_\xi(a).$$

□

Таким образом, достаточно построить последовательность независимых равномерно распределенных величин, а затем из них уже можно изготовить случайные величины с любым наперед заданным распределением.

Предложение 13.3. *Если ξ_k – последовательность независимых бернуллиевских величин, принимающих значения 0 или 1 с вероятностью $1/2$, то случайная величина*

$$\xi(\omega) = \sum_k \frac{\xi_k(\omega)}{2^k}$$

имеет равномерное распределение на отрезке $[0,1]$.

Доказательство. Пусть $t \in [0,1]$ и t_k – цифры двоичного разложения t. Событие $\{\xi(\omega) < t\}$ представляется в виде объединения попарно не пересекающихся событий

$$\{\xi_1 < t_1\} \cap \{\xi_1 = t_1, \xi_2 < t_2\} \cap \{\xi_1 = t_1, \xi_2 = t_2, \xi_3 < t_3\} \cap \ldots$$

Следовательно, используя независимость ξ_k, получаем равенство

$$P(\{\xi(\omega) < t\}) = \frac{t_1}{2} + \frac{1}{2} \cdot \frac{t_2}{2} + \ldots = t.$$

□

Следствие 13.1. *Для произвольной последовательности функций распределения F_n существует последовательность независимых случайных величин ξ_n таких, что ξ_n имеет распределение F_n.*

Доказательство. Строим последовательность независимых бернуллиевских величин η_n. Разбиваем эту последовательность на счетный набор попарно непересекающихся последовательностей η_n^k. По каждой из этих последовательностей строим величину $\xi_k = \sum_n 2^{-n}\eta_n^k$. Наконец, применяем к каждой такой величине свою функцию F_k^{-1} и получаем требуемую последовательность случайных величин. □

Генераторы случайных и псевдослучайных чисел

Как численно генерировать значения случайной величины с заданным распределением?

Сразу заметим, что достаточно научиться генерировать значения равномерно распределенной на $[0,1]$ случайной величины. Применяя к этим значениям обратную функцию к функции распределения, получим значения случайной величины с нужным нам распределением.

Пусть x – значение равномерно распределенной величины на $[0,1]$. Разобьем отрезок $[0,1]$ на две части в отношении 1 к 2. Если число x попало в большую часть отрезка, то пишем 0, а если в меньшую, то пишем 1. В результате мы смоделировали подбрасывание монеты с вероятностью орла $1/3$.

Один из способов получить значение равномерно распределенной величины состоит в использовании представления ее через независимые бернуллиевские величины.

Бросаем три раза правильную монету. Пусть в результате бросания получили 101. Выбираем любое значение из промежутка, в котором двоичная запись чисел начинается с 101, например $0,625$. Это значение равномерно распределенной величины с точностью до $1/16$. Повторяя бросания монеты, получаем серию значений, соответствующую последовательности независимых равномерно распределенных величин.

Как смоделировать случайную величину ξ, описывающую однократное бросание правильной монеты?

Для этого можно использовать, например явление радиоактивного распада. Будем регистрировать появление радиоактивных частиц за время Δt. Если число частиц четно, то значение $\xi = 0$, а если число частиц нечетно, то значение $\xi = 1$.

Число частиц, регистрируемых за время Δt, имеет распределение Пуассона. Следовательно,

$$P(\xi = 0) = \sum_{k=0}^{\infty} \frac{(\lambda\Delta t)^{2k}}{(2k)!} e^{-\lambda\Delta t} = \frac{1}{2}\Big(1 + e^{-2\lambda\Delta t}\Big).$$

Если величина $\lambda\Delta t$ достаточно велика (что накладывает существенные ограничения на скорость работы датчика), то вероятность $P(\xi = 0)$ близка к $1/2$. Для получения набора из N значений независимых бернуллиевских величин надо применить эту схему на непересекающихся интервалах Δt.

Еще один способ сгенерировать такую величину – использовать различные шумы, например случайные колебания напряжения или теплоты. Наблюдаются случайную величину η, порождаемую этим шумом. Выбираем некоторое пороговое значение C и полагаем, что $\xi = 1$ при $\eta > C$ и $\xi = 0$ при $\eta < C$. Конечно основная проблема состоит в таком выборе C, что $P(\eta < C) = 1/2$. Эта проблема возникает и при ручном моделировании ξ подбрасыванием обычной монеты, а именно, мы не можем быть уверены в том, что монета правильная. С этой трудностью можно бороться следующим образом: будем подбрасывать монету N раз. Если число орлов четно, то $\xi = 0$. Если число орлов нечетно, то $\xi = 1$. Пусть вероятность орла равна p. Тогда

$$P(\xi = 0) = \sum_k C_N^{2k} p^{2k} (1-p)^{N-2k}, \quad P(\xi = 1) = \sum_k C_N^{2k+1} p^{2k+1} (1-p)^{N-2k-1}.$$

Заметим, что $((1-p) - p)^N = P(\xi = 0) - P(\xi = 1)$. Следовательно,

$$P(\xi = 0) = \frac{1}{2}\Big(1 - (1-2p)^N\Big).$$

Вероятность $P(\xi = 0) \to 1/2$ при $N \to \infty$.

Описанный способ генерировать равномерно распределенную случайную величину имеет ряд недостатков: относительно долгое время работы и невоспроизводимость данных. В связи с этим очень распространены датчики псевдослучайных чисел, которые кажутся случайными (проходят серию статистических тестов), но на самом деле таковыми не являются. Типичный датчик строит последовательность псевдослучайных чисел следующим образом: $x_{n+1} = f(x_n)$, где f – некоторая фиксированная функция. Естественно выбирать функцию f так чтобы ее график как можно плотнее заполнял единичный квадрат, например $f(x) = \{Mx\}$, где M – достаточно большое целое число, а $\{\,\}$ – дробная часть. Кроме того, несложно показать, что если случайная величина η имеет равномерное распределение на $[0,1]$, то $\{M\eta\}$ также имеет равномерное распределение на $[0,1]$. Отметим еще раз, что последовательность x_n может лишь казаться случайной, но на самом деле полностью определена x_1 и f. Если взять $f(x) = \{Mx\}$ и $x_1 = \frac{1}{d}$, где $d > 1$ – натуральное число, то последовательность имеет вид $x_n = k_n/d$, где

$$k_{n+1} = Mk_n (\operatorname{mod} d).$$

Такой датчик называют мультипликативным. Выбор коэффициентов M и d является трудной задачей. Ясно, что последовательность k_n зацикливается и надо стараться делать цикл длиннее и следить, что последовательность пробегает все значения от 1 до $d-1$ до зацикливания. Кроме того, последовательность должна проходить различные статистические тесты, проверяющие ее «случайность».

14. Математическое ожидание дискретных случайных величины.

Пусть ξ – дискретная случайная величина на вероятностном пространстве $(\Omega, \mathcal{A}, P)$, принимающая значения $\{x_1, \ldots, x_n\}$ (предполагаем, что значения перечислены без повторений). Полагая $A_i = \xi^{-1}(\{x_i\})$ случайную величину можно записать в виде

$$\xi = x_1 \mathbb{I}_{A_1} + \ldots x_n \mathbb{I}_{A_n}.$$

Математическим ожиданием ξ называют число

$$\mathbb{E}\xi = x_1 P(A_1) + \ldots + x_n P(A_n).$$

Если случайная величина ξ имеет вид $\xi = \sum_j y_j \mathbb{I}_{B_j}$, где B_j попарно не пересекаются, но значения y_j могут повторяться, то математическое ожидание можно считать по формуле

$$\mathbb{E}\xi = y_1 P(B_1) + \ldots + y_n P(B_n).$$

Действительно, мы можем сгруппировать вместе значения $y_j = x_i$, а соответствующие события B_j объединяться в событие A_i.

Далее, пока не будет сказано обратное, мы предполагаем, что ξ и η – дискретные величины с конечным множеством значений.

Если некоторое свойство имеет место с вероятностью единица, то говорят, что оно имеет место почти наверное.

Теорема 14.1. (i) $\mathbb{E}(\alpha\xi + \beta\eta) = \alpha\mathbb{E}\xi + \beta\mathbb{E}\eta$ *(математическое ожидание линейно),*

(ii) *если* $\xi \geqslant 0$ *почти наверное, то* $\mathbb{E}\xi \geqslant 0$ *(математическое ожидание монотонно).*

(iii) *если* $\xi \geqslant 0$ *почти наверное и* $\mathbb{E}\xi = 0$, *то* $\xi = 0$ *почти наверное.*

Доказательство. Из всех трех утверждений теоремы более или менее неочевидно только первое. Пусть $\xi = \sum_i a_i \mathbb{I}_{A_i}$ и $\eta = \sum_j b_j \mathbb{I}_{B_j}$. Тогда

$$\alpha\xi + \beta\eta = \sum_{i,j} (\alpha a_i + \beta b_j) \mathbb{I}_{A_i \cap B_j}.$$

Следовательно, математическое ожидание $\mathbb{E}(\alpha\xi + \beta\eta)$ равно

$$\sum_{i,j} (\alpha a_i + \beta b_j) P(A_i \cap B_j) = \alpha \sum_i a_i \Big(\sum_j P(A_i \cap B_j) \Big) + \beta \sum_j b_j \Big(\sum_i P(A_i \cap B_j) \Big).$$

Правая часть этого равенства совпадает с $\alpha\mathbb{E}\xi + \beta\mathbb{E}\eta$. □

Проиллюстрируем доказанную теорему примером.

Расстановка ладей

В каждой клетке шахматной доски стоит плюс или минус, причем число плюсов равно числу минусов. Покажем, что можно так расставить восемь ладей, что они не бьют друг-друга и не менее четырех из них стоят на плюсах.

Будем расставлять ладьи случайным образом так, что они не бьют друг-друга. Такие расстановки равновероятны. С каждой ладьей свяжем случайную величину X_i, которая равна

1, если i-я ладья стоит на клетке с плюсом, и равна 0, если i-я ладья стоит на клетке с минусом. Количество расстановок, в которых ладья i стоит на клетке с плюсом, равно количеству расстановок, в которых ладья i стоит на клетке с минусом. Это немедленно следует из того, что количество расстановок, в которых ладья i стоит в некоторой фиксированной клетки, не зависит от этой клетки (их всегда 7!). Следовательно, $P(X_i = 1) = P(X_i = 0) = 1/2$. Так как

$$\mathbb{E}(X_1 + \ldots + X_8) = \mathbb{E}X_1 + \ldots + \mathbb{E}X_8 = 4.$$

Если всегда $X_1 + \ldots + X_8 < 4$, то на самом деле всегда $X_1 + \ldots + X_8 \leqslant 3$ и $\mathbb{E}(X_1 + \ldots + X_8) \leqslant 3$, что противоречит только что сделанным вычислениям. Следовательно, существует расстановка, когда $X_1 + \ldots + X_8 \geqslant 4$.

Следствие 14.1. *Имеет место оценка*

$$|\mathbb{E}\xi| \leqslant \mathbb{E}|\xi|.$$

Доказательство. Достаточно заметить, что $-|\xi| \leqslant \xi \leqslant |\xi|$, и применить утверждение (ii) теоремы. □

Следствие 14.2. (Неравенство Коши–Буняковского) *Имеет место неравенство*

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

причем равенство возможно тогда и только тогда, когда найдутся числа α и β такие, что $\alpha\xi + \beta\eta = 0$ почти наверное.

Доказательство. Если $\mathbb{E}\xi^2 = 0$, то утверждение очевидно верно. Пусть $\mathbb{E}\xi^2 > 0$. Рассмотрим квадратичную функцию

$$p(t) = \mathbb{E}(\eta - t\xi)^2 = \mathbb{E}\eta^2 - 2t\mathbb{E}(\eta\xi) + t^2\mathbb{E}\xi^2.$$

Из неотрицательности $p(t)$ для всех t следует, что

$$D = 4\Big(\mathbb{E}(\eta\xi)\Big)^2 - 4\mathbb{E}\eta^2 \cdot \mathbb{E}\xi^2 \leqslant 0.$$

Последнее неравенство равносильно требуемому. □

Следствие 14.3. (Неравенство Чебышева) *Если $\xi \geqslant 0$, то для всякого $C > 0$ верно неравенство*

$$P(\xi \geqslant C) \leqslant \frac{1}{C}\mathbb{E}\xi.$$

Доказательство. Пусть $A = \{\omega\colon \xi \geqslant C\}$. Тогда $\xi \geqslant C\mathbb{I}_A$ и значит

$$\mathbb{E}\xi \geqslant CP(A).$$

□

Теорема 14.2. *Если случайные величины ξ и η независимы, то*

$$\mathbb{E}(\xi\eta) = \mathbb{E}\xi \cdot \mathbb{E}\eta.$$

Доказательство. Пусть $\xi = \sum_i a_i\mathbb{I}_{A_i}$ и $\eta = \sum_j b_j\mathbb{I}_{B_j}$. Тогда

$$\xi\eta = \sum_{i,j} a_i b_j \mathbb{I}_{A_i \cap B_j}.$$

Следовательно, математическое ожидание $\mathbb{E}(\xi\eta)$ равно

$$\sum_{i,j} a_i b_j P(A_i \cap B_j).$$

Теперь воспользуемся независимостью ξ и η и преобразуем последнюю сумму:

$$\sum_{i,j} a_i b_j P(A_i \cap B_j) = \sum_{i,j} a_i b_j P(A_i)P(B_j) = \Big(\sum_i a_i P(A_i)\Big)\Big(\sum_j b_j P(B_j)\Big) = \mathbb{E}\xi \cdot \mathbb{E}\eta.$$

□

Математическое ожидание отвечает за среднее значение случайной величины. Теперь обсудим величину, которая отвечает за разброс значений случайной величины.

Дисперсией случайной (пока дискретной с конечным множеством значений) величины ξ называется число

$$\mathbb{D}\xi = \mathbb{E}(\xi - \mathbb{E}\xi)^2.$$

Предложение 14.1. (i) *Если $\mathbb{D}\xi = 0$, то $\xi = \mathbb{E}\xi$ почти наверное.*

(ii) *Для всякого числа α верно равенство $\mathbb{D}(\alpha\xi) = \alpha^2\mathbb{D}\xi$.*

(iii) *Если ξ и η независимы, то $\mathbb{D}(\xi + \eta) = \mathbb{D}\xi + \mathbb{D}\eta$.*

Доказательство. Эти свойства немедленно следуют из свойств математического ожидания. □

Для описания зависимости между случайными величинами ξ и η используют величину

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

которую называют *ковариацией* и величину

$$r = \frac{\operatorname{cov}(\xi, \eta)}{\sqrt{\mathbb{D}\xi}\sqrt{\mathbb{D}\eta}},$$

которую называют *коэффициентом корреляции.*

Рассмотрим линейное пространство случайных величин (пока дискретных с конечным множеством значений) с нулевым математическим ожиданием. На этом пространстве введем скалярное произведение

$$\langle \xi, \eta \rangle = \mathbb{E}(\xi\eta).$$

Соответственно появляется норма

$$\|\xi\| = \sqrt{\langle \xi, \xi \rangle} = \sqrt{\mathbb{E}\xi^2} = \sqrt{\mathbb{D}\xi}.$$

Коэффициент ковариации на случайных величинах с нулевым математическим ожиданием совпадает с скалярным произведением, а коэффициент корреляции имеет вид

$$r = \frac{\langle \xi, \xi \rangle}{\|\xi\|\|\eta\|}.$$

Таким образом, коэффициент корреляции – косинус угла между векторами ξ и η.

Несложно проверить, что из независимости ξ и η следует равенство нулю ковариации и коэффициента корреляции. Если же коэффициент корреляции равен по модулю единице, то (из неравенства Коши–Буняковского) имеет место линейная зависимость между ξ и η. Важно отметить, что из равенства нулю ковариации не следует независимость.

Рассмотрим пример работы с математическим ожиданием.

Модель Эрдеша–Реньи случайного графа и надежность сети

Пусть V_n – конечное множество $\{1, 2, \ldots, n\}$, элементы которого мы называем вершинами. Будем проводить между двумя различными вершинами ребро (только одно) с вероятностью p независимо от остальных пар вершин. Получающийся граф будем называть случайным графом в модели Эрдеша–Реньи. Множество элементарных исходов Ω состоит из C_n^2 ребер. Событием называется любое подмножество $E \subset \Omega$. Вероятность E задается формулой

$$P(E) = p^{|E|}(1-p)^{C_n^2 - |E|}.$$

Можно представлять, что изображен полный граф на n вершинах, но ребра проведены пунктиром. Мы обводим некоторый набор ребер, т. е. выбираем подмножество в множестве всех ребер полного графа.

Нас интересует вероятность того, что получающийся граф связен, т. е. для любых двух вершин есть соединяющий их путь. Имеет место следующий замечательный результат, который иногда называют теоремой о надежности сети.

Если $p = \frac{c \ln n}{n}$, то граф связен с вероятностью единица при $c > 1$ и не является связным с вероятностью единица при $c < 1$

Мы докажем более грубое утверждение, но с меньшим числом технических деталей.

Предложение 14.2. *Если $p = \frac{c \ln n}{n}$ и $c > 2$, то граф связен с вероятностью единица.*

Доказательство. Пусть случайная величина X_n – число компонент связности в случайном графе G, если граф не является связным, и $X_n = 0$ в случае связности G. Нам надо доказать, что $P(X_n > 1) \to 0$ при $n \to \infty$. По неравенству Чебышева

$$P(X_n > 1) \leqslant \mathbb{E}X_n.$$

Следовательно, достаточно доказать стремление к нулю $\mathbb{E}X_n$. Пусть $K_1, \ldots, K_{C_n^k}$ – все k-элементные подмножества V_n. Через $X_{n,k,i}$ обозначим случайную величину, которая равна единице в случае, когда K_i является связной компонентой, и равна нулю в случае, когда это не так. Ясно, что

$$\mathbb{E}X_n = \sum_{k=1}^{n-1} \sum_{i=1}^{C_n^k} \mathbb{E}X_{n,k,i}.$$

Заметим, что $\mathbb{E}X_{n,k,i} = P(X_{n,k,i} = 1)$, а эта вероятность оценивается через вероятность того, что вершины из множества K_i не соединены ребрами с вершинами из $V_n \setminus K_n$. Пусть $q = 1 - p$. Имеет место оценка

$$\mathbb{E}X_n \leqslant \sum_{k=1}^{n-1} C_n^k q^{k(n-k)}$$

Эта сумма симметрична и удваивая ее можно считать, что суммирование идет по $k \leqslant n/2$. При таких k имеет место неравенство $k(n-k) \geqslant kn/2$. Добавляя недостающие слагаемые мы можем написать

$$\sum_{k=1}^{n-1} C_n^k q^{k(n-k)} \leqslant 2 \sum_{k=1}^{n-1} C_n^k (q^{n/2})^k = 2(1 + q^{n/2})^n - 2 - 2q^{n^2/2}.$$

По условию $q = 1 - \frac{2a \ln n}{n}$, где $a > 1$. Имеем

$$q^{n/2} = e^{2^{-1} n \ln(1 - \frac{2a \ln n}{n})} = e^{-a \ln n + \beta_n} = \frac{1}{n^a} e^{\beta_n}, \quad \beta_n \to 0.$$

Следовательно,

$$(1 + q^{n/2})^n = \left(1 + \frac{1}{n^a} e^{\beta_n}\right)^n \to 1$$

и

$$2(1 + q^{n/2})^n - 2 - 2q^{n^2/2} \to 0.$$

Таким образом, $\mathbb{E}X_n \to 0$ и теорема доказана. $\square$

15. Математическое ожидание: общий случай.

Пусть множество значений $\{x_1, x_2, \dots\}$ дискретной величины ξ бесконечно. Положим $A_i = \xi^{-1}(\{x_i\})$. Будем говорить, что у ξ существует конечное математическое ожидание, если сходится ряд

$$\sum_{k=1}^{\infty} |x_k| P(A_k),$$

и число

$$\mathbb{E}\xi = \sum_{k=1}^{\infty} x_k P(A_k)$$

называется математическим ожиданием ξ. Напомним, что слагаемые абсолютно сходящегося ряда можно группировать и переставлять, а произведение двух абсолютно сходящихся рядов равно ряду из всех возможных попарных произведений слагаемых этих рядов, взятых в произвольном порядке. Используя это легко можно показать, что доказанные выше утверждения без изменений остаются в силе. Заметим только, что в неравенстве Коши–Буняковского, определениях дисперсии, ковариации и корреляции надо дополнительно требовать $\mathbb{E}\xi^2 < \infty$ и $\mathbb{E}\eta^2 < \infty$. Доказательства повторяются практически дословно. Единственное существенное дополнение требуется в теореме о математическом ожидании произведения независимых величин. Доказательство надо провести задом наперед:

$$\mathbb{E}\xi \cdot \mathbb{E}\eta = \Big(\sum_i a_i P(A_i)\Big)\Big(\sum_j b_j P(B_j)\Big) =$$
$$\sum_{i,j} a_i b_j P(A_i \cap B_j) = \sum_{i,j} a_i b_j P(A_i) P(B_j) =$$
$$\sum_{i,j} a_i b_j P(A_i \cap B_j) = \sum_{i,j} a_i b_j P(A_i \cap B_j) = \mathbb{E}(\xi\eta),$$

где существование последнего выражения следует из абсолютной сходимости ряда, являющегося произведением абсолютно сходящихся рядов.

Всякую случайную величину ξ можно равномерно приблизить последовательностью дискретных случайных величин:

$$\xi_n = 10^{-n}[10^n \xi].$$

Действительно, $0 \leqslant x - [x] \leqslant 1$ и $0 \leqslant \xi - \xi_n \leqslant 10^{-n}$.

Предположим, что для каждого n существует математическое ожидание $\mathbb{E}\xi_n$. Тогда числа $\mathbb{E}\xi_n$ образуют фундаментальную последовательность так как

$$|\mathbb{E}\xi_n - \mathbb{E}\xi_m| \leqslant \sup |\xi_n - \xi_m|.$$

Следовательно, существует предел $\lim_{n\to\infty} \mathbb{E}\xi_n$. Несложно проверить, что этот предел не зависит от выбора последовательности ξ_n. Математическим ожиданием ξ называется число

$$\mathbb{E}\xi = \lim_{n\to\infty} \mathbb{E}\xi_n.$$

Пусть $\xi_n = 10^{-n}[10^n \xi]$. Равенство $\xi_n(\omega) = k10^{-n}$ выполняется в точности для тех ω, для которых $k10^{-n} \leqslant \xi(\omega) < (k+1)10^{-n}$. Следовательно, имеет место равенство

$$\mathbb{E}\xi = \lim_{n\to\infty} \sum_k \frac{k}{10^n} P(\{\omega\colon k10^{-n} \leqslant \xi(\omega) < (k+1)10^{-n}\}).$$

Для математического ожидания используют еще одно обозначение

$$\int_\Omega \xi(\omega) P(d\omega).$$

Заметим, что у ограниченных случайных величин ξ всегда существует математическое ожидание. Действительно, в этом случае приближающие величины $\xi_n = 10^{-n}[10^n\xi]$ имеют конечное множество значений и конечное математическое ожидание. Единственное, что может помешать существованию математического ожидания – это отсутствие математического ожидания у приближающих дискретных величин.

Предложение 15.1. *Математическое ожидание* $\mathbb{E}\xi$ *существует тогда и только тогда, когда существует* $\mathbb{E}|\xi|$.

Доказательство. Если ξ_n – последовательность дискретных величин, равномерно приближающих ξ, то $|\xi_n|$ равномерно приближает ξ. Кроме того, если существует математическое ожидание у дискретной величины η, то по определению существует математическое ожидание у $|\eta|$. Вероно и обратное, если существует математическое ожидание $\mathbb{E}|\eta|$, то существует математическое ожидание $\mathbb{E}\eta$. Для дискретных величин это утверждение является частью определения. Так как $|[x]| \leqslant [|x|]+1$, то $|10^{-n}[10^n\eta]| \leqslant 10^{-n}[10^n|\eta|]+10^{-n}$. Следовательно, если существует конечное математическое ожидание у $10^{-n}[10^n|\eta|]$, то существует конечное математическое ожидание у $10^{-n}[10^n\eta]$. Остается отметить, что последовательность $\eta_n = 10^{-n}[10^n\eta]$ равномерно сходится к η. □

Таким образом, вопрос существования математического ожидания достаточно прояснить для неотрицательных случайных величин.

Предложение 15.2. *Пусть* $\xi \geqslant 0$. *Математическое ожидание* $\mathbb{E}\xi$ *существует тогда и только тогда, когда для некоторого* n *сходится ряд*

$$\sum_{k=1}^{\infty} \frac{k}{10^n} P(\{\omega\colon k10^{-n} \leqslant \xi(\omega) < (k+1)10^{-n}\})$$

Доказательство. Легко видеть, что для $\xi \geqslant 0$ верно неравенство

$$10^{-n}[10^n\xi] \leqslant 10^{-n-1}[10^{n+1}\xi] \leqslant 10^{-n}[10^n\xi] + 10^{-n}$$

. Следовательно, если существует математическое ожидание $\mathbb{E}\xi_n$, $\xi_n = 10^{-n}[10^n\xi]$, хотя бы для одного n, то существует математическое ожидание $\mathbb{E}\xi_n$ сразу для всех n. □

Следствие 15.1. *Если* $|\xi| \leqslant \eta$ *и существует* $\mathbb{E}\eta$, *то существует* $\mathbb{E}\xi$.

Доказательство. Пусть $\xi \geqslant 0$. Имеем $10^{-n}[10^n\xi] \leqslant 10^{-n}[10^n\eta]$. Остается заметить, что существует $\mathbb{E}10^{-n}[10^n\eta]$. □

Отметим, что прямо из определения математического ожидания вытекает, что

$$\mathbb{E}10^{-n}[10^n\xi] \leqslant \mathbb{E}\xi \leqslant \mathbb{E}10^{-n}[10^n\xi] + 10^{-n}.$$

Предложение 15.3. *Пусть* $\xi \geqslant 0$ *и существует математическое ожидание* $\mathbb{E}\xi$. *Тогда*

$$\lim_{N\to\infty} \mathbb{E}\xi\mathbb{I}_{\xi\geqslant N} = 0.$$

Доказательство. Для всякого n имеет место неравенство

$$\mathbb{E}\xi \leqslant 10^{-n} + \sum_{k\geqslant 10^n N} \frac{k}{N} P(\{\omega\colon k10^{-n} \leqslant \xi(\omega) < (k+1)10^{-n}\}).$$

Для всякого $\varepsilon > 0$ выбираем сначала n такое, что $10^{-n} < \varepsilon$, а затем находим N_0 такое, что для всех $N \geqslant N_0$ сумма в правой части неравенства меньше ε. □

Пусть μ_ξ – распределение ξ. Тогда

$$P(\{\omega\colon k10^{-n} \leqslant \xi(\omega) < (k+1)10^{-n}\}) = \mu_\xi([k10^{-n}, (k+1)10^{-n})).$$

Следовательно,

$$\lim_{n\to\infty} \sum_k \frac{k}{10^n} P(\{\omega\colon k10^{-n} \leqslant \xi(\omega) < (k+1)10^{-n}\}) = \lim_{n\to\infty} \sum_k \frac{k}{10^n} \mu_\xi([k10^{-n}, (k+1)10^{-n})),$$

где справа написано определение математического ожидания случайной величины $\eta(x) = x$ на вероятностном пространстве $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu_\xi)$. Таким образом, доказано равенство

$$\mathbb{E}\xi = \int_{\mathbb{R}} x\mu_\xi(dx),$$

простое обобщение которого содержит следующая теорема.

Теорема 15.1. *Пусть f – борелевская функция. Математическое ожидание случайной величины $f(\xi)$ на вероятностном пространстве $(\Omega, \mathcal{A}, P)$ существует тогда и только тогда, когда существует математическое ожидание случайной величины $f(x)$ на $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu_\xi)$. Более того, имеет место равенство*

$$\mathbb{E}f(\xi) = \int_{\mathbb{R}} f(x)\, \mu_\xi(dx).$$

Пусть теперь распределение μ_ξ задается плотностью ϱ_ξ. Напомним, что это означает, что

$$\mu_\xi([a, b]) = \int_a^b \varrho_\xi(x)\, dx$$

для всякого отрезка $[a, b]$. Как и ранее мы предполагаем, что ϱ_ξ интегрируема в несобственном смысле по Риману на $\mathbb{R}$.

Предложение 15.4. *Пусть задана непрерывная или кусочно непрерывная функция f. Тогда математическое ожидание $\mathbb{E}f(\xi)$ существует тогда и только тогда, когда абсолютно сходится несобственный интеграл*

$$\int_{-\infty}^{+\infty} f(x)\varrho_\xi(x)\, dx.$$

Более того, в случае сходимости

$$\mathbb{E}f(\xi) = \int_{-\infty}^{+\infty} f(x)\varrho_\xi(x)\, dx.$$

Доказательство. Предположим, что f является ступенчатой функцией, т. е. существует разбиение прямой на попарно непересекающиеся полуинтервалы Δ_n, на каждом из которых эта функция постоянна и равна m_n. Тогда $f(\xi)$ – дискретная случайная величина и

$$\mathbb{E}f(\xi) = \sum_n m_n P(\xi \in \Delta_n).$$

По условию

$$P(\xi \in \Delta_n) = \int_{\Delta_n} \varrho_\xi(x)\, dx.$$

Следовательно,

$$\mathbb{E}f(\xi) = \sum_n \int_{\Delta_n} f(x)\varrho_\xi(x)\, dx = \int_{-\infty}^{+\infty} f(x)\varrho_\xi(x)\, dx.$$

Остается заметить, что абсолютная сходимость несобственного интеграла равносильна сходимости ряда

$$\sum_n \int_{\Delta_n} |f(x)|\varrho_\xi(x)\, dx = \sum_n |m_n| P(\xi \in \Delta_n).$$

Случай произвольной кусочно-непрерывной f выводится из уже рассмотренного с помощью равномерного приближения f ступенчатой функцией $\square$

Аналогичным образом можно доказать утверждение для случайного вектора $(\xi_1, \ldots, \xi_n)$:

$$\mathbb{E}f(\xi_1, \ldots, \xi_n) = \int_{\mathbb{R}^n} f(x)\varrho(x)\, dx,$$

где ϱ – плотность распределения этого вектора.

В качестве примера рассмотрим вычисление математического ожидания и дисперсии нормально распределенной величины ξ с параметрами μ и σ:

$$\mathbb{E}\xi = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{+\infty} xe^{-\frac{(x-\mu)^2}{2\sigma}}\, dx = \mu + \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{+\infty} ye^{-\frac{y^2}{2\sigma}}\, dy = \mu,$$

$$\mathbb{D}\xi = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{+\infty} (x-\mu)^2 e^{-\frac{(x-\mu)^2}{2\sigma}}\, dx = \sigma^2\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} y^2 e^{-\frac{y^2}{2}}\, dy = \sigma^2,$$

где последнее равенство обосновывается так

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} y\Big(e^{-\frac{y^2}{2}}\Big)'\, dy = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} e^{-\frac{y^2}{2}}\, dy = 1.$$

Поскольку математическое ожидание произвольной ξ определяется как предел математических ожиданий от дискретных случайных величин, то все перечисленные выше свойства математического ожидания доказываются простым предельным переходом с использованием приближения $\xi_n = 10^{-n}[10^n\xi]$. Определения дисперсии, ковариация и коэффициента корреляции совпадают с соответствующими определениями для дискретных величин.

Если случайный вектор (ξ, η) имеет нормальное распределение с плотностью

$$\varrho(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-q^2}} e^{-\frac{1}{2(1-q^2)}\left(\frac{(x-a)^2}{\sigma_1^2} - 2q\frac{(x-a)(x-b)}{\sigma_1\sigma_2} + \frac{(x-b)^2}{\sigma_2^2}\right)},$$

то коэффициент q – это коэффициент корреляции. Легко видеть, что $q = 0$ тогда и только тогда, когда ξ и η независимы. Напомним, что в общем случае из независимости следует равенство нулю коэффициента корреляции, а обратно нет.

Рассмотрим еще один пример. Случайная величина ξ имеет распределение Коши, если ее распределение задано плотностью

$$\varrho_\xi(x) = \frac{1}{\pi(1+x^2)}.$$

Легко видеть, что

$$\int_{-\infty}^{+\infty} \frac{|x|}{\pi(1+x^2)}\, dx$$

расходится и значит у ξ нет математического ожидания.

16. Закон больших чисел.

Ранее было установлено, что для последовательности независимых случайных величин $\xi_1, \xi_2, \ldots$ выражения

$$\frac{\xi_1 + \ldots + \xi_n}{n}$$

либо сходятся с вероятностью единица, либо с вероятностью единица не сходятся. Теперь мы готовы в достаточно общей ситуации показать, что имеет место сходимость и даже найти предел.

Предложение 16.1. *Пусть случайная величина такова, что* $\mathbb{E}\xi^2 < \infty$. *Тогда*

$$P(|\xi - \mathbb{E}\xi| \geqslant \varepsilon) \leqslant \frac{\mathbb{D}\xi}{\varepsilon^2}.$$

Доказательство. Применяем неравенство Чебышева к случайной величине $|\xi - \mathbb{E}\xi|^2$. $\square$

Следствие 16.1. (Закон больших чисел в слабой форме) *Пусть $\{\xi_n\}_n$ – последовательность независимых случайных величин таких, что $\mathbb{E}\xi_n^2 < \infty$. Обозначим $\mathbb{E}\xi_n = \mu_n$ и $\mathbb{D}\xi_n = \sigma_n^2$ Если*

$$\lim_{n\to\infty} \frac{\sigma_1^2 + \ldots + \sigma_n^2}{n^2} = 0,$$

то для всякого $\varepsilon > 0$

$$\lim_{n\to\infty} P\Big(\Big|\frac{\xi_1 + \ldots + \xi_n}{n} - \frac{\mu_1 + \ldots + \mu_n}{n}\Big| \geqslant \varepsilon\Big) = 0$$

Доказательство. Для всякого $\varepsilon > 0$ имеет место неравенство

$$P\Big(\Big|\frac{\xi_1 + \ldots + \xi_n}{n} - \frac{\mu_1 + \ldots + \mu_n}{n}\Big| \geqslant \varepsilon\Big) \leqslant \frac{\sigma_1^2 + \ldots + \sigma_n^2}{n^2\varepsilon^2},$$

из которого немедленно следует доказываемое утверждение. $\square$

Если $\mathbb{E}\xi_1 = \mathbb{E}\xi_2 = \ldots = \mu$ и $\mathbb{D}\xi_1 = \mathbb{D}\xi_2 = \ldots = \sigma^2$, то

$$\frac{\sigma_1^2 + \ldots + \sigma_n^2}{n^2} = \frac{\sigma^2}{n} \to 0$$

и выполняется утверждение следствия:

$$P\Big(\Big|\frac{\xi_1 + \ldots + \xi_n}{n} - \mu\Big| \geqslant \varepsilon\Big) \leqslant \frac{\sigma^2}{n\varepsilon^2} \to 0, \quad n \to \infty.$$

Предположим теперь, что ξ_i – независимые бернуллиевские величины с вероятностью успеха p. Величины $S_n = \xi_1 + \ldots + \xi_n$ – число успехов при n бросаниях монеты. Заметим, что

$$\mathbb{E}\xi_n = p, \quad \mathbb{D}\xi_n = pq, \quad q = 1 - p.$$

Следовательно,

$$P\Big(\Big|\frac{S_n}{n} - p\Big| \geqslant \varepsilon\Big) \leqslant \frac{pq}{n\varepsilon^2} \to 0, \quad n \to \infty.$$

Таким образом, доля успехов при n бросаниях монеты стремится к p в том смысле, что вероятность их отличия более чем на ε стремится к нулю для всякого $\varepsilon > 0$. Говорят, что S_n/n стремится к p по вероятности.

Оценка больших уклонений в схеме Бернулли. Неравенство Чернова.

Пусть $\xi_1, \ldots, \xi_n$ – бернуллиевские независимые величины с вероятностью успеха p. Как и выше, полагаем $S_n = \xi_1 + \ldots + \xi_n$. По неравенству Чебышева

$$P(S_n - pn \geqslant z) \leqslant e^{-\lambda(np+z)}\mathbb{E}e^{\lambda S_n}, \quad z \geqslant 0.$$

Из-за независимости ξ_i правую часть можно преобразовать к виду

$$e^{-\lambda(np+z)}(pe^\lambda + q)^n.$$

Итак,

$$P(S_n - pn \geqslant z) \leqslant \Big[(1 + p(e^\lambda - 1))e^{-\lambda(p+s)}\Big]^n, \quad s = \frac{z}{n}.$$

Выражение в квадратных скобках равно $\mathbb{E}e^{\lambda(\xi_1 - p - s)}$ и является выпуклой функцией по λ. Пусть $\lambda(s)$ – точка минимума этого выражения. Тогда

$$-(p+s)(1 + p(e^\lambda - 1)) + pe^\lambda = 0, \quad e^{\lambda(s)} = \frac{(p+s)q}{p(q-s)}.$$

Подставляя $\lambda(s)$, приходим к равенству

$$(1 + p(e^\lambda - 1))e^{-\lambda(p+s)} = e^{-H(p+s)}, \quad H(x) = x\ln\frac{x}{p} + (1-x)\ln\frac{1-x}{1-p}.$$

Так как $H(p) = H'(p) = 0$ и $H''(x) \geqslant 4$, то $H(p+s) \geqslant 2s^2$. Окончательно получаем неравенство

$$P(S_n - pn \geqslant z) \leqslant e^{-2z^2/n}.$$

Аналогично выводится оценка $P(S_n - pn \leqslant -z) \leqslant e^{-2z^2/n}$. Объединяя эти оценки и полагая $z = n\delta > 0$, получаем известное неравенство Чернова:

$$P\Bigl(\Bigl|\frac{S_n}{n} - p\Bigr| \geqslant \delta\Bigr) \leqslant 2e^{-2n\delta^2}$$

Таким образом, доля успехов очень плотно с ростом n концентрируется около вероятности успеха.

Теорема Вейерштрасса о приближении непрерывной функции многочленом.

Теорема 16.1. (Вейерштрасс) *Пусть $f \in C([0,1])$. Для всякого $\varepsilon > 0$ существует многочлен P_ε такой, что*

$$\max_{[0,1]} |f(x) - P_\varepsilon(x)| < \varepsilon.$$

Доказательство. Пусть S_n – число успехов в n бросаниях монеты с вероятностью успеха p. Пусть $\varepsilon > 0$. Оценим разность

$$|f(p) - \mathbb{E}f(S_n/n)|.$$

Положим $M = \max|f|$. Так как $f(p)$ – константа, то

$$|f(p) - \mathbb{E}f(S_n/n)| \leqslant \mathbb{E}|f(p) - f(S_n/n)|.$$

Найдем $\delta > 0$ такое, что $|f(x) - f(y)| < \varepsilon$ для всех $|x-y| < \delta$. Пусть $A_\delta = \{\omega\colon |p - S_n/n| \geqslant \delta\}$. Имеем

$$\mathbb{E}|f(p) - f(S_n/n)| = \mathbb{E}|f(p) - f(S_n/n)|\mathbb{I}_{A_\delta} + \mathbb{E}|f(p) - f(S_n/n)|\mathbb{I}_{\overline{A}_\delta}.$$

Второе слагаемое меньше ε так как $|p - S_n/n| < \delta$ и, значит, $|f(p) - f(S_n/n)| < \varepsilon$. Оценим первое слагаемое:

$$\mathbb{E}|f(p) - f(S_n/n)|\mathbb{I}_{A_\delta} \leqslant 2MP(A_\delta).$$

Напомним, что $P(A_\delta) \leqslant n^{-1}\delta^{-2}$. Мы получили оценку

$$|f(p) - \mathbb{E}f(S_n/n)| \leqslant \varepsilon + \frac{1}{n\delta^2}.$$

Выбираем n столь большим, что $n^{-1}\delta^{-2} < \varepsilon$, и получаем $\max_p |f(p) - \mathbb{E}f(S_n/n)| < 2\varepsilon$. Остается заметить, что $\mathbb{E}f(S_n/n)$ является многочленом от p. □

Следствие 16.2. *Если непрерывная функция f является 2π периодической, то для всякого ε существует тригонометрический полином*

$$T_N(x) = \sum_{k=-N}^{N} c_k e^{ikx}$$

такой, что $\max|f(x) - T_N(x)| < \varepsilon$.

Доказательство. Во первых заметим, что $\cos^k x$ и $\sin^k x$ являются тригонометрическими полиномами. Покажем, что если функция f четная, то ее можно приблизить многочленом от $\cos x$. Действительно, функция $f(\arccos y)$ является непрерывной функцией на $[0,1]$. По теореме Вейерштрасса существует многочлен P такой, что $\max|f(\arccos y) - P(y)| < \varepsilon$. Тогда $\max_{[0,\pi]} |f(x) - P(\cos x)| < \varepsilon$. Теперь научимся сводить общий случай к уже рассмотренному. Предположим, что f нечетная функция. Тогда $f(x)\sin x$ является четной. Значит мы умеем приближать $f(x)\sin^2 x$, где f – произвольная непрерывная 2π периодическая функция. Поскольку можно приблизить $f(x+\pi/2)\sin^2(x+\pi/2)$, то можно приблизить $f(x)\cos^2 x$.

Так как $f(x) = f(x)\cos^2 x + f(x)\sin^2 x$, то мы научились приближать тригонометрическим многочленом функцию f. $\square$

Линейным преобразованием теорема Вейерштрасса переносится на произвольный отрезок, а ее следствие на любые непрерывные периодические функции с той поправкой, что надо брать $e^{2\pi ikx/T}$, где T – период f.

Лемма 16.1. (Лемма Бореля–Кантелли) *Пусть A_i такие события, что $\sum_i P(A_i) < \infty$. Тогда*

$$P\Big(\cap_k \cup_{n\geqslant k} A_n\Big) = 0.$$

Доказательство. Достаточно заметить, что

$$P\Big(\cap_k \cup_{n\geqslant k} A_n\Big) \leqslant P(\cup_{n\geqslant k} A_n) \leqslant \sum_{n\geqslant k} P(A_n) \to 0, \quad k\to\infty.$$

$\square$

Теорема 16.2. (Закон больших чисел в усиленной форме) *Предположим, что $\{\xi_n\}$ – последовательность независимых одинаково распределенных случайных величин таких, что $\mathbb{E}\xi_n^4 < \infty$. Пусть $\mathbb{E}\xi_n = \mu$. Тогда с вероятностью единица*

$$\lim_{n\to\infty} \frac{\xi_1 + \ldots + \xi_n}{n} = \mu.$$

Доказательство. Положим

$$A_{n,r} = \{\omega\colon \Big|\frac{\xi_1 + \ldots + \xi_n}{n} - \mu\Big| \geqslant 1/r\}.$$

Событие, состоящее из исходов, для которых $(\xi_1 + \ldots + \xi_n)/n$ не сходится к μ, представляется в виде

$$\bigcup_{r\geqslant 1} \bigcap_{N\geqslant 1} \bigcup_{n>N} A_{n,r}.$$

Теперь заметим, что по неравенству Чебышева

$$P(A_{n,r}) \leqslant \frac{r^4 \mathbb{E}|(\xi_1 - \mu) + \ldots + (\xi_n - \mu)|^4}{n^4}.$$

Используя независимость ξ_i, получаем

$$\mathbb{E}|(\xi_1 - \mu) + \ldots + (\xi_n - \mu)|^4 = n\mathbb{E}|\xi_1 - \mu|^4 + 6C_n^2(\mathbb{E}|\xi_1 - \mu|^2)^2 \leqslant Cn^2.$$

Следовательно, $P(A_{n,r}) \leqslant Cr^4 n^{-2}$ и ряд $\sum_n P(A_{n,r})$ сходится. По лемме Бореля–Кантелли

$$P(\bigcap_{N\geqslant 1} \bigcup_{n>N} A_{n,r}) = 0.$$

Из этого следует, что событие, состоящее из исходов, для которых $(\xi_1 + \ldots + \xi_n)/n$ не сходится к μ, имеет вероятность 0. $\square$

Равномерное распределение. Метод Монте–Карло. Критерий Вейля.

Пусть ξ_i – последовательность независимых случайных величин, имеющих равномерное распределение на $[0,1]$. Тогда для всякого полуинтервала $[a,b) \subset [0,1]$ через $L_n(\omega)$ обозначим количество чисел в наборе $\xi_1(\omega), \ldots, \xi_n(\omega)$, которые попадают в отрезок $[a,b)$. Из усиленного закона больших чисел следует, что с вероятностью единица

$$\frac{L_n}{n} \to (b-a).$$

Действительно, для последовательности независимых случайных величин $\mathbb{I}_{[a,b)}(\xi_i)$ имеем

$$L_n = \mathbb{I}_{[a,b)}(\xi_1) + \ldots + \mathbb{I}_{[a,b)}(\xi_n), \quad \mathbb{E}\mathbb{I}_{[a,b)}(\xi_i) = b - a.$$

Более того, эти же рассуждения можно применить к произвольной непрерывной функции f, а именно рассмотреть последовательность $f(\xi_i)$. Тогда с вероятностью единица

$$\frac{f(\xi_1)+\ldots+f(\xi_n)}{n} \to \int_0^1 f(x)\,dx.$$

Это наблюдение можно использовать для вычисления интеграла. Такой подход называется методом Монте–Карло.

При моделировании последовательности значений x_i независимых равномерно распределенных случайных величин ξ_i, часто равномерность нас интересует именно с точки зрения частоты попадания членов этой последовательности в заданный полуинтервал $[a,b)$.

Будем говорить, что последовательность (обычная числовая) $\{x_n\}$ чисел из отрезка $[0,1]$ имеет *равномерное распределение*, если

$$\frac{L_n}{n} \to b-a,$$

где L_n – количество чисел среди $x_1,\ldots,x_n$, принадлежащих $[a,b)$.

Предложение 16.2. *Последовательность $\{x_n\}$ имеет равномерное распределение тогда и только тогда, когда для всякой непрерывной функции f на $[0,1]$ верно равенство*

$$\lim_{n\to\infty}\frac{f(x_1)+\ldots+f(x_n)}{n} = \int_0^1 f(x)\,dx.$$

Доказательство. Предположим, что для всякой непрерывной функции f

$$\lim_{n\to\infty}\frac{f(x_1)+\ldots+f(x_n)}{n} = \int_0^1 f(x)\,dx.$$

Достаточно показать, что это равенство верно для $f=\mathbb{I}_{[a,b)}$.

Найдем две непрерывные функции g_ε и h_ε такие, что

$$g_\varepsilon \leqslant \mathbb{I}_{[a,b)} \leqslant h_\varepsilon$$

и

$$(b-a)-\varepsilon \leqslant \int_0^1 g_\varepsilon\,dx \leqslant \int_0^1 h_\varepsilon\,dx \leqslant (b-a)+\varepsilon.$$

Для функций g_ε и h_ε пределы существуют и равны интегралам от них. Остается заметить, что

$$\frac{g_\varepsilon(x_1)+\ldots+g_\varepsilon(x_n)}{n} \leqslant \frac{\mathbb{I}_{[a,b)}(x_1)+\ldots+\mathbb{I}_{[a,b)}(x_n)}{n} \leqslant \frac{h_\varepsilon(x_1)+\ldots+h_\varepsilon(x_n)}{n}.$$

Предположим теперь, что $\{x_n\}$ имеет равномерное распределение. Пусть задано некоторое разбиение $[a,b)$ на полуинтервалы Δ_i и функция f постоянна на каждом таком полуинтервале Δ_i и равна m_i. Пусть L_n^i – количество элементов из набора $x_1,\ldots,x_n$, которые попали в Δ_i. Тогда

$$\frac{f(x_1)+\ldots+f(x_n)}{n} = \sum_i \frac{L_n^i}{n}\cdot m_i \to \sum_i m_i|\Delta_i| = \int_0^1 f(x)\,dx.$$

Для завершения доказательства достаточно заметить, что всякую непрерывную функцию можно равномерно приблизить кусочно постоянной функцией. □

Отметим, что в доказанном утверждении можно дополнительно требовать от функции f условия $f(0)=f(1)$, т. е. считать, что f непрерывная функция с периодом 1.

Следствие 16.3. (Критерий Вейля) *Последовательность $\{x_n\}$ имеет равномерное распределение тогда и только тогда, когда*

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}e^{2\pi i m x_k}=0.$$

для всякого целого $m\neq 0$.

Доказательство. Это утверждение выводится из предыдущего с помощью равномерного приближения произвольной непрерывной 1-периодической функции f на $[0,1]$ тригонометрическим многочленом. □

Важный пример равномерно распределенной последовательности: $\{na\}$, где a – иррациональное число. Для такой последовательности указанная в критерии Вейля сумма вычисляется очень просто – сумма геометрической прогрессии.

Первая цифра степени двойки.

Может ли 2^n начинаться с 3? Каких степеней 2^n больше тех, которые начинаются с 3, или тех, которые начинаются с 7?

Число 2^n начинается с цифры A тогда и только тогда, когда найдется такое целое неотрицательное число k, что

$$A10^k\leqslant 2^n<(A+1)10^k.$$

Эти неравенства можно переписать так $\log_{10}A\leqslant n\log_{10}2-k\leqslant\log_{10}(A+1)$ Так как последовательность $\{n\log_{10}2\}$ равномерно распределена, то частота выполнения этих неравенств стремится к $\log_{10}(1+1/A)$. В частности для тройки и семерки $\log_1 0(1+1/3)>\log_1 0(1+1/7)$ и, следовательно, степени двойки чаще начинаются с тройки.

17. Характеристическая функция и слабая сходимость распределений. ЦПТ.

Пусть задана последовательность случайных величин ξ_n. Всюду далее мы предполагаем, что ξ_n имеют конечный второй момент, причем

$$\sup_n\mathbb{E}\xi_n^2<\infty.$$

Из этого условия с помощью неравенства Чебышева выводится оценка

$$P(|\xi_n|\geqslant A)\leqslant\frac{C}{A^2},$$

где $C=\sup_n\mathbb{E}\xi_n^2$ не зависит от n. В терминах распределений μ_{ξ_n} это означает, что

$$\mu_{\xi_n}(\mathbb{R}\setminus[-A,A])\leqslant CA^{-2}$$

и всегда можно выбрать столь большой отрезок $[-A,A]$, что сразу для всех μ_{ξ_n} мера этого отрезка почти единица, а его дополнения почти ноль.

Пусть кроме последовательности $\{\xi_n\}$ задана еще случайная величина ξ. Будем говорить, что последовательность распределений μ_{ξ_n} *сходится слабо* к распределению μ_ξ, если

$$\mathbb{E}f(\xi_n)=\int_{\mathbb{R}}f(x)\,\mu_{\xi_n}(dx)\to\int_{\mathbb{R}}f(x)\,\mu_\xi(dx)=\mathbb{E}f(\xi)$$

при $n\to\infty$ для всякой непрерывной ограниченной функции f.

Предложение 17.1. *Пусть ξ и η – случайные величины. Если для всякой ограниченной и непрерывной функции f верно равенство $\mathbb{E}f(\xi)=\mathbb{E}f(\eta)$, то $\mu_\xi=\mu_\eta$. Это означает, что последовательность распределений может слабо сходится только к одному распределению.*

Доказательство. Напомним, что функция $F_\xi(t) = \mathbb{E}\mathbb{I}_{(-\infty,t]}(\xi)$ непрерывна справа. Зафиксируем t. Положим

$$g_\delta(x) = \begin{cases} 1, & x \leqslant t, \\ \delta^{-1}(t+\delta-x), & t < x \leqslant t+\delta, \\ 0, & x > t+\delta. \end{cases}$$

Ясно, что $F_\xi(t) \leqslant \mathbb{E}g_\delta(\xi) \leqslant F_\xi(t+\delta)$. Следовательно, $\mathbb{E}g_\delta(\xi)$ стремится к $F_\xi(t)$ при $\delta \to 0$. Это означает, что функции распределения ξ и η совпадают. □

Предложение 17.2. *Пусть F_{ξ_n} и F_ξ – функции распределения случайных величин ξ_n, ξ. Если последовательность μ_{ξ_n} сходится слабо к распределению μ_ξ, то $\lim_{n\to\infty} F_{\xi_n}(t) = F_\xi(t)$ в каждой точке t, в которой F_ξ непрерывна.*

Доказательство. Пусть t – точка непрерывности $F_\xi(t)$. Заметим, что $F_\xi = \mathbb{E}\mathbb{I}_{(-\infty,t]}(\xi)$. Для всякого $\delta > 0$ определим непрерывные функции

$$g_\delta(x) = \begin{cases} 1, & x < t-\delta, \\ \delta^{-1}(t-x), & t-\delta \leqslant x \leqslant t, \\ 0, & x > t. \end{cases} \quad h_\delta(x) = \begin{cases} 1, & x < t \\ \delta^{-1}(t+\delta-x), & t \leqslant x \leqslant t+\delta, \\ 0, & x > t+\delta. \end{cases}$$

Ясно, что

$$\mathbb{I}_{(-\infty,t-\delta]} \leqslant g_\delta(t) \leqslant \mathbb{I}_{(-\infty,t]} \leqslant h_\delta(t) \leqslant \mathbb{I}_{(-\infty,t+\delta]}.$$

Следовательно,

$$\mathbb{E}g_\delta(\xi_n) \leqslant F_{\xi_n}(t) \leqslant \mathbb{E}h_\delta(\xi_n).$$

Устремляя $n \to \infty$ получаем

$$\mathbb{E}g_\delta(\xi) \leqslant \liminf F_{\xi_n}(t) \leqslant \limsup F_{\xi_n}(t) \leqslant \mathbb{E}h_\delta(\xi).$$

Теперь заметим, что

$$F_\xi(t-\delta) \leqslant \mathbb{E}g_\delta(\xi), \quad \mathbb{E}h_\delta(\xi) \leqslant F_\xi(t+\delta).$$

Устремляя $\delta \to 0$ приходим к равенству $\lim_{n\to\infty} F_{\xi_n}(t) = F_\xi(t)$. □

Как и в критерии Вейля произвольную непрерывную ограниченную функцию f можно заменить на e^{ix}.

Характеристической функцией случайной величины ξ называется функция

$$\varphi(x) = \mathbb{E}e^{ix\xi}.$$

Предложение 17.3. *Если для каждого x имеет место равенство*

$$\lim_{n\to\infty} \mathbb{E}e^{ix\xi_n} = \mathbb{E}e^{ix\xi},$$

то распределения μ_{ξ_n} слабо сходятся к μ_ξ.

Доказательство. Пусть f – ограниченная непрерывная функция и $M = \sup|f|$. Пусть $\varepsilon > 0$. Существует такой отрезок $[-A, A]$, что

$$\mu_{\xi_n}([-A,A]) \geqslant 1-\varepsilon, \quad \mu_\xi([-A,A]) \geqslant 1-\varepsilon.$$

Пусть непрерывная функция f_ε совпадает с f на $[-A, A]$, на концах отрезка $[-A-1, A+1]$ равна нулю, а затем продолжена как периодическая функция с периодом $T = 2A+2$. Конечно можно построить f_ε так, что $\max|f_\varepsilon| \leqslant M$. Заметим, что

$$|\mathbb{E}f_\varepsilon(\xi_n) - \mathbb{E}f(\xi_n)| \leqslant 2M\varepsilon, \quad |\mathbb{E}f_\varepsilon(\xi) - \mathbb{E}f(\xi)| \leqslant 2M\varepsilon.$$

Поскольку периодическую функцию f_ε можно равномерно приблизить тригонометрическим многочленом, а для тригонометрических многочленов предел существует и равен правильному значению по условию, то $\lim_{n\to\infty}\mathbb{E}f_\varepsilon(\xi_n)=\mathbb{E}f_\varepsilon(\xi)$. Следовательно,

$$\mathbb{E}f(\xi)-2M\varepsilon\leqslant\liminf\mathbb{E}f(\xi_n)\leqslant\limsup\mathbb{E}f(\xi_n)\leqslant\mathbb{E}f(\xi)+2M\varepsilon.$$

Устремляя $\varepsilon\to 0$ приходим к равенству $\lim_{n\to\infty}\mathbb{E}f(\xi_n)=\mathbb{E}f(\xi)$. □

Повторяя рассуждения с приближением периодической функции тригонометрическим многочленом, можно доказать следующее утверждение.

Предложение 17.4. *Если у двух случайных величин совпадают характеристические функции, то эти величины имеют одинаковые распределения.*

Найдем характеристическую функцию случайной величины ξ, имеющей нормальное распределенной с параметрами 0 и 1.

По определению

$$\varphi(x)=\mathbb{E}e^{ix\xi}=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}\cos(xy)e^{-y^2/2}\,dy.$$

Заметим, что

$$\varphi'(x)=-\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}y\sin(xy)e^{-y^2/2}\,dy=-\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}x\cos(xy)e^{-y^2/2}\,dy=-x\varphi(x).$$

Решая дифференциальное уравнение, находим

$$\varphi(x)=e^{-x^2/2}.$$

Здесь мы учли, что $\varphi(0)=1$.

Предложение 17.5. *Если случайные величины ξ и η независимы, то*

$$\varphi_{\xi+\eta}(x)=\varphi_\xi(x)\cdot\varphi_\eta(x).$$

Доказательство. Немедленно следует из равенств

$$\mathbb{E}e^{ix(\xi+\eta)}=\mathbb{E}e^{ix\xi}e^{ix\eta}=\mathbb{E}e^{ix\xi}\mathbb{E}e^{ix\eta}.$$

□

Предложение 17.6. *Если случайная величина ξ такова, что $\mathbb{E}|\xi|^k<\infty$ для некоторого натурального числа k, то φ имеет непрерывную k-ю производную, и*

$$\varphi^{(k)}(0)=i^k\mathbb{E}\xi^k$$

Доказательство. Заметим, что выполняется оценка

$$\Bigl|\frac{e^{ihy}-1}{h}\Bigr|\leqslant|y|.$$

Рассмотрим выражение

$$\frac{\varphi(x+h)-\varphi(x)}{h}=\mathbb{E}e^{ix\xi}\Bigl(\frac{e^{ih\xi}-1}{h}\Bigr)=\mathbb{E}e^{ix\xi}\Bigl(\frac{e^{ih\xi}-1}{h}\Bigr).$$

Следовательно,

$$\Bigl|\frac{\varphi(x+h)-\varphi(x)}{h}-i\mathbb{E}(e^{ix\xi}\xi)\Bigr|\leqslant\mathbb{E}\Bigl|\frac{e^{ih\xi}-1}{h}-i\xi\Bigr|\mathbb{I}_{|\xi|\leqslant N}+2\mathbb{E}|\xi|\mathbb{I}_{|\xi|>N}.$$

Первое слагаемое в правой части при фиксированном N сходится к нулю из-за равномерной сходимости к нулю при $h\to 0$ выражения

$$\frac{e^{ihy}-1}{h}-iy$$

при $|y| \leqslant N$. Второе слагаемое стремится к нулю при $N \to +\infty$ из-за $\mathbb{E}|\xi| < \infty$. Таким образом,

$$\lim_{h\to 0} \frac{\varphi(x+h)-\varphi(x)}{h} = \mathbb{E}(i\xi e^{ix\xi}).$$

Производные более высокого порядка рассматриваются аналогично. □

Теорема 17.1. (Центральная предельная теорема) *Пусть* $\{\xi_n\}$ *– последовательность независимых одинаково распределенных случайных величин, причем* $\mathbb{E}\xi_1 = \mu$ *и* $\mathbb{D}\xi_1 = \sigma^2$. *Тогда для всех* x

$$\lim_{n\to\infty} P\Big(\frac{\xi_1+\ldots+\xi_n - n\mu}{\sigma\sqrt{n}} \leqslant x\Big) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-y^2/2}\, dy.$$

Доказательство. Переходя к величинам $\xi_i - \mu$ можно далее считать, что $\mu = 0$. Пусть φ – характеристическая функция ξ_1. Тогда характеристическая функция случайной величины

$$\frac{\xi_1+\ldots+\xi_n}{\sigma\sqrt{n}}$$

равна

$$\varphi_n(x) = \Big(\varphi\Big(\frac{x}{\sigma\sqrt{n}}\Big)\Big)^n.$$

По формуле Тейлора

$$\varphi\Big(\frac{x}{\sigma\sqrt{n}}\Big) = 1 - \frac{x^2}{2n} + o(\frac{1}{n}).$$

Следовательно,

$$\varphi_n(x) = \Big(1 - \frac{x^2}{2n} + o(\frac{1}{n})\Big)^n \to e^{-x^2/2}, \quad n \to \infty.$$

Остается заметить, что $e^{-x^2/2}$ – характеристическая функция нормального распределения. □

Теорема Муавра–Лапласа является частным случаем центральной предельной теоремы.

Литература

1. Ширяев А.Н. Вероятность. – М.: Наука, Главная редакция физико-математической литературы, 1980.

2. Тутубалин В.Н. Теория вероятностей и случайных процессов. Учеб. пособие. – М.:Изд-во МГУ, 1992.

3. Ламперти Дж. Вероятность. – М.: Наука, Главная редакция физико-математической литературы, 1973.

4. Чистяков В.П. Курс теории вероятностей. – 5-е издание – М.: Агар, 2000.

5. Феллер В. Введение в теорию вероятностей и ее приложения. Т1,2.-М.:Мир,1964,1967.

6. Лагутин М.Б. Наглядная математическая статистика. 2-е изд., испр. - М.: 2009.

7. Алон Н., Спенсер Дж. Вероятностный метод. – М.: БИНОМ. Лаборатория знаний, 2015.

8. Секей Г. Парадоксы в теории вероятностей и математической статистике. – М., Мир, 1990.

9. Колмогоров А.Н., Журбенко И.Г., Прохоров А.В. Введение в теорию вероятностей. (вып 23 серии «Библиотечка квант»)–М.: Наука, 1982.

10. Шень А. Вероятность: примеры и задачи. – М.: МЦНМО, 2007.

11. Райгородский А.М. Модели случайных графов. – М.: МЦНМО, 2011.

12. Ильинский Д.Г., Райгородский А.М., Скопенков А.Б. Независимость и доказательства существования в комбинаторике. Математическое просвещение, Третья серия. — 2015. — Т. 19. — С. 163–178.

13. Боровков А.А. Теория вероятностей. 2-изд. – М.: Наука, Главная редакция физико-математической литературы, 1986.

14. Бабенко М.А., Левин М.В. Введение в теорию алгоритмов и структур данных. М.: МЦНМО, 2016.

15. Грэхем Р.Л., Кнут Д.Э., Паташник О. Конкретная математика. Основание информатики. — М.: Мир, 1998.

16. Верещагин Н.К., Успенский В.А., Шень А. Колмогоровская сложность и алгоритмическая случайность. М.: МЦНМО, 2013.

17. Коблиц Н. Курс теории чисел и криптографии. М.: Научное изд. «ТВП», 2001.