Казанский федеральный университет
Институт Математики и Механики
имени Н.И. Лобачевского
Кафедра математического анализа

С.Г. Халиуллин

Теория вероятностей

Учебное пособие

Казань — 2021

УДК 519.2

Принято на заседании Учебно-методической комиссии
Института математики и механики им. Н.И. Лобачевского
Протокол 1 от 23-го сентиября 2021 года

Рецензенты:
Доктор физико-математических наук, профессор кафедры
математической статистики И.Н. Володин

Халиуллин С.Г.
Теория вероятностей / С.Г. Халиуллин / Учебное пособие — Казань: Казан. ун-т, 2021. – 113 с.

Курс теории вероятностей основан на лекциях, читаемых автором в Казанском университете для математиков (бакалавров) Института Математики и Механики имени Н.И. Лобачевского. Здесь приведены основные результаты классической и современной теории вероятностей, важные понятия иллюстрированы примерами.

Курс будет полезен как студентам, изучающим математику, так и тем, кто ищет решение многочисленных прикладных задач.

Оглавление

Глава 1

Элементарное введение

1.1 Введение

Возникновение теории вероятностей связано с комбинаторными задачами азартных игр. Отдельные задачи шансов выиграть в азартной игре рассматривались ещё в XV-XVI веках в эпоху итальянского ренессанса (Дж. Кардано, Н. Тарталья, Л. Пачоли). Первые общие методы решения таких задач были даны, по-видимому, в переписке между Б. Паскалем и П. Ферма, начавшейся в 1654 году, а также в первой книге по теории вероятностей «О расчётах в азартной игре», опубликованной Х. Гюйгенсом в 1659 году.

Истинная история теории вероятностей начинается с работы Я. Бернулли «Искусство предположения», опубликованной в 1713 году. Дальнейшее развитие идей и методов теории вероятностей связано с именами А. Муавра, П.-С. Лапласа, С.Д. Пуассона, К.Ф. Гаусса (XVIII-XIX в.в.).

Следующий важный этап развития теории вероятностей связан с именами русских математиков XIX века П.Л. Чебышёва, А.А. Маркова, А.М. Ляпунова.

Основополагающая работа по теории вероятностей «Основные понятия теории вероятностей» была опубликована в 1933 году А.Н. Колмогоровым. XX век можно считать веком взлёта научной мысли, в том числе и прогресса теории вероятностей.

1.2 Случайные явления, частота события

Все явления окружающего мира (природа или социум)мы будем представлять как результат некоторого (мнимого) эксперимента. Конечно, при проведении эксперимента мы всегда обеспечиваем некоторый комплекс условий и не меняем его от эксперимента к эксперименту. Будем различать два вида таких явлений по такому признаку: некоторые из них при вполне определённых условиях эксперимента всегда происходят, другие же происходят не всегда. Первые из них мы назовём детерминированными, вторые — стохастическими (от греческого слова стохос, означающего хаос) или недетерминированными. Предметом теории вероятностей, однако, являются не все стохастические события, а только, так называемые, случайные события, которые характеризуются следующими свойствами:

1. Отсутствие детерминистической регулярности (наблюдения за ними в неизменных условиях не всегда приводят к одним и тем же результатам);

2. Наличие статистической регулярности (она проявляется в статистической устойчивости частот).

Примерами детерминированных событий являются, например, следующие: если подбросить какое-то материальное тело, то оно неизбежно упадёт на землю, и мы даже знаем, что это происходит под действием силы тяжести; если нагреть воду до ста градусов Цельсия, то она превратится в пар.

Классическими примерами случайных событий являются появления «герба» и «решки» при подбрасывании монеты и появление цифр от единицы до шести при подбрасывании игральной кости. С последними событиями мы часто будем встречаться в дальнейшем.

Ещё во времена античной философии наметились два подхода к понятию случайности. Один из них связан с именем Демокрита, который считал, что мир строго детерминирован, а «случайное» отождествлял с непознанным. «Люди сотворили себе кумира из случая как прикрытие присущего им недомыслия», — говорил он. Эпикур же и его последователи считали случайность объективной. Точка зрения Демокрита господствовала в философии вплоть до XIX века.

Вернёмся к упомянутой устойчивости частот и поясним это понятие на простом примере. Пусть «честно» подбрасывается «правильная» монета. Мы, конечно же, отнесём результат этого эксперимента к стохастическим явлениям, то есть заранее не сможем однозначно указать результат опыта. Однако, как было замечено, если провести большое число «независимых» подбрасываний, то частота выпадения, к примеру, «герба» близка к 1/2:

$$\mu_n(\Gamma) = \frac{n_\Gamma}{n} \sim \frac{1}{2},$$

где n_Γ — число опытов, при которых выпал «герб», n — общее число опытов. В этом и заключается статистическая устойчивость частот, которая и делает правдоподобной гипотезу о возможности количественной оценки «случайности». В истории известно несколько проведённых серий таких опытов. Например, Жорж-Луи де Бюффон (XVIII век) бросал монету 4040 раз, при этом получил 2024 «герба», К. Пирсон в 24000 опытах получил 12012 «гербов». Нетрудно видеть, что частоты выпадения «герба» в этих сериях достаточно близко подходят к одной второй.

Оказывается, это явление носит более общий характер: частота $\mu_n(A)$ случайного события A в последовательности повторяемых в одинаковых условиях опытов «приближается» к некоторому числу $p \in [0, 1]$. Считать это число вероятностью события было предложено Р. фон Мизесом в начале XX века.

1.3 Классическое определение вероятности

Вполне естественной кажется идея выделения в эксперименте со случайными исходами (событиями) таких исходов, посредством которых можно было получить любой исход эксперимента. При этом, конечно, нужно, чтобы эти исходы были бы простейшими.

Определение 1.1. *Множеством элементарных исходов (событий) опыта* $\Omega = \{\omega\}$ *называется множество исходов, обладающих следующими свойствами:*
1) все эти элементарные события различны;
2) наступление одного из исходов исключает наступление всех остальных;
3) в ходе эксперимента одно из этих событий неизбежно наступает.

Замечание 1.1. *Множество элементарных событий, как любое другое множество, может быть конечным, счётным или несчётным.*

Замечание 1.2. *Подмножество* A *множества* Ω *(строго говоря, не всякое в общем случае) называется случайным событием. Будем считать, что происходит событие* A*, если происходит какое-то элементарное событие* $\omega \in A$*. Если* $\omega \in A$*, то говорят, что элементарное событие* ω *благоприятствует событию* A*.*

Приведём простые примеры пространств элементарных событий.

Пример 1.1. *Симметричная монета бросается один раз. В этом случае* $\Omega = \{\omega_\Gamma, \omega_P\}$*.*

Пример 1.2. *Правильная игральная кость бросается один раз. Тогда* $\Omega = \{\omega_1, \ldots \omega_6\}$*, где* $\omega_i, (i = \overline{1,6})$*, означает выпадение* i *очков.*

Пример 1.3. *Правильная монета бросается то тех пор, пока не выпадет «герб». Тогда* $\Omega = \{\omega_\Gamma, \omega_{P\Gamma}, \omega_{PP\Gamma}, \ldots\}$*, где, к примеру,* $\omega_{P\Gamma}$ *означает выпадение вначале «решки», а затем «герба».*

Пример 1.4. *На отрезке* $[0,1]$ *случайным образом выбирается точка. Тогда* $\Omega = [0,1]$*.*

Во многих экспериментах с конечным числом элементарных событий из соображений симметрии можно предположить, что все элементарные события равновероятны (можно при этом привлечь частотный подход). Итак,

Определение 1.2. *Пусть*
1. $|\Omega| = N < \infty$ *(здесь* $|A|$ *означает число элементов множества* A*);*
2. Все элементарные исходы равновероятны.

Тогда каждому событию $A \in \Omega$ *ставится в соответствие число*

$$\mathbf{P}(A) = \frac{|A|}{|\Omega|},$$

называемое вероятностью события A*.*

Пример 1.5. *Бросаются две игральные кости. Какова вероятность того, что на обеих костях выпадет одинаковое число очков?*

Решение.Каждый элементарный исход может быть представлен парой чисел $\Omega = \{(i,j) : i,j = \overline{1,6}\}$. Из соображений симметрии ясно, что все элементарные исходы равновероятны и $|\Omega| = 36$.

При этом событие $A = \{(i,j) : i = j; i = \overline{1,6}\}$. Очевидно, что $|A| = 6$. Тогда

$$\mathbf{P}(A) = \frac{6}{36} = \frac{1}{6}.$$

□

С невыполнением требования равновероятности элементарных событий связан известный парадокс де Мере.

Пример 1.6 (Парадокс де Мере)**.** *В результате многократных наблюдений игры в кости шевалье де Мере подметил, что при одновременном подбрасывании трёх игральных костей более часто выпадает комбинация, дающая в сумме 11 очков, чем комбинация, дающая в сумме 12 очков, хотя, с его точки зрения, этих комбинаций одинаковое количество.*

Решение.Де Мере рассуждал так: 11 очков можно получить шестью разными способами (6-4-1, 6-3-2, 5-5-1, 5-4-2, 5-3-3, 4-4-3), и столькими же способами можно получить 12 очков (6-5-1, 6-4-2, 6-3-3, 5-5-2, 5-4-3, 4-4-4). Ошибку де Мере заметил Б.Паскаль, который указал, что рассматриваемые им исходы не являются равновероятными. Например, комбинация (6-4-1) с учётом того, на каких именно костях выпадает определённое количество очков, выпадает, когда наступает один из шести исходов (6,4,1), (6,1,4), (4,6,1), (4,1,6), (1,6,4), (1,4,6), а комбинация (4-4-4) выпадает только в одном случае (4,4,4). □

1.3.1 Элементы комбинаторики

Правило суммы и произведения. *Если объект A можно выбрать n различными способами, объект B можно выбрать m различными способами, то объект A или B можно выбрать $n+m$ различными способами, объект A и B можно выбрать $n \cdot m$ различными способами.*

Пусть $\mathscr{A} = \{a_1, a_2, \ldots, a_n\}$ — множество, содержащее n элементов. Обозначим через $\omega = (i_1, i_2, \ldots, i_r)$ набор из r элементов множества $\mathscr{A}$. Предположим также, что все такие наборы равновероятны. Положим

$$\widetilde{\Omega} = \{\omega = (i_1, i_2, \ldots, i_r) : i_k \in \mathscr{A}, k = \overline{1, r}\},$$

$$\Omega = \{\omega = (i_1, i_2, \ldots, i_r) : i_k \in \mathscr{A}, k = \overline{1, r},\ i_1, i_2, \ldots, i_r\ \}.$$

Множество $\widetilde{\Omega}$ при этом будем называть схемой случайного выбора с возвращением, а множество Ω — схемой случайного выбора без возвращения. Для интерпретации этих схем будем использовать «урновую схему», с которой придётся иметь дело и в будущем. Итак, в некоторой урне содержатся одинаковые шары, пронумерованных от единицы до n. Эксперимент состоит в следующем: мы последовательно вынимаем из урны r шаров и фиксируем его номер. Если при этом каждый вынутый шар будет возвращён в урну, то мы получим схему $\tilde{\Omega}$, в противном случае — схему Ω.

Определение 1.3. *Размещениями с повторениями из n элементов r элементов называется упорядоченный набор длины r, в котором элементы могут повторяться.*

Понятно, что в этом случае используется схема случайного выбора с возвращением. Заметим, что здесь может быть, что $r \geqslant n$. Число таких размещений равно

$$\widetilde{A}_n^r = n^r.$$

Определение 1.4. *Размещениями без повторений из n элементов r элементов называется упорядоченный набор длины r, в котором все элементы различны.*

В этом случае используется схема случайного выбора без возвращения. Число таких размещений равно

$$A_n^r = \frac{n!}{(n-r)!}.$$

Определение 1.5. *Размещения без повторений из n элементов по r элементов, где $r = n$, называется перестановками из n.*

Число таких перестановок равно $n!$.

Если в определениях 1.3 и 1.4 брать неупорядоченные наборы, то говорят о сочетаниях с повторениями и без повторений.

Число сочетаний без повторений из n элементов r элементов равно

$$C_n^r = \frac{n!}{r!(n-r)!}.$$

Число сочетаний с повторениями из n элементов r элементов равно

$$\widetilde{C}_n^r = C_{n+r-1}^r.$$

Пример 1.7. *У человека в кармане находится n ключей, из которых только один подходит к замку́. Он пытается подобрать ключ, причём проверенные ключи он убирает в другой карман. Какова вероятность того, что нужный ключ появится при k-ом ($k = \overline{1, n}$) извлечении?*

Решение.

Элементарными событиями здесь являются последовательности длины n, состоящие из различных и неповторяющихся ключей, то есть, перестановки из n. Значит, $|\Omega| = n!$. Последовательности, в которых на k-ом месте стоит нужный ключ являются благоприятствующими нашему искомому событию A. Ясно что $|A| = (n-1)!$. Тогда

$$\mathbf{P}(A) = \frac{(n-1)!}{n!} = \frac{1}{n}.$$

Интересно, что эта вероятность не зависит от k. □

Пример 1.8 (гипергеометрическое распределение)**.** *В урне тщательно перемешано N одинаковых шаров, среди которых M белых и $(N - M)$ чёрных. Наудачу извлекается комплект из n шаров (без учёта порядка извлечения). Вычислить вероятность того, что среди извлечённых шаров окажется ровно m белых.*

Решение.

Очевидно, что элементарными событиями здесь являются сочетания без повторений из N по n. Таким образом, $|\Omega| = C_N^n$. Благоприятствующие элементарные события состоят в том, мы вынимаем ровно m белых из M и ровно $n-m$ чёрных шаров из $N-M$. Число таких исходов равно $C_M^m \cdot C_{N-M}^{n-m}$. Тогда вероятность искомого события равна

$$p_m = \frac{C_M^m \cdot C_{N-M}^{n-m}}{C_N^n}, \quad 0 \leqslant m \leqslant \min(n, M).$$

Набор вероятностей (p_m) будем называть гипергеометрическим распределением. $\square$

1.4 Геометрические вероятности

Следуя классическому определению, мы находим вероятность события как долю тех элементарных событий, которые приводят к наступлению события, среди всех элементарных событий. Аналогичный подход можно применять в более сложной ситуации, когда пространство элементарных исходов не является конечным или счётным множеством. Уточним постановку. Пусть Ω является ограниченным подмножеством n-мерного евклидова пространства, в котором задана конечная мера μ. Рассмотрим в Ω систему подмножеств $\mathscr{F}$, измеримых относительно этой меры с смысле Жордана. Тогда для $\forall A \in \mathscr{F}$ рассмотрим событие, состоящее в том, что наудачу взятая точка из Ω попадёт в A. Это событие будем обозначать той же буквой A. Положим

$$\mathbf{P}(A) = \frac{\mu(A)}{\mu(\Omega)}.$$

В этом случае говорят, что на множестве Ω задано равномерное распределение (некоторый аналог равновероятности элементарных событий в конечном множестве). Смысл же такого определения состоит в том, что события имеют одинаковую вероятность, если они имеют одинаковую геометрическую меру. Естественно, что такой схеме выбор математической модели уже не столь очевиден, как это было в классической модели. Далее мы приведём пример, из которого будет следовать, что вероятность события будет зависеть от выбранной математической модели.

Пример 1.9 (Парадокс Бертрана). *В круге радиуса R случайно проведена хорда. Обозначим через ξ её длину. Найти вероятность $\mathbf{P}\{\xi < R\}$.*

Решение.

Рассмотрим три случая того, как мы будем понимать выражение «случайно проведена хорда». Это даст нам три различные математические модели.

Случай 1. Середину хорды совместим со случайно выбранной точкой в круге (часто говорят при этом, что середина хорды имеет равномерное распределение в круге). В этом случае $\Omega = \{(x, y) : x^2 + y^2 \leqslant R^2\}$, $\{\xi < R\} = \{(x, y) : \frac{3}{4}R^2 \leqslant x^2 + y^2 \leqslant R^2\}$. Значит,

$$\mathbf{P}\{\xi < R\} = 1/4.$$

Случай 2. Направление хорды фиксировано, а её середина равномерно распределена на диаметре, перпендикулярном направлению хорды.

В этом случае $\Omega = [0, 2R]$. Решая несложную геометрическую задачу, получим

$$\mathbf{P}\{\xi < R\} = 1 - \sqrt{3}/2.$$

Случай 3. Один конец хорды фиксирован, а второй конец равномерно распределён на окружности. Здесь $\Omega = \{(x, y) : x^2 + y^2 = R^2\}$. Легко видеть, что в этом случае

$$\mathbf{P}\{\xi < R\} = 1/3.$$

□

Пример 1.10 (Игла Бюффона)**.** *Плоскость расчерчена параллельными прямыми, расстояние между которыми равно $2a$. На неё наудачу брошена игла длины $2l$, $(l < a)$. Найти вероятность того, что игла пересечёт какую-то прямую.*

Решение.

Пусть u — расстояние от середины иглы до ближайшей прямой, φ — угол наклона иглы к положительному направлению прямой. Пара чисел (u, φ) задаёт положение прямой с точностью до конкретной прямой. Тогда в качестве пространства элементарных событий возьмём

$$\Omega = \{(u, \varphi) : 0 \leqslant u \leqslant a, 0 \leqslant \varphi \leqslant \pi\}.$$

Пересечение иглы с прямой произойдёт тогда, когда $u \leqslant l \sin\varphi$. Интересующее нас событие

$$A = \{(u, \varphi) : 0 \leqslant u \leqslant l \sin\varphi, 0 \leqslant \varphi \leqslant \pi\}.$$

Нетрудно вычислить, что

$$\mathbf{P}(A) = \frac{2l}{\pi a}.$$

Этот пример интересен тем, что позволяет, объединяя частотную интерпретацию вероятности и геометрическую вероятность, приблизительно посчитать значение числа π. С этой целью игла бросается n раз. Пусть в m случаях произошло событие A. Тогда из приближенного равенства $\mathbf{P}(A) \approx \dfrac{m}{n}$ получим

$$\pi \approx 2 \cdot \frac{l}{a} \cdot \frac{n}{m}.$$

□

1.5 Понятие вероятностного пространства

Теперь мы обобщим все те интуитивные представления о случайных событиях и их вероятностях.

Определение 1.6. *Пространством элементарных событий* Ω *назовём произвольное непустое множество. Элементы этого множества будем называть элементарными событиями.*

Замечание 1.3. *Напомним, что в реальном опыте элементарным событиям соответствуют взаимоисключающие друг друга исходы этого опыта.*

В дальнейшем мы будем различать два типа вероятностных пространств — дискретные и непрерывные.

I. **Дискретное вероятностное пространство.**

Пусть $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n, \ldots\}$ — конечное или счётное множество, $\mathscr{F}$ — множество всех подмножеств множества Ω. Пусть далее $(p_n)_{n=1}^{\infty}$ — неотрицательные числа, такие что

$$\sum_{n=1}^{\infty} p_n = 1.$$

При этом каждому ω_n мы поставим в соответствие число p_n, $n = 1, 2, \ldots,$ полагая $p_n = \mathbf{P}(\{\omega_n\})$. Теперь для каждого $A \in \mathscr{F}$ положим

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

В этом случае тройка $(\Omega, \mathscr{F}, \mathbf{P})$ называется дискретным вероятностным пространством.

II. **Непрерывное вероятностное пространство.**

Пусть в общем случае $\Omega = \mathbb{R}^n$ — n-мерное евклидово пространство, $p = p(x_1, x_2, \ldots, x_n)$ — неотрицательная измеримая по Риману функция на Ω. Будем предполагать существование n-мерного интеграла от функции $p = p(x_1, x_2, \ldots, x_n)$ и

$$\int \underset{\mathbb{R}^n}{\cdots} \int p(x_1, x_2, \ldots, x_n)\, dx_1 \ldots dx_n = 1.$$

Через $\mathscr{F}$ обозначим некоторый класс подмножеств множества Ω, являющихся измеримыми по Жордану. Положим для любого $A \in \mathscr{F}$

$$\mathbf{P}(A) = \int\limits_{A} \cdots \int p(x_1, x_2, \ldots, x_n)\, dx_1 \ldots dx_n.$$

Тройка $(\Omega, \mathscr{F}, \mathbf{P})$ в этом случае называется непрерывным вероятностным пространством.

Замечание 1.4. *Как в случае дискретного вероятностного пространства, так и в случае непрерывного, элементы класса $\mathscr{F}$ подмножеств множества Ω называются событиями.*

1.6 Операции над событиями

Для абстрактного вероятностного пространства, введённого в предыдущем параграфе, алгебраические операции над событиями соотнесём с алгебраическими операциями над множествами. Приведём здесь соответствующую таблицу.

Теория множеств	Теория вероятностей	Название	Интерпретация
Ω	Ω	Достоверное событие	Неизбежно наступает
$\emptyset$	$\emptyset$	Невозможное событие	Никогда не наступает
$A \cap B$	$A \cdot B$	Произведение событий	Наступают вместе
$A \cup B$	$A + B$, $A \cup B$	Сумма событий	Наступает либо A, либо B, либо оба вместе
$A \subset B$	$A \subset B$	A влечёт B	B наступает всякий раз, когда наступает A
$A \setminus B$	$A \setminus B$	Разность событий	Наступает A и не наступает событие B
$A = B$	$A = B$	События эквивалентны	A наступает тогда и только тогда, когда наступает B
$\overline{A}$	$\overline{A}$	Противоположное событие	A наступает тогда и только тогда, когда не наступает $\overline{A}$

Введём здесь ещё несколько понятий. Если $A \cdot B = \emptyset$, то события A и B называются несовместными. В этом случае для обозначения суммы всегда будем писать $A + B$.

1.7 Аксиомы теории вероятностей

Пусть Ω — произвольное непустое множество, $\mathscr{F}$ — некоторая система подмножеств множества Ω.

Определение 1.7. *Система $\mathscr{F}$ подмножеств множества Ω называется алгеброй, если*
A1. $\Omega \in \mathscr{F}$*;*
A2. Если $A \in \mathscr{F}$*, то* $\overline{A} \in \mathscr{F}$
A3. Если $A \in \mathscr{F}, B \in \mathscr{F}$*, то* $A \cdot B \in \mathscr{F}$*;*

Из этих аксиом немедленно следуют следующие свойства алгебры:
1. $\emptyset \in \mathscr{F}$;
2. Если $A \in \mathscr{F}, B \in \mathscr{F}$, то $A \cup B \in \mathscr{F}$, $A \setminus B \in \mathscr{F}$.

Определение 1.8. *Система $\mathscr{F}$ подмножеств множества Ω называется σ-алгеброй, если*
SA1. $\Omega \in \mathscr{F}$*;*
SA2. Если $A \in \mathscr{F}$*, то* $\overline{A} \in \mathscr{F}$
SA3. Если $A_1, A_2, \dots \in \mathscr{F}$*,, то* $\bigcap_{i=1}^{\infty} A_i \in \mathscr{F}$*,* $\bigcup_{i=1}^{\infty} A_i \in \mathscr{F}$*;*

Замечание 1.5. *Всюду далее алгебру (σ-алгебру) $\mathscr{F}$ будем называть алгеброй (σ-алгеброй) событий.*

Определение 1.9. *Пара $(\Omega, \mathscr{F})$ называется измеримым пространством.*

Рассмотрим измеримое пространство $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$, где $\mathbb{R}$ — числовая прямая, $\mathscr{B}(\mathbb{R})$ — борелевская σ-алгебра числовой прямой. Остановимся на свойствах борелевской σ-алгебры, как одном из наиболее важных моментов для нашего курса.

Напомним, что борелевская σ-алгебра числовой прямой порождается следующей системой подмножеств $\mathscr{B}_0 = \{\emptyset, (-\infty, x) : x \in \mathbb{R}\}$. Ясно, что это даже не алгебра, но с её помощью возможно построение σ-алгебры:

$$\mathscr{B}(\mathbb{R}) := \sigma(\mathscr{B}_0) = \bigcap_{\alpha} \mathscr{F}^{\alpha},$$

где $\mathscr{F}^{\alpha}$ — σ-алгебра, содержащая систему $\mathscr{B}_0$, а пересечение берётся по всем таким σ-алгебрам. При этом σ-алгебра $\sigma(\mathscr{B}_0)$ оказывается наименьшей такой σ-алгеброй и называется, σ-алгеброй, порождённой системой $\mathscr{B}_0$, — борелевской σ-алгеброй. Отметим здесь, что эту же σ-алгебру мы получим, если в качестве системы $\mathscr{B}_0$ возьмём систему всех интервалов или отрезков числовой прямой.

Если говорить об определении вероятности на системе $\mathscr{F}$, то проще сделать это на более простом объекте, каковым является алгебра событий. Так мы и поступим. Но часто в возникающих задачах (особенно, в задачах, связанных со случайными процессами) требуется, чтобы вероятность была задана на σ-алгебре. Этот вопрос мы обсудим чуть позже.

Определение 1.10 (А.Н. Колмогоров)**.** *Числовая функция, определённая на алгебре событий $\mathscr{F}$, называется вероятностью или вероятностной мерой, если выполнены следующие аксиомы:*
P1. (аксиома неотрицательности) $\mathbf{P}(A) \geqslant 0$ *для* $\forall A \in \mathscr{F}$*;*
P2. (аксиома нормировки) $\mathbf{P}(\Omega) = 1$*;*
P3. (аксиома конечной аддитивности) Если $AB = \emptyset$*, то*

$$\mathbf{P}(A + B) = \mathbf{P}(A) + \mathbf{P}(\boldsymbol{B});$$

P4. (аксиома счётной аддитивности) Если последовательность событий $\{A_n\}_{n=1}^{\infty}$ *такова, что* $A_i A_j = \emptyset\,(i \neq j)$ *и* $\sum_{i=1}^{\infty} A_i \in \mathscr{F}$*, то*

$$\mathbf{P}\left(\sum_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbf{P}(A_i)$$

При этом тройку $(\Omega, \mathscr{F}, \mathbf{P})$ будем называть абстрактным вероятностным пространством.

Сформулируем ещё одну аксиому — аксиому непрерывности. Она расширит наши представления о вероятности.

P5. (аксиома непрерывности) Для любой невозрастающей последовательности событий $A_1 \supseteq A_2 \supseteq \cdots \supseteq A_n \supseteq \ldots$*, такой, что* $\bigcap_{n=1}^{\infty} A_n = A$ *(при этом иногда будем писать* $A_n \searrow A$*,* $n \to \infty$*) имеет место равенство*

$$\lim_{n\to\infty} \mathbf{P}(A_n) = \mathbf{P}(A).$$

Теорема 1.1. *Функция множеств* $\mathbf{P} = \mathbf{P}(A)$*,* $A \in \mathscr{F}$*, счётно-аддитивна тогда и только тогда, если она конечно-аддитивна и непрерывна.*

Доказательство. Другими словами, надо доказать, что аксиомы $P4$ и $(P3+P5)$ эквивалентны. Пусть выполнена аксиома $P4$ счётной аддитивности. А, значит, выполнена и аксиома $P3$. Возьмём последовательность событий (A_n), удовлетворяющую условиям аксиомы $P5$: $A_1 \supseteq A_2 \supseteq \cdots \supseteq A_n \supseteq \ldots$ и $\bigcap_{n=1}^{\infty} A_n = A$. Образуем новую последовательность событий, полагая

$$C_k = A_k\overline{A}_{k+1},\ (k = 1, 2, \ldots).$$

Заметим, что все C_k и A несовместны. Тогда для любого n

$$A_n = A + \bigcup_{k=n}^{\infty} C_k.$$

В частности, $A_1 = A + \bigcup_{k=1}^{\infty} C_k$. В силу аксиомы $P4$ имеем:

$$\mathbf{P}(A_1) = \mathbf{P}(A) + \sum_{k=1}^{\infty} \mathbf{P}(C_k).$$

Так как в левой части последнего равенства стоит число, то ряд в правой части равенства сходится, и, следовательно, хвост этого ряда стремится к нулю. Тогда

$$\lim_{n\to\infty} \mathbf{P}(A_n) = \mathbf{P}(A) + \lim_{n\to\infty} \left(\sum_{k=n}^{\infty} \mathbf{P}(C_k) \right) = \mathbf{P}(A).$$

Обратно, пусть выполнены аксиомы $P3$ и $P5$. Рассмотрим последовательность событий (A_n), удовлетворяющих аксиоме $P4$: $A_iA_j = \emptyset\,(i \neq j)$ и $\sum_{i=1}^{\infty} A_i \in \mathscr{F}$. Тогда в силу аксиомы $P3$

$$\mathbf{P}\left(\sum_{i=1}^{\infty} A_i \right) = \mathbf{P}\left(\sum_{i=1}^{n} A_i \right) + \mathbf{P}\left(\sum_{i=n+1}^{\infty} A_i \right).$$

С другой стороны, используя последнее равенство, имеем

$$\begin{aligned} \sum_{i=1}^{\infty} \mathbf{P}(A_i) = \lim_{n\to\infty} \sum_{i=1}^{n} \mathbf{P}(A_i) = \lim_{n\to\infty} \mathbf{P}\left(\sum_{i=1}^{n} A_i \right) = \\ \lim_{n\to\infty} \left\{ \mathbf{P}\left(\sum_{i=1}^{\infty} A_i \right) - \mathbf{P}\left(\sum_{i=n+1}^{\infty} A_i \right) \right\}. \end{aligned} \tag{1.1}$$

Так как последовательность событий $B_n = \sum_{i=n+1}^{\infty} A_i$ не возрастает и $\bigcap_{n=1}^{\infty} B_n = \emptyset$, то из аксиомы $P5$ следует, что

$$\lim_{n\to\infty} \mathbf{P}\left(\sum_{i=n+1}^{\infty} A_i \right) = \lim_{n\to\infty} \mathbf{P}(B_n) = \mathbf{P}(\emptyset) = 0.$$

(То, что $\mathbf{P}(\emptyset) = 0$, мы обсудим чуть позже). Значит, из формулы (1.1) следует, что

$$\mathbf{P}\left(\sum_{i=1}^{\infty} A_i \right) = \sum_{i=1}^{\infty} \mathbf{P}(A_i).$$

$\square$

Свойства вероятности

1. $\mathbf{P}(\emptyset) = 0$.
2. $\mathbf{P}(\overline{A}) = 1 - \mathbf{P}(A)$.
3. Если $A \subseteq B$, то $\mathbf{P}(A) \leqslant \mathbf{P}(B)$.
4. $\mathbf{P}(A) \leqslant 1$ для всех $A \in \mathscr{F}$.
5. $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(AB)$.
6. $\mathbf{P}(\bigcup_{k=1}^{n} A_k) = \sum_k \mathbf{P}(A_k) - \sum_{k<l} \mathbf{P}(A_k A_l) + \sum_{k<l<m} \mathbf{P}(A_k A_l A_m) - \ldots + (-1)^{n-1}\mathbf{P}(A_1 A_2 \ldots A_n)$.
7. Если последовательность событий (A_n) не убывает, то есть $A_1 \subseteq A_2 \subseteq \cdots \subseteq A_n \subseteq \ldots$ и $\bigcup_{n=1}^{\infty} A_n = A$, то

$$\lim_{n\to\infty} \mathbf{P}(A_n) = \mathbf{P}(A).$$

Докажем некоторые из этих свойств.

1. Из равенства $\Omega = \Omega + \emptyset$ и аксиомы конечной аддитивности получаем: $\mathbf{P}(\Omega) = \mathbf{P}(\Omega) + \mathbf{P}(\emptyset)$. Отсюда $\mathbf{P}(\emptyset) = 0$.

2. Точно также из равенства $\Omega = A + \overline{A}$ и аксиомы конечной аддитивности получим: $1 = \mathbf{P}(A) + \mathbf{P}(\overline{A})$. Отсюда $\mathbf{P}(\overline{A}) = 1 - \mathbf{P}(A)$.

3. Пусть $A \subseteq B$. Тогда $B = A + B\overline{A}$. Значит, $\mathbf{P}(B) = \mathbf{P}(A) + \mathbf{P}(B\overline{A})$ и, так как $\mathbf{P}(B\overline{A}) \geqslant 0$, то $\mathbf{P}(B) \geqslant \mathbf{P}(A)$.

4. Очевидно из 3) и аксиомы нормировки.

5. Запишем $A \cup B = A + (B \setminus AB)$. Но $\mathbf{P}(B \setminus AB) = \mathbf{P}(B) - \mathbf{P}(AB)$. Следовательно, $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(AB)$.

Как мы уже отмечали, класс алгебр значительно шире класса σ-алгебр, во многих же задачах теории вероятностей приходится накладывать дополнительные условия — требовать, чтобы класс событий был σ-алгеброй. Если $\mathscr{F}$ является алгеброй, на которой задана вероятностная мера, то нельзя ли перенести эту меру на σ-алгебру, порождённую этой алгеброй? Положительный ответ на этот вопрос даёт следующая теорема из анализа, приводимая здесь без доказательства.

Теорема 1.2 (Теорема Каратеодори)**.** *Пусть* $\mathbf{P}$ *— вероятность на алгебре событий* $\mathscr{F}$, $\sigma(\mathscr{F})$ *—* σ*-алгебра, порождённая этой алгеброй. Тогда существует единственная вероятностная мера* $\mathbf{P}^*$*, определённая на* $\sigma(\mathscr{F})$*, причём*

$$\mathbf{P}^*(A) = \mathbf{P}(A)$$

для всех $A \in \mathscr{F}$*. Вероятность* $\mathbf{P}^*$ *называется при этом продолжением вероятности* $\mathbf{P}$ *с алгебры на* σ*-алгебру.*

1.8 Условная вероятность. Независимость

Начнём с простого примера.

Пример 1.11. *Пусть опыт заключается в том, что три раза бросается монета. Вероятность того, что при этом «герб» выпадет ровно один раз в классической схеме равна* $3/8$. *Обозначим это событие через* A. *Предположим далее, что об исходе эксперимента дополнительно известно, что произошло событие* $B = \{$*число выпавших «гербов» нечётно*$\}$. *Какова вероятность события* A *с учётом этой информации? Построим новое пространство элементарных исходов с учётом того, что событие* B *произошло:* $\Omega|B = \{\Gamma PP, P\Gamma P, PP\Gamma, \Gamma\Gamma\Gamma\}$. *Тогда согласно классической схеме вероятность события* A *при условии, что произошло событие* B, *равна* $\mathbf{P}(A|B) = 3/4$. *Назовём эту вероятность условной вероятностью.*

Из этого примера мы можем сделать некоторые выводы. Во-первых, в нетривиальном случае события A и B должны быть совместными. Во-вторых, интуитивно понятно, что на значение условной вероятности влияет вероятность события B (при вычислении по классической схеме мы использовали $\Omega|B$). Введём строгое определение.

Пусть $(\Omega, \mathscr{F}, \mathbf{P})$ — произвольное вероятностное пространство.

Определение 1.11. *Рассмотрим* $A, B \in \mathscr{F}$, $\mathbf{P}(B) > 0$. *Условная вероятность события* A *при условии, что произошло событие* B, *определяется по формуле*

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

Теперь мы дадим интуитивно понятное определение независимости событий — одного из важнейших в теории вероятностей.

Определение 1.12. *Пусть* $A, B \in \mathscr{F}$, $\mathbf{P}(B) > 0$. *События* A *и* B *называются независимыми, если*

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

Замечание 1.6. *Кажется, что здесь есть какой-то подвох — в определении 1.12 говорится только о независимости события* A *от* B. *Но на самом деле, если событие* A *не зависит от* B, *то и событие* B *не зависит от* A, *если* $\mathbf{P}(A) > 0$.

Действительно, пусть $\mathbf{P}(A|B) = \mathbf{P}(A)$. По определению условной вероятности $\frac{\mathbf{P}(AB)}{\mathbf{P}(B)} = \mathbf{P}(A)$. Отсюда $\mathbf{P}(AB) = \mathbf{P}(A)\mathbf{P}(B)$. Тогда $\mathbf{P}(B|A) = \frac{\mathbf{P}(AB)}{\mathbf{P}(A)} = \frac{\mathbf{P}(A)\mathbf{P}(B)}{\mathbf{P}(A)} = \mathbf{P}(B)$.

Смысл этого определения лежит на поверхности: наступление одного события не влияет на вероятность наступления другого, но оно не симметрично относительно событий. Предыдущие рассуждения привели нас к следующему, удобному для использования, определению независимости событий.

Определение 1.13. *Пусть $A, B \in \mathscr{F}$, $\mathbf{P}(A) > 0$, $\mathbf{P}(B) > 0$. События A и B называются независимыми, если*

$$\mathbf{P}(AB) = \mathbf{P}(A)\mathbf{P}(B).$$

Мы уже показали, что из определения 1.12 следует определение 1.13. Обратное более, чем очевидно.

Замечание 1.7. *Определение 1.13 не только симметрично, но и имеет смысл также и для невозможного события $\emptyset$, то есть в этом определении допускается, что события A и B могут иметь вероятность, равную нулю. При этом, невозможное и достоверное события не зависят от любых других событий, в том числе, от себя самих.*

1.8.1 Свойства независимых событий

Предложение 1.1. *1. Если события A и B независимы, то независимы также события $\overline{A}$ и B, A и $\overline{B}$, $\overline{A}$ и $\overline{B}$;*

2. Пусть события A и B_1 независимы, события A и B_2 независимы, при этом события B_1 и B_2 несовместны. Тогда независимы события A и $(B_1 + B_2)$.

3. Пусть события A и B несовместны и $\mathbf{P}(A) > 0, \mathbf{P}(B) > 0$. Тогда события A и B зависимы.

Доказательство. 1. Проверим независимость, к примеру, событий $\overline{A}$ и B.
$\mathbf{P}(\overline{A}B) = \mathbf{P}(B \setminus AB) = \mathbf{P}(B) - \mathbf{P}(AB) = \mathbf{P}(B) - \mathbf{P}(A)\mathbf{P}(B) = \mathbf{P}(B)(1 - \mathbf{P}(A)) = \mathbf{P}(B) \cdot \mathbf{P}(\overline{A})$.
2. $\mathbf{P}(A(B_1+B_2)) = \mathbf{P}(AB_1+AB_2)) = \mathbf{P}(AB_1)+\mathbf{P}(AB_2)) = \mathbf{P}(A)\mathbf{P}(B_1)+\mathbf{P}(A)\mathbf{P}(B_2)) = \mathbf{P}(A)(\mathbf{P}(B_1)+\mathbf{P}(B_2)) = \mathbf{P}(A)\mathbf{P}((B_1+B_2))$
3. $0 = \mathbf{P}(AB) \neq \mathbf{P}(A)\mathbf{P}(B) > 0$. □

Наряду с понятием независимости событий существует ещё и понятие условной независимости.

Определение 1.14. *События A и B называются условно независимыми относительно события C, если*

$$\mathbf{P}(AB|C) = \mathbf{P}(A|C)\mathbf{P}(B|C).$$

В теории марковских цепей именно это свойство выполняется для «прошлого» и «будущего» при фиксированном «настоящем». Об этом мы поговорим в курсе случайных процессов.

1.8.2 Вероятность произведения событий

Рассмотрим $A, B \in \mathscr{F}$, $\mathbf{P}(B) > 0$. Записывая формулу условной вероятности в несколько другой форме, получим формулу

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

По индукции можно доказать следующую, более общую формулу, называемую ещё формулой умножения вероятностей:

$$\mathbf{P}(A_1A_2\cdots A_n) = \mathbf{P}(A_1)\mathbf{P}(A_2|A_1)\mathbf{P}(A_3|A_1A_2)\cdots\mathbf{P}(A_n|A_1\ldots A_{n-1}).$$

1.8.3 Формула полной вероятности. Формула Байеса

Определение 1.15. *Пусть $(\Omega, \mathscr{F}, \mathbf{P})$ — произвольное вероятностное пространство, тогда любая конечная или бесконечная система из N $(2 \leqslant N \leqslant \infty)$ случайных событий $H_1, H_2, \ldots, H_N \in \mathscr{F}$ называется полной системой (или группой) случайных событий (полной группой гипотез), если выполнены следующие три свойства:*

1. $\mathbf{P}(H_i) > 0, i = 1, 2, \ldots, N$;
2. $H_iH_j = \emptyset\, (i \neq j)$;
3. $\sum_{i=1}^{N} H_i = \Omega$.

Теорема 1.3 (Формула полной вероятности)**.** *Пусть на произвольном вероятностном пространстве $(\Omega, \mathscr{F}, \mathbf{P})$ определена полная группа событий $(H_i)_{i=1}^{N}$. Тогда для любого случайного события $A \in \mathscr{F}$ его безусловная вероятность равна*

$$\mathbf{P}(A) = \sum_{i=1}^{N} \mathbf{P}(H_i)\mathbf{P}(A|H_i).$$

Доказательство. Поскольку $\Omega = \sum_{i=1}^{N} H_i$, то $A = \sum_{i=1}^{N}(AH_i)$. Поэтому

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

□

Теорема 1.4 (Формула Байеса)**.** *Пусть выполнены условия теоремы 1.3. Тогда для любого события* $A \in \mathscr{F}, \mathbf{P}(A) > 0$ *справедлива следующая формула:*

$$\mathbf{P}(H_j|A) = \frac{\mathbf{P}(H_j)\mathbf{P}(A|H_j)}{\sum_{i=1}^{N} \mathbf{P}(H_i)\mathbf{P}(A|H_i)},\ (j = 1, 2, \ldots .N).$$

Доказательство. Очевидно следует из формулы условной вероятности и формулы полной вероятности. □

Замечание 1.8. *Вероятности в формулах полной вероятности и Байеса называются соответственно априорной и апостеориорной вероятностями.*

Пример 1.12. *Пусть в шкатулке находятся две монеты, из которых одна симметричная (вероятность «герба» равна* $1/2$*), а другая — несимметричная (вероятность «герба» равна* $1/3$*). Наудачу выбирается и подбрасывается одна монета. Предположим, что выпал «герб». Какова вероятность того, что выбранная монета симметрична?*

Решение.

Обозначим через

$$H_1 = \{\text{выбранная монета симметрична}\},$$

$$H_2 = \{\text{выбранная монета несимметрична}\},$$

$$A = \{\text{при подбрасывании выбранной монеты выпал «герб»}\}.$$

События H_1 и H_2 образуют полную группу событий, причём, $\mathbf{P}(H_1) = \mathbf{P}(H_2) = 1/2$. Вычислим вероятность события A по формуле полной вероятности. Из условий задачи имеем: $\mathbf{P}(A|H_1) = 1/2,\ \mathbf{P}(A|H_2) = 1/3$. Тогда

$$\mathbf{P}(A) = \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{3} = \frac{5}{12}.$$

Теперь, используя формулу Байеса, имеем:

$$\mathbf{P}(H_1|A) = \frac{\frac{1}{2}\cdot\frac{1}{2}}{\frac{5}{12}} = \frac{3}{5}.$$

□

Пример 1.13 (Задача контроля качества)**.** *Известно, что* 90% *выпускаемых заводом изделий отвечает стандарту. Существующая система контроля (выборочная) признаёт стандартную продукцию годной с вероятностью* $0,95$*, а нестандартную — с вероятностью* $0,03$*. Найти вероятность того, что изделие признанное годным, отвечает стандарту.*

Решение.

Обозначим через

$$H_1 = \{\text{изделие отвечает стандарту}\},$$

$$H_2 = \{\text{изделие не отвечает стандарту}\},$$

$$A = \{\text{изделие признано годным}\}.$$

По условиям задачи имеем: $\mathbf{P}(H_1) = 0,9$, $\mathbf{P}(H_2) = 0,1$, $\mathbf{P}(A|H_1) = 0,95$, $\mathbf{P}(A|H_2) = 0,03$. Значит, по формуле Байеса

$$\mathbf{P}(H_1|A) = \frac{0,95\cdot 0,9}{0,95\cdot 0,9 + 0,1\cdot 0,03} \approx 0,997.$$

□

1.8.4 Независимость в совокупности и попарная

Определение 1.16. *События* $A_1, A_2, \ldots, A_n$ *называются независимыми в совокупности, если для любого набора индексов* $(i_1, i_2, \ldots, i_k)$, $k = 2, 3, \ldots, n$*:*

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

В том случае, когда $k = 2$*, такая независимость называется попарной.*

Замечание 1.9. *Заметим, что из попарной независимости, вообще говоря, ещё не следует независимости в совокупности. Приведём известный пример, подтверждающий это утверждение.*

Пример 1.14 (Пирамидка Бернштейна)**.** *Бросается правильный тетраэдр, три грани которого окрашены в красный, синий и зелёный цвета, а основание во все три цвета. События K, S, Z имеют место, если на грани тетраэдра, на которую он упал, имеется соответствующий цвет. Проверить независимость (в совокупности и попарную) событий K, S, Z.*

Решение.

Непосредственно проверяется, что события K, S, Z попарно независимы, и в то же время

$$\mathbf{P}(KSZ) = \frac{1}{4} \neq \frac{1}{8} = \mathbf{P}(K)\mathbf{P}(S)\mathbf{P}(Z),$$

то есть, не являются независимыми в совокупности. □

1.8.5 Независимость σ-алгебр. Схема Бернулли

Пусть $(\Omega, \mathscr{F}, \mathbf{P})$ — произвольное вероятностное пространство. Пусть далее $\mathscr{F}_1, \mathscr{F}_2, \ldots, \mathscr{F}_n$ — σ-подалгебры σ-алгебры $\mathscr{F}$.

Определение 1.17. *σ-алгебры $\mathscr{F}_1, \mathscr{F}_2, \ldots, \mathscr{F}_n$ называются независимыми, если для любых событий $A_1 \in \mathscr{F}_1, A_2 \in \mathscr{F}_2, \ldots, A_n \in \mathscr{F}_n$*

$$\mathbf{P}(A_1 A_2 \cdots A_n) = \mathbf{P}(A_1)\mathbf{P}(A_2) \cdots \mathbf{P}(A_n).$$

Замечание 1.10. *Любопытно, что добавка «в совокупности» оказывается в этом случае излишней, так как из этого определения следует независимость в совокупности событий $A_i (i = 1, 2, \ldots, n)$. Мы требуем выполнения лишь одного равенства, но для всех наборов, состоящих из одного события в каждой из σ-алгебр.*

В свете последнего определения рассмотрим следующую классическую модель.

Предположим, что некоторый эксперимент (опыт) проводится n раз, причём мы фиксируем при этом наступление некоторого события A. Если в результате опыта происходит событие A, то будем говорить, что произошёл «успех», если не произошло, — «неудача». Обозначим через $p = \mathbf{P}(A)$ (вероятность «успеха»), через $q = 1 - p$ (вероятность «неудачи»). Результат n-кратного повторения опыта запишем в виде упорядоченного набора $(\omega_1, \omega_2, \ldots, \omega_n)$, где $\omega_i = 1$, если произошёл «успех», $\omega_i = 0$, если произошла «неудача». Пространство элементарных исходов тогда имеет следующую структуру:

$$\Omega = \{\omega : \omega = (\omega_1, \omega_2, \ldots, \omega_n), \omega_i = 0; 1, i = 1, 2, \ldots, n\}.$$

Припишем каждому элементарному событию $\omega = (\omega_1, \omega_2, \ldots, \omega_n)$ вероятность
$$p(\omega) = p^{\sum_{i=1}^n \omega_i} \cdot q^{n-\sum_{i=1}^n \omega_i}.$$

Покажем, что такой способ задания «весов» $p(\omega)$ является корректным. Для этого достаточно проверить, что $\sum\limits_{\omega\in\Omega} p(\omega) = 1$. Рассмотрим все те элементарные события $\omega = (\omega_1, \omega_2, \ldots, \omega_n)$, для которых $\sum\limits_{i=1}^n \omega_i = k$, $k = 0, 1, 2, \ldots, n$, то есть, в серии n испытаний произошло ровно k «успехов». Понятно, что k «успехов» по n местам можно расположить C_n^k способами. Поэтому

$$\sum_{\omega\in\Omega} p(\omega) = \sum_{k=0}^n C_n^k p^k q^{n-k} = (p+q)^n = 1.$$

Далее в Ω рассмотрим алгебру всех подмножеств $\mathscr{F}$, и для каждого $A \in \mathscr{F}$ положим
$$\mathbf{P}(A) = \sum_{\omega\in A} p(\omega).$$

Таким образом, мы получили некоторое дискретное вероятностное пространство $(\Omega, \mathscr{F}, \mathbf{P})$.

Рассмотрим в алгебре $\mathscr{F}$ следующие события:

$$A_k = \{\omega = (\omega_1, \omega_2, \ldots, \omega_n) : \omega_k = 1\},\ \overline{A_k} = \{\omega = (\omega_1, \omega_2, \ldots, \omega_n) : \omega_k = 0\}.$$

Очевидно, что эти события зависят только от k-го испытания. Рассмотрим далее подалгебры $\mathscr{F}_1, \mathscr{F}_2, \ldots \mathscr{F}_n$, где

$$\mathscr{F}_k = \{\emptyset, A_k, \overline{A_k}, \Omega\}.$$

Покажем, что тогда эти алгебры независимы.

Ясно, что

$$\begin{aligned}\mathbf{P}(A_k) = \sum_{\{\omega:\omega_k=1\}} p(\omega) = \sum_{\{\omega:\omega_k=1\}} p^{\sum_{i=1}^n \omega_i} \cdot q^{n-\sum_{i=1}^n \omega_i} = \\ p \cdot \sum_{(\omega_1,\ldots,\omega_{k-1},\omega_{k+1},\ldots,\omega_n)} p^{\sum_{i\neq k} \omega_i} \cdot q^{n-\sum_{i\neq k} \omega_i} = \\ p \cdot \sum_{m=0}^{n-1} C_{n-1}^m p^m q^{n-m-1} = p(p+q)^{n-1} = p.\end{aligned} \tag{1.2}$$

Аналогично (1.2) показывается, что $\mathbf{P}(\overline{A_k}) = q$, $\mathbf{P}(A_k A_l) = p^2$, $(k \neq l)$, и так далее. Таким образом, алгебры $\mathscr{F}_1, \mathscr{F}_2, \dots \mathscr{F}_n$ независимы.

Всё это даёт нам основание называть полученную модель $(\Omega, \mathscr{F}, \mathbf{P})$ моделью, отвечающей «n независимым испытаниям с двумя исходами и вероятностью «успеха» p». Эту модель систематически изучал Я. Бернулли, в связи с чем её называют схемой Бернулли.

Естественным обобщением этой модели служит так называемая «полиномиальная схема» — модель, отвечающая n независимым испытаниям с m исходами и вероятностями этих исходов $p_1, p_2, \dots, p_m$, такими, что $\sum_{i=1}^m p_i = 1$. Подробнее об этой схеме мы поговорим позже.

1.9 Предельные теоремы в схеме Бернулли

Рассмотрим введённую ранее схему Бернулли $(\Omega, \mathscr{F}, \mathbf{P})$ и обозначим через μ_n — число «успехов» в n независимых испытаниях с вероятностью «успеха» p. Ранее мы показали, что

$$\mathbf{P}\{\mu_n = k\} = C_n^k p^k q^{n-k}.$$

Во многих задачах приходится сталкиваться с тем, что число испытаний n достаточно велико, поэтому необходимы асимптотические формулы для подсчёта этой вероятности. Рассмотрим здесь несколько теорем.

Теорема 1.5 (Локальная теорема Муавра-Лапласа)**.** *Пусть* $0 < p < 1$, *последовательность* $\{\frac{k-np}{\sqrt{npq}}\}$ *равномерно ограничена по* n *и* k *при* $n \to \infty$. *Тогда*

$$\mathbf{P}\{\mu_n = k\} = \frac{1}{\sqrt{2\pi npq}} e^{-\frac{(k-np)^2}{2npq}} \cdot \left(1 + O\left(\frac{1}{\sqrt{n}}\right)\right).$$

Доказательство. Доказательство теоремы существенно опирается на известную формулу Стирлинга: при $n \to \infty$

$$n! \approx \sqrt{2\pi n} \left(\frac{n}{e}\right)^n e^{\theta(n)}, \quad \text{где } |\theta(n)| \leqslant \frac{1}{2n}.$$

Обозначим $x_k = \frac{k-n}{\sqrt{npq}}$. Тогда $k = np + x_k\sqrt{npq}$, $n - k = nq - x_k\sqrt{npq}$.

Рассмотрим

$$\ln \mathbf{P}\{\mu_n = k\} = \ln\left(C_n^k p^k q^{n-k}\right) = \ln C_n^k + k\ln p + (n-k)\ln q. \tag{1.3}$$

Имеем: $C_n^k = \frac{n!}{k!(n-k)!}$. Пользуясь формулой Стирлинга, запишем

$$\ln n! \approx \ln\sqrt{2\pi n} + n\ln n - n,$$

$$\ln k! \approx \ln\sqrt{2\pi k} + (np + x_k\sqrt{npq})\ln(np + x_k\sqrt{npq}) - np - x_k\sqrt{npq},$$

$$\ln(n-k)! \approx \ln\sqrt{2\pi(n-k)} + (nq - x_k\sqrt{npq})\ln(nq - x_k\sqrt{npq}) - nq + x_k\sqrt{npq}.$$

Преобразуем последние логарифмы

$$\ln(np + x_k\sqrt{npq}) = \ln\left[np \cdot \left(1 + x_k\sqrt{\frac{q}{np}}\right)\right] = \ln(np) + \ln\left(1 + x_k\sqrt{\frac{q}{np}}\right).$$

Точно также

$$\ln(nq - x_k\sqrt{npq}) = \ln(nq) + \ln\left(1 - x_k\sqrt{\frac{p}{nq}}\right).$$

При $n \to \infty$ $\quad x_k\sqrt{\frac{q}{np}} \to 0$, $x_k\sqrt{\frac{p}{nq}} \to 0$. Поэтому, применяя формулу Тейлора к функции $\ln(1+x) = x - x^2/2 + O(x^3)$ в окрестности нуля к последним логарифмам, и подставляя в формулу 1.3, получим

$$\ln \mathbf{P}\{\mu_n = k\} \approx -\frac{x_k^2}{2} - \ln\sqrt{2\pi} - \ln\sqrt{npq}.$$

Отсюда

$$\mathbf{P}\{\mu_n = k\} \approx \frac{1}{\sqrt{2\pi npq}} e^{-\frac{(k-np)^2}{2npq}}. \tag{1.4}$$

□

Замечание 1.11. *При $n \to \infty$ сумма остаточных членов стремится к нулю при любых фиксированных значениях p и q, $0 < p < 1$. Однако, при конечных (но больших) значениях n сумма остаточных членов может быть достаточно большой, если p или q малы.*

Хорошее приближение формула 1.4 даёт при $p = q = 1/2$. Здесь можно показать, оценивая сумму остаточных членов, что порядок остаточного члена будет $O(1/n)$.

Теорема 1.6 (Интегральная теорема Муавра-Лапласа)**.** *Пусть $0 < p < 1$. Тогда в условиях предыдущей теоремы*

$$\mathbf{P}\{k_1 \leqslant \mu_n \leqslant k_2\} \to \frac{1}{\sqrt{2\pi}} \int_{x_1}^{x_2} e^{-\frac{x^2}{2}} dx, \quad (n \to \infty)$$

равномерно по x_1 и x_2, где $x_i = (k_i - np)/\sqrt{npq}$, $(i = 1, 2)$.

Доказательство. Ясно, что

$$\mathbf{P}\{k_1 \leqslant \mu_n \leqslant k_2\} = \sum_{k=k_1}^{k_2} \mathbf{P}\{\mu_n = k\}.$$

Используя локальную теорему Муавра-Лапласа, имеем

$$S_n = \sum_{k=k_1}^{k_2} \mathbf{P}\{\mu_n = k\} \approx \sum_{x_i \in [x_1, x_2]} \frac{1}{\sqrt{2\pi}} e^{-\frac{x_i^2}{2}} \cdot \frac{1}{\sqrt{npq}} = \sum_{x_i \in [x_1, x_2]} \frac{1}{\sqrt{2\pi}} e^{-\frac{x_i^2}{2}} \cdot \Delta x_i,$$

где $\Delta x_i = 1/\sqrt{npq}$. Поэтому S_n представляется как интегральная сумма для функции $\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ на отрезке $[x_1, x_2]$. Значит,

$$\lim_{n\to\infty} S_n = \frac{1}{\sqrt{2\pi}} \int_{x_1}^{x_2} e^{-\frac{x^2}{2}} dx,$$

то есть,

$$\mathbf{P}\{k_1 \leqslant \mu_n \leqslant k_2\} \approx \frac{1}{\sqrt{2\pi}} \int_{x_1}^{x_2} e^{-\frac{x^2}{2}} dx.$$

Равномерная сходимость S_n к своему пределу может быть доказана сходимостью к нулю остаточного члена

$$R_n = \frac{1}{\sqrt{2\pi}} \int_{x_1}^{x_2} e^{-\frac{x^2}{2}} \alpha_n(x) dx + \rho_n,$$

где

$$\lim_{n\to\infty} \max_{x_1 \leqslant x \leqslant x_2} \alpha_n(x) = 0,$$

ρ_n — бесконечно малая (подробнее можно посмотреть в [3]).

$\square$

Введём следующее обозначение:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{x^2}{2}} dx.$$

Тогда интегральную теорему Муавра-Лапласа можно записать в виде:

$$\mathbf{P}\{k_1 \leqslant \mu_n \leqslant k_2\} \approx \Phi\left(\frac{k_2 - np}{\sqrt{npq}}\right) - \Phi\left(\frac{k_1 - np}{\sqrt{npq}}\right).$$

Значения функции $\Phi(x)$ можно взять из таблиц так называемой нормальной функции распределения. Отметим здесь, что интегральная теорема Муавра-Лапласа имеет более общий характер, но об этом мы поговорим позднее.

О применимости приближённых формул Муавра-Лапласа мы уже говорили выше. Эти две теоремы, скорее, имеют качественный характер, и ими надо пользоваться с осторожностью. Значительный круг задач связан с вычислением вероятности $\mathbf{P}\{\mu_n = k\}$ при значениях вероятности p близкими к нулю или единице (тогда близко к нулю будет q). Если при этом число испытаний достаточно велико, то приближённую формулу нашёл Пуассон.

Рассмотрим последовательность серий испытаний, в которой «успехи» одной серии взаимно не зависят между собой имеют каждое вероятность p_n. Через μ_n обозначается число фактически появившихся «успехов» n-й сериии.

Теорема 1.7 (Теорема Пуассона)**.** *Если* $p_n \to 0$ *при* $n \to \infty$*, причём* $np \to \lambda,\ \lambda > 0$*, то для каждого* $k = 0, 1, 2, \dots$

$$\mathbf{P}\{\mu_n = k\} \to \frac{\lambda^k}{k!}e^{-\lambda} \quad (n \to \infty).$$

Доказательство. По предположениям теоремы

$$p = p_n = \frac{\lambda}{n} + o\left(\frac{1}{n}\right).$$

Тогда для каждого фиксированного k и достаточно больших n

$$\mathbf{P}\{\mu_n = k\} = C_n^k p_r q_{n-k} =$$
$$= \frac{n(n-1)\dots(n-k+1)}{k!}\left(\frac{\lambda}{n} + o\left(\frac{1}{n}\right)\right)^k \left(1 - \frac{\lambda}{n} + o\left(\frac{1}{n}\right)\right)^{n-k} =$$
$$= \frac{n(n-1)\dots(n-k+1)}{k!}\frac{1}{n^k}(\lambda + o(1))^k \left(1 - \frac{\lambda}{n} + o\left(\frac{1}{n}\right)\right)^{n-k} =$$
$$= \frac{(\lambda + o(1))^k}{k!}(1 - \frac{1}{n})\dots(1 - \frac{k+1}{n})\left(1 - \frac{\lambda}{n} + o\left(\frac{1}{n}\right)\right)^n \times$$
$$\times \left(1 - \frac{\lambda}{n} + o\left(\frac{1}{n}\right)\right)^{-k} \to \frac{\lambda^k}{k!}e^{-\lambda}, \quad (n \to \infty).$$

□

Пример 1.15. *Вероятность попадания в цель при каждом выстреле равна* 0,001*. Найти вероятность попадания в цель двумя и более выстрелами при залпе* 5000 *выстрелов.*

Решение.

Имеем $p = 0,001$, $n = 5000$, $\lambda = 5$. Так как вероятность p близка к нулю, то воспользуемся приближённой формулой Пуассона.

$$\mathbf{P}\{\mu_n \geqslant 2\} = 1 - \mathbf{P}\{\mu_n = 0\} - \mathbf{P}\{\mu_n = 1\} \approx 1 - \frac{6}{e^5}.$$

□

Пример 1.16. *Найти вероятность того, что число выпадений «6» при* 12000 *бросаниях правильной игральной кости заключено между* 1900 *и* 2150.

Решение.

По условию $p = 1/6, n = 12000, np = 2000, npq \approx 1667$. Воспользуемся интегральной формулой Муавра-Лапласа.

$$\mathbf{P}\{1900 \leqslant \mu_n \leqslant 2150\} \approx \Phi(15/4) - \Phi(-5/2) = \Phi(15/4) - (1 - \Phi(5/2)) = 0,937.$$

$\square$

Глава 2

Случайные величины и их распределения

2.1 Случайные величины и функции распределения

2.1.1 Случайная величина

Пусть $(\Omega, \mathscr{F}, \mathbf{P})$ — некоторое вероятностное пространство, $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ — борелевская прямая (числовая прямая с борелевской σ-алгеброй).

Определение 2.1. *Числовая функция $\xi : \Omega \to \mathbb{R}$ называется случайной величиной, если для любого $B \in \mathscr{B}(\mathbb{R})$*

$$\xi^{-1}(B) = \{\omega : \xi(\omega) \in B\} \in \mathscr{F}.$$

Другими словами, случайной величиной называется измеримая функция $\xi : \Omega \to \mathbb{R}$.

Замечание 2.1. *Для компактности формул будем часто писать $\{\xi \in B\}$ вместо $\{\omega : \xi(\omega) \in B\}$, $\{\xi < x\}$ вместо $\{\omega : \xi(\omega) < x\}$, и т.д.*

Пример 2.1. *Монета бросается один раз. Пространство элементарных исходов содержит два элемента: $\Omega = \{\omega_\Gamma, \omega_P\}$. Зададим на Ω отображение ξ следующим образом: $\xi(\omega_\Gamma) = 1$, $\xi(\omega_P) = 0$. Покажем, что заданное таким образом отображение является случайной величиной.*

Решение.

Понятно, что алгебра событий в нашем случае состоит из четырёх элементов: $\mathscr{F} = \{\emptyset, \{\omega_\Gamma\}, \{\omega_P\}, \Omega\}$. Рассмотрим произвольное множество

$B \in \mathscr{B}(\mathbb{R})$. Если множество B не содержит ни нуля, ни единицы, то $\xi^{-1}(B) = \emptyset \in \mathscr{F}$; если содержит только единицу, то $\xi^{-1}(B) = \{\omega_\Gamma\} \in \mathscr{F}$; если содержит только нуль, то $\xi^{-1}(B) = \{\omega_P\} \in \mathscr{F}$; если содержит и нуль, и единицу, то $\xi^{-1}(B) = \Omega \in \mathscr{F}$.

□

Пример 2.2. *Точка случайно бросается на квадрат $\{(x,y) : 0 \leqslant x \leqslant 1, 0 \leqslant y \leqslant 1\}$. В качестве пространства элементарных событий здесь выступает сам квадрат $\Omega = \{\omega = (x,y) : 0 \leqslant x \leqslant 1, 0 \leqslant y \leqslant 1\}$, $\mathscr{F}$ — сужение борелевской σ-алгебры $\mathbb{R}^2$ на квадрат. Положим для любого $\omega = (x,y) \in \Omega :\ \xi(\omega) = \sqrt{x^2+y^2}$. Тогда заданное таким образом отображение является случайной величиной.*

Для того, чтобы утверждение примера 2.2 стало очевидным, докажем следующую теорему.

Теорема 2.1. *Для того, чтобы отображение $\xi : \Omega \to \mathbb{R}$ было случайной величиной, необходимо и достаточно, чтобы для любого $x \in \mathbb{R}$*

$$\xi^{-1}((-\infty, x)) = \{\xi < x\} \in \mathscr{F}.$$

Доказательство. Необходимость очевидна, поскольку $(-\infty, x) \in \mathscr{B}(\mathbb{R})$.

Достаточность. Обозначим $\mathscr{D}_1 = \{(-\infty, x) : x \in \mathbb{R}\}, \mathscr{D}_2 = \{B \in \mathscr{B}(\mathbb{R}) : \xi^{-1}(B) \in \mathscr{F}\}$. Так как операция «взятия прообраза» сохраняет все алгебраические операции, в том числе и счётные, то система $\mathscr{D}_2$ является σ-алгеброй и

$$\mathscr{D}_1 \subseteq \mathscr{D}_2 \subseteq \mathscr{B}(\mathbb{R}).$$

Тогда

$$\sigma(\mathscr{D}_1) \subseteq \sigma(\mathscr{D}_2) = \mathscr{D}_2 \subseteq \mathscr{B}(\mathbb{R}).$$

Поскольку $\sigma(\mathscr{D}_1) = \mathscr{B}(\mathbb{R})$, то $\mathscr{D}_2 = \mathscr{B}(\mathbb{R})$. □

Возвращаясь к примеру 2.2: из теоремы 2.1 непосредственно следует, что множество $\{(x,y) : \sqrt{x^2+y^2} < t\} \in \mathscr{F}$ для всех $t \in \mathbb{R}$.

2.1.2 Функции от случайных величин

Пусть функция $\varphi : \mathbb{R} \to \mathbb{R}$ является боревской (измеримой по Борелю) функцией, $\xi = \xi(\omega)$ — некоторая случайная величина.

Теорема 2.2. *Борелевская функция от случайной величины есть случайная величина.*

Доказательство. Пусть $B \in \mathscr{B}(\mathbb{R})$. Обозначим через $\eta(\omega) = \varphi(\xi(\omega))$. Рассмотрим

$$\{\eta \in B\} = \{\varphi(\xi) \in B\} = \{\xi \in \varphi^{-1}(B)\} \in \mathscr{F},$$

так как $\varphi^{-1}(B) \in \mathscr{B}(\mathbb{R})$. □

2.1.3 Функция распределения

Определение 2.2. *Функция*

$$F_\xi(x) = \mathbf{P}\{\xi < x\}, \quad x \in \mathbb{R},$$

называется функцией распределения случайной величины ξ.

Приведём примеры непосредственного построения функции распределения.

Пример 2.3. *Возьмём случайную величину из примера 2.1. Ясно, что* $\mathbf{P}\{\xi = 0\} = \mathbf{P}\{\xi = 1\} = 1/2$.

Решение.

Построение функции распределения очевидно следует из разбора примера 2.1:

$$F_\xi(x) = \begin{cases} 0, & x \leqslant 0; \\ \frac{1}{2}, & 0 < x \leqslant 1; \\ 1, & x > 1. \end{cases}$$

□

Пример 2.4. *Пусть на отрезок* $[a, b]$ *числовой прямой случайно бросается точка. Здесь* $\Omega = [a, b]$, $\mathscr{F} = \mathscr{B}([a, b])$ *— борелевская* σ*-алгебра отрезка* $[a, b]$, *в качестве* $\mathbf{P}$ *рассмотрим геометрическую вероятность (равномерное распределение). Определим случайную величину*

$$\xi(\omega) = \omega, \ \omega \in [a, b].$$

Решение.

Очевидно, что функция ξ измерима, то есть, является случайной величиной. Построим функцию распределения этой случайной величины. Ясно из условий задачи, что если $x \leqslant a$, то $F_\xi(x) = 0$. Если $x > b$, то $F_\xi(x) = 1$. Пусть теперь $x \in (a, b]$. Тогда событие $\{\xi < x\}$ означает, что точка попала в интервал (a, x). Из геометрических соображений вероятность этого события равна

$$\mathbf{P}\{\xi < x\} = \frac{x-a}{b-a}.$$

Итак,

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

□

Замечание 2.2. *Многие авторы (например, А.Н. Ширяев, [6, 7]) предпочитают называть функцией распределения случайной величины ξ функцию $\mathbf{P}\{\xi \leqslant x\}$. Отличие этой функции от введённой нами ранее функции лишь в том, что меняется одно из её свойств. Обратите на это внимание.*

2.1.4 Свойства функции распределения

Пусть $F_\xi(x)$ — функция распределения случайной величины ξ.

1. Функция $F_\xi(x)$ не убывает;
2. Функция $F_\xi(x)$ непрерывна слева для всех $x \in \mathbb{R}$;
3. $\lim\limits_{x\to-\infty} F_\xi(x) = 0$, $\lim\limits_{x\to+\infty} F_\xi(x) = 1$;
4. Функция $F_\xi(x)$ имеет не более, чем счётное число точек разрыва.

Доказательство. 1. Пусть $x_1 \leqslant x_2$. Тогда легко видеть, что $(-\infty, x_1) \subseteq (-\infty, x_2)$ и $\xi^{-1}((-\infty, x_1)) \subseteq \xi^{-1}((-\infty, x_2))$. Используя свойство вероятности 3, имеем $\mathbf{P}\{\xi < x_1\} \leqslant \mathbf{P}\{\xi < x_2\}$. Следовательно, $F_\xi(x_1) \leqslant F_\xi(x_2)$.

2. Возьмём $x_0 \in \mathbb{R}$ и последовательность x_n монотонно сходящуюся к x_0 слева: $x_n \uparrow x_0$. Обозначим

$$A_0 = \{\xi < x_0\}, \quad A_n = \{\xi < x_n\},\ n = 1, 2, \dots$$

Ясно, что при этом $A_1 \subseteq A_2 \subseteq \dots \subseteq A_n \subseteq \dots$ и $\bigcup\limits_{n=1}^{\infty} A_n = A_0$. Тогда из аксиомы непрерывности следует, что $\lim\limits_{n\to\infty} F_\xi(x_n) = \lim\limits_{n\to\infty} \mathbf{P}(A_n) = \mathbf{P}(A_0) = F_\xi(x_0)$.

Из монотонности функции $F_\xi(x)$ и полученной выше сходимости следует, что соответствующая сходимость есть для произвольной последовательности $x_n \to x_0-$: $\lim\limits_{x_n\to x_0-} F_\xi(x_n) = F_\xi(x_0)$.

3. Покажем, например, что $\lim\limits_{x\to-\infty} F_\xi(x) = 0$. Возьмём невозрастающую последовательность $x_n \downarrow -\infty$. Тогда, в тех же обозначениях, что и в пункте 2, $A_1 \supseteq A_2 \supseteq \dots \supseteq A_n \supseteq \dots$, и $\bigcap\limits_{n=1}^{\infty} A_n = \emptyset$. Из непрерывности вероятности следует, что $F_\xi(-\infty) = \lim\limits_{n\to\infty} F_\xi(x_n) = \lim\limits_{n\to\infty} \mathbf{P}(A_n) = \mathbf{P}(\emptyset) = 0$. В силу монотонности функции $F_\xi(x)$ заключаем, что $\lim\limits_{x\to-\infty} F_\xi(x) = 0$.

4. Имеем $0 \leqslant F_\xi(x) \leqslant 1$. Обозначим через M_n множество точек разрыва функции $F_\xi(x)$ с величиной скачка в точках разрыва $\geqslant \frac{1}{n}$. Ясно, что тогда множество M_n конечно и содержит не более, чем n точек. А, значит, множество точек разрыва функции $F_\xi(x)$ не может быть более, чем счётным. □

Замечание 2.3. *Из определения и свойств функции распределения $F_\xi(x)$ легко вывести формулы для вычисления вероятностей*

$\mathbf{P}\{a < \xi < b\} = F_\xi(b) - F_\xi(a+)$; $\mathbf{P}\{a \leqslant \xi \leqslant b\} = F_\xi(b+) - F_\xi(a)$;

$\mathbf{P}\{a \leqslant \xi < b\} = F_\xi(b) - F_\xi(a)$; $\mathbf{P}\{a < \xi \leqslant b\} = F_\xi(b+) - F_\xi(a+)$.

2.1.5 Распределение случайной величины

Определение 2.3. *Числовая функция $G = G(x)$, $x \in \mathbb{R}$, $G : \mathbb{R} \to \mathbb{R}$ называется функцией распределения, если она обладает свойствами:*

1. $G(x)$ не убывает;

2. $G(x)$ непрерывна слева;

3. $G(-\infty) = 0$, $G(+\infty) = 1$.

Теорема 2.3. *Пусть $G = G(x)$ — функция распределения. Тогда существует вероятностное пространство $(\Omega, \mathscr{F}, \mathbf{P})$, существует случайная величина $\xi = \xi(\omega), \omega \in \Omega$, такие, что $F_\xi(x) = G(x), x \in \mathbb{R}$.*

Доказательство. В качестве Ω возьмём числовую прямую $\mathbb{R}$, в качестве σ-алгебры событий $\mathscr{F}$ рассмотрим борелевскую σ-алгебру числовой прямой $\mathscr{B}(\mathbb{R})$.

Остаётся задать вероятность. Зададим вначале вероятность $\mathbf{P}_0$ на интервалах вида $(-\infty, x)$ следующим образом:

$$\mathbf{P}_0((-\infty, x)) = G(x).$$

Продолжим функцию множеств $\mathbf{P}_0$ на алгебру $\mathscr{A}$, порождённую интервалами вида $[a, b)$, по следующей формуле:

$$\mathbf{P}_0(A) = \sum_{k=1}^{n} [G(b_k) - G(a_k)],$$

где $A = \sum\limits_{i=1}^{n} [a_k, b_k)$, $[a_k, b_k) \cap [a_l, b_l) = \emptyset$, $(k \neq l)$.

Тогда $\mathbf{P}_0$ определяет конечно-аддитивную функцию множеств, заданную на алгебре $\mathscr{A}$. Если же показать, что эта функция к тому же счётно-аддитивна, то существование искомой вероятности будет следовать из теоремы Каратеодори.

Покажем, что функция $\mathbf{P}_0$ непрерывна в $\emptyset$. Возьмём невозрастающую последовательность $A_1 \supseteq A_2 \supseteq \dots \supseteq A_n \supseteq \dots$, $\bigcap\limits_{n=1}^{\infty} A_n = \emptyset$ и покажем, что $\lim\limits_{n\to\infty} \mathbf{P}_0(A_n) = 0$.

Предположим вначале, что все эти множества содержатся в некотором замкнутом интервале: $A_n \subseteq [-N, N], N < \infty$, $(n = 1, 2, \dots)$

Из непрерывности функции $G(x)$ слева имеем

$$\mathbf{P}_0([a, b')) = G(b') - G(a) \to G(b) - G(a) = \mathbf{P}_0([a, b))$$

при $b' \uparrow b$.

Так как множества A_n представляют из себя конечные суммы интервалов вида $[a, b)$, то для каждого A_n существует множество $B_n \in \mathscr{A}$ такое, что замыкание $[B_n] \subseteq A_n$ и

$$\mathbf{P}_0(A_n) - \mathbf{P}_0(B_n) \leqslant \frac{\varepsilon}{2^n},$$

где ε — некоторое наперёд заданное положительное число.

Поскольку $\bigcap\limits_{n=1}^{\infty} A_n = \emptyset$, то и $\bigcap\limits_{n=1}^{\infty} [B_n] = \emptyset$. Множества $[B_n]$ замкнуты, следовательно, найдётся такое конечное число $n_0 = n_0(\varepsilon)$, что

$$\bigcap_{n=1}^{n_0} [B_n] = \emptyset.$$

Последнее следует из того, что $[-N, N]$ является компактом, а система множеств $\{[-N, N] \setminus [B_n]\}_{n \geqslant 1}$ образует открытое покрытие этого компакта. Хорошо известно, что тогда найдётся конечное подпокрытие:

$$\bigcup_{n=1}^{n_0} ([-N, N] \setminus [B_n]) = [-N, N].$$

Отсюда следует, что $\bigcap\limits_{n=1}^{n_0} [B_n] = \emptyset$.

Так как $A_{n_0} \subseteq A_{n_0-1} \subseteq \dots \subseteq A_1$, то

$$\mathbf{P}_0(A_{n_0}) = \mathbf{P}_0\left(A_{n_0} \setminus \bigcap_{k=1}^{n_0} B_k\right) + \mathbf{P}_0\left(\bigcap_{k=1}^{n_0} B_k\right) = \mathbf{P}_0\left(A_{n_0} \setminus \bigcap_{k=1}^{n_0} B_k\right)$$

$$\leqslant \mathbf{P}_0\left(\bigcup_{k=1}^{n_0}(A_k \setminus B_k)\right) \leqslant \sum_{k=1}^{n_0} \mathbf{P}_0(A_k \setminus B_k) \leqslant \sum_{k=1}^{n_0} \frac{\varepsilon}{2^k} \leqslant \varepsilon.$$

Следовательно $\lim\limits_{n\to\infty} \mathbf{P}_0(A_n) = 0$, а, значит, функция P_0 является счётно-аддитивной.

Откажемся теперь от условия, что все $A_n \subseteq [-N, N]$ для некоторого N. Зададим произвольное $\varepsilon > 0$ и выберем такое N, что $\mathbf{P}_0([-N, N]) > 1 - \varepsilon/2$. Так как

$$A_n = A_n \cap [-N, N] + A_n \cap \overline{[-N, N]},$$

то

$$\mathbf{P}_0(A_n) = \mathbf{P}_0(A_n \cap [-N, N]) + \mathbf{P}_0(A_n \cap \overline{[-N, N]}) \leqslant \mathbf{P}_0(A_n \cap [-N, N]) + \varepsilon/2.$$

Рассуждая как и выше, заменяя при этом A_n на $A_n \cap [-N, N]$, получим, что для достаточно больших n $\mathbf{P}_0(A_n \cap [-N, N]) \leqslant \varepsilon/2$. Тем самым $\lim\limits_{n\to\infty} \mathbf{P}_0(A_n) = 0$.

Теперь непрерывную (следовательно, счётно-аддитивную) функцию множеств $\mathbf{P}_0$ по теореме Каратеодори продолжим до вероятности $\mathbf{P}$, заданной на всей σ-алгебре борелевских множеств.

Далее, в качестве искомой случайной величины ξ возьмём тождественное отображение числовой прямой на себя: $\xi(\omega) = \omega$. Тогда

$$F_\xi(x) = \mathbf{P}\{\omega : \xi(\omega) < x\} = \mathbf{P}\{\omega : \omega < x\} = \mathbf{P}(-\infty, x) = G(x).$$

$\square$

Замечание 2.4. *Построенная в теореме 2.3 вероятность на борелевской σ-алгебре числовой прямой*

$$\mathbf{P}(B) =: \mathbf{P}_\xi(B) = \mathbf{P}\{\xi^{-1}(B)\}, \ B \in \mathscr{B}(\mathbb{R}),$$

называется ***распределением*** *или* ***законом распределения*** *вероятностей. Это распределение может быть связано с какой-то случайной величиной, а может быть и не связано. Но в последнем случае, следуя теореме 2.3, мы всегда можем такую случайную величину указать.*

2.2 Типы распределений

Пусть $\xi = \xi(\omega)$ — произвольная случайная величина на вероятностном пространстве $(\Omega, \mathscr{F}, \mathbf{P})$, $\mathbf{P}_\xi$ — её распределение вероятностей на числовой прямой. Многообразие математических моделей случайных величин есть многообразие их функций распределения, поэтому проведём классификацию случайных величин и распределений вероятностей, классифицируя функции распределения.

Существует следующая классификация функций распределения.

1. Дискретные функции распределения (дискретные распределения вероятностей, случайные величины дискретного типа).

2. Абсолютно непрерывные функции распределения (абсолютно непрерывные распределения вероятностей, случайные величины непрерывного типа).

3. Сингулярные функции распределения (сингулярные распределения вероятностей, случайные величины сингулярного типа).

Мы уже говорили, что любое распределение вероятностей может рассматриваться как распределение вероятностей некоторой случайной величины. Поэтому дадим определения типов распределений на языке случайных величин.

2.2.1 Дискретные распределения вероятностей

Определение 2.4. *Скажем, что случайная величина $\xi = \xi(\omega)$, определённая на вероятностном пространстве $(\Omega, \mathscr{F}, \mathbf{P})$, имеет дискретное распределение вероятностей, если существует конечное или счётное множество $x_1, x_2, \ldots, x_n, \ldots$ (без предельных точек) таких, что*

$$\mathbf{P}\{\xi = x_n\} = p_n > 0,\ n = 1, 2, \ldots;\ \sum_{n=1}^{\infty} p_n = 1.$$

Ясно, что функция распределения такой случайной величины будет терпеть разрывы в точках x_k и $p_k = F_\xi(x_k+) - F_\xi(x_k)$ — это скачок функции $F_\xi(x)$ в точке x_k, $k = 1, 2, \ldots$. Понятно также, что если x — точка непрерывности функции $F_\xi(x)$, то $\mathbf{P}\{\xi = x\} = 0$.

Приведём примеры дискретных распределений.

1. **Вырожденное распределение**. Если случайная величина такова, что

$$\mathbf{P}\{\xi = C\} = 1,\ C - const,$$

то распределение этой случайной величины называется вырожденным, и сама случайная величина называется вырожденной.

2. **Распределение Бернулли** с параметром $p \in (0, 1)$: $\xi \sim \mathbf{B}(p)$.

Здесь
$$\mathbf{P}\{\xi = 1\} = p,\ \mathbf{P}\{\xi = 0\} = 1 - p.$$
Эту случайную величину мы могли наблюдать в единичном опыте Бернулли (схема Бернулли) с вероятностью «успеха» $p > 0$.

3. **Биномиальное распределение** с параметрами n и p: $\xi \sim \mathbf{B}(n, p)$.

Случайную величину с этим распределением мы уже видели в схеме Бернулли. Это — число «успехов» в n независимых испытаниях Бернулли:
$$\mathbf{P}\{\xi = k\} = C_n^k p^k q^{n-k},\ \ k = 0, 1, \dots, n,\ p + q = 1.$$

4. **Геометрическое распределение** с параметром $p > 0$: $\xi \sim \mathbf{G}(p)$.
$$\mathbf{P}\{\xi = k\} = q^{k-1}p,\ \ k = 1, 2, \dots.$$
В этом случае значения случайной величины — это номер опыта в последовательности независимых испытаний Бернулли, в котором впервые выпал «успех».

5. **Гипергеометрическое распределение**: $\xi \sim \mathbf{Gg}(N, M, n)$.
$$\mathbf{P}\{\xi = m\} = \frac{C_M^m \cdot C_{N-M}^{n-m}}{C_N^n},\ \ m = 0, 1, \dots, \min(M, n).$$
Эту случайную величину мы получили в примере 1.8.

6. **Распределение Пуассона** с параметром $\lambda > 0$: $\xi \sim \mathbf{P}(\lambda)$.
$$\mathbf{P}\{\xi = k\} = \frac{\lambda^k}{k!}e^{-\lambda},\ \ k = 0, 1, 2, \dots.$$
Это распределение появилось у нас в предельной теореме Пуассона.

2.2.2 Абсолютно непрерывные распределения вероятностей

Определение 2.5. *Распределение случайной величины ξ называется абсолютно непрерывным (или случайная величина имеет абсолютно* *непрерывный тип распределения), если существует неотрицательная функция $p_\xi(x)$ такая, что для всех $x \in \mathbb{R}$*

$$F_\xi(x) = \int\limits_{-\infty}^{x} p_\xi(t)dt.$$

Функция $p_\xi(x)$ при этом называется ***функцией плотности распределения вероятностей****.*

Будем далее предполагать, что функция $p_\xi(x)$ может иметь только конечное число точек разрыва.

Свойства функции плотности

1. $p_\xi(x) \geqslant 0,\ x \in \mathbb{R}$;
2. $\int\limits_{-\infty}^{+\infty} p_\xi(x)dx = 1$;
3. $F'_\xi(x) = p_\xi(x)$ для всех точек непрерывности функции $p_\xi(x)$.

Доказательство. Первое и второе свойства очевидны из определения. Докажем третье свойство. Пусть x — точка непрерывности функции $p_\xi(x)$. Тогда, используя свойства функции распределения и теорему о среднем, получим

$$\mathbf{P}\{x \leqslant \xi < x+\Delta x\} = F_\xi(x+\Delta x)-F_\xi(x) = \int\limits_{x}^{x+\Delta x} p_\xi(t)dt = p_\xi(c)\Delta x+o(\Delta x),$$

где $x \leqslant c < x + \Delta x$. Значит,

$$F'_\xi(x) = \lim_{\Delta x \to 0} \frac{F_\xi(x+\Delta x) - F_\xi(x)}{\Delta x} = p_\xi(x).$$

□

Примеры абсолютно непрерывных распределений

Ясно, что абсолютно непрерывное распределение однозначно определяется либо функцией распределения, либо функцией плотности. Мы будем задавать такие распределения функциями плотности.

1. Равномерное распределение на отрезке $[a,b]$. Будем писать в этом случае $\xi \sim U([a,b])$.
$$p(x) = \begin{cases} \frac{1}{b-a}, & x \in [a,b]; \\ 0, & x \notin [a,b]. \end{cases}$$

2. Нормальное распределение с параметрами μ и σ^2: $\xi \sim \mathscr{N}(\mu,\sigma^2)$.
$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathbb{R}.$$
Если $\mu = 0, \sigma^2 = 1$, то такое распределение называют стандартным нормальным распределением $\mathscr{N}(0,1)$. Нормальное распределение также называют распределением Гаусса или гауссовским распределением.

3. Показательное (экспоненциальное) распределение с параметром $\alpha > 0$: $\xi \sim E(\alpha)$.
$$p(x) = \begin{cases} 0, & x \leqslant 0; \\ \alpha e^{-\alpha x}, & x > 0. \end{cases}$$

4. Распределение Коши (стандартное):
$$p(x) = \frac{1}{\pi} \cdot \frac{1}{1+x^2}, \ x \in \mathbb{R}.$$

5. Распределение Парето с параметром (a,α):
$$p(x) = \begin{cases} 0, & x \leqslant a; \\ \frac{\alpha}{a} \cdot (\frac{a}{x})^{\alpha+1}, & x > a. \end{cases}$$

2.2.3 Сингулярный тип распределения

Этот тип распределений характеризуется тем, что функция распределения $F(x)$ непрерывна, но точки роста $F(x)$ образуют множество нулевой меры Лебега.

Напомним что точка x называется точкой роста функции $F(x)$, если для $\forall \varepsilon > 0 \quad F(x+\varepsilon) - F(x-\varepsilon) > 0$.

Мы рассмотрим здесь лишь пример сингулярного распределения на числовой прямой и построим кривую Кантора, которая служит примером такой функции распределения.

Пример 2.5. *Лестница Кантора.*

Положим $F(x) = 0$, *если* $x \leqslant 0; F(x) = 1$, *если* $x > 1$. *На отрезке* $[0,1]$ *кривая строится следующим образом: отрезок* $[0,1]$ *разбиваем точками* $\frac{1}{3}, \frac{2}{3}$ *на три равные части. На интервале* $(\frac{1}{3}, \frac{2}{3})$ *полагаем* $F(x) = \frac{1}{2}$*:*

$$F(x) = \begin{cases} 1/2, & x \in (1/3, 2/3), \\ 0, & x = 0, \\ 1, & x = 1. \end{cases}$$

В точках, не принадлежащих внутреннему сегменту, $F(x)$ *определяется по непрерывности.*

Оставшиеся два отрезка снова разобьём на три равные части и на внутренних интервалах функция $F(x)$ *определяется как постоянная, равная среднему арифметическому между соседними, уже определёнными значениями* $F(x)$*:*

$$F(x) = \begin{cases} 1/2, & x \in (1/3, 2/3), \\ 1/4, & x \in (1/9, 2/9), \\ 3/4, & x \in (7/9, 8/9), \\ 0, & x = 0, \\ 1, & x = 1. \end{cases}$$

В точках, не принадлежащих таким внутренним сегментам, $F(x)$ *определяется по непрерывности.*

Этот процесс продолжается до бесконечности.

Ясно, что суммарная длина «внутренних» сегментов, на которых $F(x)$ *постоянна, равна* $\frac{1}{3} + \frac{2}{9} + \frac{4}{27} + \cdots = 1$.

В этом случае говорят, что мера Лебега на $[0,1]$ *и вероятность, заданная посредством канторовой кривой* $F(x)$*, сингулярны или ортогональны.*

Не останавливаясь более на вопросе о возможных типах функций распределения, ограничимся лишь замечанием о том, что на самом деле указанными тремя типами исчерпываются все такие функции. Точнее, произвольная функция распределения может быть представлена в виде $\alpha_1 F_1 + \alpha_2 F_2 + \alpha_3 F_3$, где F_1 — дискретная, F_2 — абсолютно непрерывная, F_3 — сингулярная функции распределения, α_i — неотрицательные числа, $\alpha_1 + \alpha_2 + \alpha_3 = 1$.

2.3 Числовые характеристики случайных величин

2.3.1 Замечания об интеграле Лебега

Пусть $(\Omega, \mathscr{F}, \mathbf{P})$ — некоторое вероятностное пространство. Здесь $\mathbf{P}$ является счётно-аддитивной конечной мерой на $\mathscr{F}$, значит, мы можем рассмотреть интеграл Лебега

$$\int\limits_{\Omega} g(\xi(\omega))\mathbf{P}(d\omega)$$

для любой измеримой по Борелю функции $g = g(x)$ и любой случайной величины $\xi = \xi(\omega)$. Выше мы видели, что случайная величина $\xi = \xi(\omega)$ индуцирует вероятность $\mathbf{P}_\xi$ на числовой прямой равенством:

$$\mathbf{P}_\xi([a, b)) = \mathbf{P}\{\xi \in [a, b)\} = F_\xi(b) - F_\xi(a).$$

Посредством этой вероятности исходный интеграл можно записать в виде (произведя замену $x = \xi(\omega)$ в интеграле Лебега):

$$\int\limits_{\Omega} g(\xi(\omega))\mathbf{P}(d\omega) = \int\limits_{\mathbb{R}} g(x)\mathbf{P}_\xi(dx).$$

Интеграл в правой части называется интегралом Лебега-Стильтьесса от функции $g(x)$ по функции $F_\xi(x)$ и записывается

$$\int\limits_{\mathbb{R}} g(x) dF_\xi(x).$$

Это выражение имеет смысл, если

$$\int\limits_{\mathbb{R}} |g(x)| dF_\xi(x) < \infty.$$

Из определения интеграла Лебега-Стильтьесса легко вывести следующие формулы для вычисления интеграла Лебега в двух особо важных случаях:

а). **Случай дискретного распределения.**

В этом случае функция распределения $F_\xi(x)$ имеет ступенчатый вид. Тогда

$$\int_{\mathbb{R}} g(x)dF_\xi(x) = \sum_k g(x_k)(F(x_k+0) - F(x_k)) = \sum_k g(x_k)\mathbf{P}\{\xi = x_k\},$$

где $x_1, x_2, \ldots, x_n, \ldots$ — точки разрыва функции $F_\xi(x)$.

b). **Случай непрерывного распределения.**

В этом случае $F_\xi(x) = \int\limits_{-\infty}^{x} p_\xi(t)dt$.

Если здесь $p_\xi(x)$ и $g(x)$ интегрируемы по Риману, то

$$\int_{\mathbb{R}} g(x)dF_\xi(x) = \int_{\mathbb{R}} g(x)p_\xi(x)dx.$$

2.3.2 Математическое ожидание

Пусть $(\Omega, \mathscr{F}, \mathbf{P})$ — вероятностное пространство, $\xi = \xi(\omega)$ — некоторая случайная величина.

Определение 2.6. *Математическим ожиданием случайной величины ξ называется число*

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

если интеграл в правой части существует.

Учитывая предыдущие замечания,

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

и мы можем конкретизировать это определение на случай дискретного и непрерывного распределений.

Дискретный тип

Здесь заданы вещественные числа $x_1, x_2, \ldots, x_n, \ldots$ и $p_k = \mathbf{P}\{\xi = x_k\}$, причём $\sum\limits_k p_k = 1$.

В этом случае

$$\mathbb{E}\xi = \sum_k x_k \mathbf{P}\{\xi = x_k\} = \sum_k x_k p_k.$$

Непрерывный тип

Случайная величина непрерывного типа задаётся с помощью функции плотности $p_\xi(x) \geqslant 0$. Тогда

$$\mathbb{E}\xi = \int\limits_{-\infty}^{+\infty} x p_\xi(x) dx.$$

Свойства математических ожиданий

Основные свойства математических ожиданий совпадают со свойствами сумм и интегралов.

$\mathbb{E}1$) Если c — const, то $\mathbb{E}c = c$;

$\mathbb{E}2$) $\mathbb{E}(c\xi_1 + \xi_2) = c\mathbb{E}\xi_1 + \mathbb{E}\xi_2$, если все интегралы существуют (c — const);

$\mathbb{E}3$) Если $a \leqslant \xi \leqslant b$, то $a \leqslant \mathbb{E}\xi \leqslant b$, в частности, $\mathbb{E}\xi \leqslant \mathbb{E}|\xi|$;

$\mathbb{E}4$) Если $\xi \geqslant 0$ и $\mathbb{E}\xi = 0$, то $\xi = 0$ с вероятностью 1, то есть $\mathbf{P}\{\xi = 0\} = 1$;

$\mathbb{E}5$) Вероятность события A может быть записана в терминах математических ожиданий:

$$\mathbf{P}(A) = \mathbb{E}I_A(\omega),$$

где $I_A(\omega)$ — индикатор события A.

Доказательства этих свойств очевидны.

Приведем примеры непосредственного вычисления математических ожиданий различных случайных величин.

1. Случайная величина с распределением Бернулли:

$$\mathbf{P}\{\xi = 1\} = p, \quad \mathbf{P}\{\xi = 0\} = 1 - p.$$

Тогда

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

2. Случайная величина, имеющая геометрическое распределение:

$$\mathbf{P}\{\xi = k\} = q^{k-1}p, \quad k = 1, 2, \ldots, \ q = 1 - p.$$

Тогда, используя метод интегрирования под знаком суммы, получим

$$\mathbb{E}\xi = \sum_{k=1}^{\infty} k \cdot q^{k-1} p = p \sum_{k=1}^{\infty} k q^{k-1} = p \left[\frac{d}{dx} \sum_{k=1}^{\infty} \int k x^{k-1} dx \right]_{x=q} =$$

$$p \left[\frac{d}{dx} \sum_{k=1}^{\infty} x^k \right]_{x=q} = p \left[\frac{d}{dx} (\frac{x}{1-x}) \right]_{x=q} = \frac{p}{(1-q)^2} = \frac{1}{p}.$$

Итак, $\mathbb{E}\xi = 1/p$.

3. Случайная величина, имеющая биномиальное распределение с параметрами n и p:

$$\mathbf{P}\{\xi = k\} = C_n^k p^k q^{n-k},\ k = 0, 1, \ldots, n.$$

Тогда

$$\mathbb{E}\xi = \sum_{k=0}^{n} k \cdot C_n^k p^k q^{n-k} = \left[k \cdot C_n^k = n \cdot C_{n-1}^{k-1} \right] =$$

$$= \sum_{k=1}^{n} n \cdot C_{n-1}^{k-1} p^k q^{n-k} = np \sum_{k=1}^{n} C_{n-1}^{k-1} p^{k-1} q^{n-k} = np(p+q)^{n-1} = np.$$

4. Случайная величина, имеющая распределение Пуассона с параметром $\lambda > 0$:

$$\mathbf{P}\{\xi = k\} = \frac{\lambda^k}{k!} e^{-\lambda},\ k = 0, 1, 2, \ldots$$

Тогда

$$\mathbb{E}\xi = \sum_{k=0}^{\infty} k \cdot \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \lambda \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda \cdot e^{-\lambda} e^{\lambda} = \lambda.$$

5. Случайная величина, имеющая равномерное распределение на $[a, b]$:

Здесь

$$p_\xi(x) = \frac{1}{b-a} I_{[a,b]}(x).$$

Тогда

$$\mathbb{E}\xi = \int_a^b x \cdot \frac{1}{b-a} dx = \frac{a+b}{2}.$$

6. Случайная величина, имеющая нормальное распределение с параметрами μ и σ^2:

   Здесь
$$p_\xi(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \ \ x \in \mathbb{R}.$$

   Тогда
$$\mathbb{E}\xi = \frac{1}{\sqrt{2\pi}\sigma} \int\limits_{-\infty}^{+\infty} x e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \left[\frac{x-\mu}{\sigma} = t \right] =$$
$$\frac{1}{\sqrt{2\pi}}\sigma \int\limits_{-\infty}^{+\infty} t e^{-\frac{t^2}{2}} dt + \mu \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{+\infty} e^{-\frac{t^2}{2}} dt = \mu.$$

7. Случайная величина, имеющая показательное распределение с параметорм $\alpha > 0$:
$$p_\xi(x) = \alpha e^{-\alpha x} I_{\{x>0\}}.$$

   Тогда
$$\mathbb{E}\xi = \int\limits_{0}^{+\infty} \alpha x e^{-\alpha x} dx = \frac{1}{\alpha}.$$

8. Случайная величина, имеющая стандартное распределение Коши:
$$p_\xi(x) = \frac{1}{\pi} \cdot \frac{1}{1+x^2}, \ \ x \in \mathbb{R}.$$

   Как известно, в этом случае интеграл
$$\int\limits_{-\infty}^{+\infty} \frac{x}{1+x^2} dx$$

   расходится, а, значит, случайная величина не имеет среднего. В этом случае можно говорить лишь о существовании последнего интеграла в смысле главного значения Коши:
$$v.p. \frac{1}{\pi} \int\limits_{-\infty}^{+\infty} \frac{x dx}{1+x^2} = 0.$$

2.3.3 Дисперсия

Определение 2.7. *Дисперсией случайной величины ξ называется число*

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

если интеграл в правой части существует.

Заметим сразу, что можно найти другую формулу для вычисления $\mathbb{D}\xi$:

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

Свойства дисперсии

D1. $\mathbb{D}\xi \geqslant 0$; $\mathbb{D}\xi = 0 \Leftrightarrow \mathbf{P}\{\xi = c\} = 1, c - const.$
D2. $\mathbb{D}[c\xi] = c^2\mathbb{D}\xi,\ \mathbb{D}(\xi + c) = \mathbb{D}\xi.$
D3. $\mathbb{D}\xi = \min\limits_b \mathbb{E}(\xi - b)^2.$

Доказательство. D1. Ясно, что $\mathbb{D}\xi = \mathbb{E}(\xi-\mathbb{E}\xi)^2 \geqslant 0$. Далее пусть $\mathbf{P}\{\xi = c\} = 1$. Тогда $\mathbb{E}\xi = c$, $\mathbb{E}\xi^2 = c^2$. Значит, $\mathbb{D}\xi = 0$.

Обратно, пусть $\mathbb{D}\xi = 0$. Так как $(\xi - \mathbb{E}\xi)^2 \geqslant 0$, то по свойствам среднего:

$$\mathbf{P}\{\xi - \mathbb{E}\xi = 0\} = 1 \Leftrightarrow \mathbf{P}\{\xi = \mathbb{E}\xi\} = 1.$$

D2. Очевидно следует из определения.

D3. Рассмотрим

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

Очевидно, что точкой минимума этой функции является точка $b = \mathbb{E}\xi$. Тогда

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

□

Определение 2.8. *Величина $\sigma = \sqrt{\mathbb{D}\xi}$ называется стандартным отклонением случайной величины ξ.*

Отметим здесь, что дисперсия $\mathbb{D}\xi$ или $\sqrt{\mathbb{D}\xi}$ характеризуют величину рассеяния (разброса) значений случайной величины относительно её математического ожидания.

2.3.4 Моментные характеристики случайных величин

Определение 2.9. *Моментом порядка k, $(k > 0)$ случайной величины ξ называется число*

$$\alpha_k = \mathbb{E}\xi^k.$$

Величина

$$\mu_k = \mathbb{E}(\xi - \mathbb{E}\xi)^k$$

называется центральным моментом порядка k.

Очевидно, что момент первого порядка — это математическое ожидание, а центральный момент второго порядка — дисперсия.

Приведем некоторые другие характеристики случайных величин.

Определение 2.10. *Число*

$$\gamma_1 = \frac{\mu_3}{\sigma^3}$$

называется коэффициентом асимметрии случайной величины ξ.

Число

$$\gamma_2 = \frac{\mu_4}{\sigma^4} - 3$$

называется коэффициентом эксцесса случайной величины ξ.

Из определения γ_1 (или μ_3) видно, что если $\gamma_1 = 0$, то положительные и отрицательные части $(\xi - \mathbb{E}\xi)^3$ «уравновешены» относительно нуля. Это означает, что значения случайной величины ξ «одинаково разбросаны» относительно её математического ожидания.

Чтобы выяснить смысл величины γ_2 подсчитаем моменты для нормального распределения $N(\mu, \sigma^2)$.

Ясно, что

$$\mathbb{E}(\xi - \mathbb{E}\xi)^{2k-1} = 0, \ \ (k = 1, 2, ...).$$

Вычислим

$$\mu_k = \mathbb{E}(\xi - \mathbb{E}\xi)^{2k} = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{+\infty}(x-\mu)^{2k}e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx = \left|\frac{x-\mu}{\sigma} = t\right| =$$
$$= \frac{\sigma^{2k}}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}t^{2k}e^{-\frac{t^2}{2}}dt = -\frac{\sigma^{2k}}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}t^{2k-1}d(e^{-\frac{t^2}{2}}) =$$
$$= -\frac{\sigma^{2k}}{\sqrt{2\pi}}\left(t^{2k-1}e^{-\frac{t^2}{2}}\Big|_{-\infty}^{+\infty} - (2k-1)\int_{-\infty}^{+\infty}t^{2k-2}e^{-\frac{t^2}{2}}dt\right) =$$
$$= \frac{\sigma^{2k}(2k-1)}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}t^{2k-2}e^{-\frac{t^2}{2}}dt = \cdots = \sigma^{2k}(2k-1)!!$$

В частности, $\mu_4 = 3\sigma^4$, и, значит, $\gamma_2 = \mu_4/\sigma^4 - 3 = 0$. Это означает, что кривая плотности нормального распределения является своеобразным «эталоном», по которому сверяют «пикообразность» кривой функции плотности других распределений.

Определение 2.11. *Пусть $F_\xi(x)$ — функция распределения случайной величины ξ, $0 < p < 1$. p-квантилью x_p распределения случайной величины ξ называется число, определяемое по формуле*

$$x_p = \inf\{x : F_\xi(x) \geqslant p\}.$$

Заметим здесь, что если $F_\xi(x)$ — строго монотонная функция распределения случайной величины ξ, то

$$x_p = F_\xi^{-1}(p),$$

то есть, является решением уравнения $F_\xi(x) = p$.

Если в определении 2.11 $p = 1/2$, то p-квантиль называется медианой.

Наконец, введем ещё одно полезное понятие.

Определение 2.12. *Пусть $p_\xi(x)$ — функция плотности случайной величины ξ непрерывного типа. Модой x_m распределения случайной величины ξ называется точка достижения локального максимума функцией плотности $p_\xi(x)$:*

$$x_m = \arg\max_x p_\xi(x).$$

Если такой максимум единственен, то говорят, что распределение случайной величины ξ унимодально.

Если ξ — случайная величина дискретного типа, модой называется такое (может, и не существует) значение x_m случайной величины ξ, что

$$\mathbf{P}\{\xi = x_m\} = \max_k \mathbf{P}\{\xi = x_k\}.$$

Мода играет некоторую роль для симметричных унимодальных распределений, которые могут не иметь математического ожидания. Выше мы видели, что случайная величина, имеющая распределение Коши с функцией плотности

$$p_\xi(x) = \frac{1}{\pi}\frac{1}{1+x^2}, \quad x \in \mathbb{R},$$

не имеет среднего. Легко видеть, что точка $x = 0$ является модой распределения и характеризует, как легко понять, центр «масс» этого распределения.

2.3.5 Неравенство Чебышёва. Правило «трёх сигм»

Теорема 2.4. *Пусть $\xi = \xi(\omega)$ — неотрицательная случайная величина. Тогда для любого $\varepsilon > 0$*

$$\mathbf{P}\{\xi \geqslant \varepsilon\} \leqslant \frac{\mathbb{E}\xi}{\varepsilon}.$$

Доказательство. Заметим, что случайная величина ξ представляется в виде

$$\xi = \xi I_{\{\xi \geqslant \varepsilon\}} + \xi I_{\{\xi < \varepsilon\}} \geqslant \xi I_{\{\xi \geqslant \varepsilon\}} \geqslant \varepsilon I_{\{\xi \geqslant \varepsilon\}}.$$

Поэтому из свойств математического ожидания

$$\mathbb{E}\xi \geqslant \varepsilon \mathbb{E}(I_{\{\xi \geqslant \varepsilon\}}) = \varepsilon \mathbf{P}\{\xi \geqslant \varepsilon\}.$$

Следовательно,

$$\mathbf{P}\{\xi \geqslant \varepsilon\} \leqslant \frac{\mathbb{E}\xi}{\varepsilon}.$$

□

Следствие 2.1. *Для произвольной случайной величины $\xi = \xi(\omega)$*

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

Доказательство. Положим $\eta = (\xi - \mathbb{E}\xi)$. По теореме 2.4

$$\mathbf{P}\{\eta^2 \geqslant \varepsilon^2\} \leqslant \frac{\mathbb{E}\eta^2}{\varepsilon^2}.$$

Но

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

□

Из неравенства Чебышёва легко вывести так называемое «правило трёх сигм»:

Теорема 2.5. *Для произвольной случайной величины $\xi = \xi(\omega)$ вероятность попадания её значений в интервал $(\mu - 3\sigma, \mu + 3\sigma)$ больше или равна $8/9$, где $\mu = \mathbb{E}\xi, \sigma^2 = \mathbb{D}\xi$.*

Доказательство. Непосредственно следует из следствия 2.1

$$\mathbf{P}\{\mu - 3\sigma < \xi < \mu + 3\sigma\} = \mathbf{P}\{|\xi - \mu| < 3\sigma\} =$$

$$= 1 - \mathbf{P}\{|\xi - \mu| \geqslant 3\sigma\} \geqslant 1 - \frac{\mathbb{D}\xi}{9\sigma^2} = 1 - \frac{\sigma^2}{9\sigma^2} = \frac{8}{9}.$$

□

2.4 Случайные векторы. Независимость

Пусть $(\Omega, \mathscr{F}, \mathbf{P})$ — некоторое вероятностное пространство, $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$ — n-мерное евклидово пространство с борелевской σ-алгеброй.

Определение 2.13. *Упорядоченный набор $\xi(\omega) = (\xi_1(\omega), \xi_2(\omega), \dots, \xi_n(\omega))$ случайных величин назовём n-мерным случайным вектором.*

Согласно определению случайный вектор ставит в соответствие любой точке $\omega \in \Omega$ некоторую точку n-мерного евклидова пространства $\mathbb{R}^n$.

Аналогично тому, как это было для случайных величин, можно доказать, что $\xi = (\xi_1, \xi_2, \dots, \xi_n)$ является случайным вектором тогда и только тогда, если для любого $B \in \mathscr{B}(\mathbb{R}^n)$

$$\xi^{-1}(B) \in \mathscr{F}.$$

Точно также, как случайная величина задаёт распределение на числовой прямой, случайный вектор задаёт распределение в $\mathbb{R}^n$: для любого $B \in \mathscr{B}(\mathbb{R}^n)$

$$\mathbf{P}_\xi(B) = \mathbf{P}\{\xi^{-1}(B)\}$$

Определение 2.14. *Функция*

$$F_{\xi_1, \xi_2, \dots, \xi_n}(x_1, x_2, \dots, x_n) = \mathbf{P}\{\xi_1 < x_1, \xi_2 < x_2, \dots, \xi_n < x_n\}$$

называется функцией распределения случайного вектора $(\xi_1, \xi_2, \dots, \xi_n)$.

Замечание 2.5. *Функцию $F_{\xi_1, \xi_2, \dots, \xi_n}(x_1, x_2, \dots, x_n)$ часто называют совместной функцией распределения случайных величин $\xi_1, \xi_2, \dots, \xi_n$.*

Определение 2.15. *Случайный вектор $(\xi_1, \xi_2 \dots \xi_n)$ называется случайным вектором дискретного типа, если существует не более, чем счётное множество (x_{ij}), $(i = 1, 2, \dots, n;\ j = 1, 2, \dots)$ вещественных чисел, такое что*

$$\mathbf{P}\{\xi_i = x_{ij}\} = p_{ij} > 0, \quad \sum_{i,j} p_{ij} = 1.$$

Определение 2.16. *Случайный вектор $(\xi_1, \xi_2 \dots \xi_n)$ называется случайным вектором непрерывного типа, если существует неотрицательная функция $p_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n)$, такая что*

$$F_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n) = \int\limits_{-\infty}^{x_1} \cdots \int\limits_{-\infty}^{x_n} p_{\xi_1, \dots, \xi_n}(t_1, \dots, t_n) dt_1 \dots dt_n.$$

Функция $p_{\xi_1\ldots\xi_n}(x_1,\ldots,x_n)$ при этом называется n-мерной функцией плотности случайного вектора $(\xi_1,\ldots,\xi_n)$ или совместной функцией плотности случайных величин $\xi_1,\ldots,\xi_n$.

Свойства n-мерной функции распределения

1) $\lim\limits_{x_1\to+\infty} F_{\xi_1,\ldots\xi_n}(x_1,\ldots x_n) = F_{\xi_2,\ldots\xi_n}(x_2,\ldots,x_n)$;
2) $\lim\limits_{x_1\to-\infty} F_{\xi_1,\ldots\xi_n}(x_1,\ldots x_n) = 0$;

То, что к бесконечности стремится x_1, несущественно, здесь может быть любая координата.

3) С вероятностью единица имеет место равенство:

$$\frac{\partial^n F_{\xi_1\ldots\xi_n}(x_1,\ldots,x_n)}{\partial x_1\ldots\partial x_n} = p_{\xi_1\ldots\xi_n}(x_1,\ldots,x_n).$$

Доказательство этих свойств точно такое же, как и в одномерном случае.

По многомерной функции плотности можно восстановить одномерные функции плотности компонент случайных величин $\xi_1,\ldots,\xi_n$.

Действительно,

$$F_{\xi_1}(x_1) = \int\limits_{-\infty}^{x_1}\left(\int\limits_{-\infty}^{+\infty}\cdots\int\limits_{-\infty}^{+\infty} p_{\xi_1,\ldots,\xi_n}(t_1,\ldots,t_n)dt_2\ldots dt_n\right)dt_1 = \int\limits_{-\infty}^{x_1} p_{\xi_1}(t_1)dt_1.$$

Значит,

$$p_{\xi_1}(x_1) = \int\limits_{-\infty}^{+\infty}\cdots\int\limits_{-\infty}^{+\infty} p_{\xi_1\ldots\xi_n}(t_1,\ldots,t_n)dt_2\ldots dt_n.$$

Определение 2.17. *Пусть $\xi = \xi(\omega)$ — произвольная случайная величина. $\sigma-$алгебра*

$$\mathscr{F}^{\xi} = \xi^{-1}(\mathscr{B}(\mathbb{R})) =: \{\xi^{-1}(B) : B \in \mathscr{B}(\mathbb{R})\}$$

называется $\sigma-$алгеброй, порождённой случайной величиной $\xi = \xi(\omega)$.

Определение 2.18. *Случайные величины $\xi_1,\xi_2,\ldots,\xi_n$ называются независимыми, если $\sigma-$алгебры $\mathscr{F}^{\xi_1},\mathscr{F}^{\xi_2},\ldots,\mathscr{F}^{\xi_n}$ независимы.*

Другими словами, случайные величины $\xi_1,\xi_2,\ldots,\xi_n$ независимы, если для любых $B_1,B_2,\ldots,B_n \in \mathscr{B}(\mathbb{R})$

$$\mathbf{P}\{\xi_1\in B_1,\xi_2\in B_2,\ldots,\xi_n\in B_n\} = \mathbf{P}\{\xi_1\in B_1\}\mathbf{P}\{\xi_2\in B_2\}\ldots\mathbf{P}\{\xi_n\in B_n\}$$

Теорема 2.6. *Для того, чтобы случайные величины $\xi_1, \xi_2, \ldots, \xi_n$ были независимы, необходимо и достаточно, чтобы для всех $(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$*

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

Доказательство. Необходимость очевидна, так как достаточно взять $B_i = (-\infty, x_i),\ i = \overline{1, n}$.

Достаточность. Докажем это утверждение для случая $n = 2$, то есть, то, что из условия $F_{\xi_1,\xi_2}(x_1, x_2) = F_{\xi_1}(x_1) \cdot F_{\xi_2}(x_2)$ следует, что $\mathbf{P}(\xi_1 \in B_1, \xi_2 \in B_2) = \mathbf{P}\{\xi_1 \in B_1\} \cdot \mathbf{P}\{\xi_2 \in B_2\}$. Для произвольного n доказательство может быть проведено по индукции.

Докажем вначале, что

$$\mathbf{P}\{a_1 \leqslant \xi_1 < b_1, a_2 \leqslant \xi_2 < b_2\} = F_{\xi_1,\xi_2}(b_1+, b_2+) - F_{\xi_1,\xi_2}(a_1, b_2+) - \\ - F_{\xi_1,\xi_2}(b_1+, a_2) + F_{\xi_1,\xi_2}(a_1, a_2).$$

Действительно,

$$\{a_1 \leqslant \xi_1 < b_1, a_2 \leqslant \xi_2 < b_2\} = \{\xi_1 < b_1, \xi_2 < b_2\} \setminus [\{\xi_1 \leqslant a_1, \xi_2 < b_2\} \cup \\ \cup \{\xi_1 < b_1, \xi_2 \leqslant a_2\}] \cup \{\xi_1 \leqslant a_1, \xi_2 \leqslant a_2\}.$$

Тогда

$$\mathbf{P}\{a_1 \leqslant \xi_1 < b_1, a_2 \leqslant \xi_2 < b_2\} = F_{\xi_1,\xi_2}(b_1+, b_2+) - F_{\xi_1,\xi_2}(a_1, b_2+) - \\ - F_{\xi_1,\xi_2}(b_1+, a_2) + F_{\xi_1,\xi_2}(a_1, a_2).$$

Итак, для прямоугольников $B = B_1 \times B_2$, где $B_i = [a_i, b_i), (i = 1, 2)$ имеем

$$\mathbf{P}\{(\xi_1, \xi_2) \in B\} = \mathbf{P}\{a_1 \leqslant \xi_1 < b_1, a_2 \leqslant \xi_2 < b_2\} = F_{\xi_1}(b_1+) \cdot F_{\xi_2}(b_2+) - \\ - F_{\xi_1}(a_1) \cdot F_{\xi_2}(b_2+) - F_{\xi_1}(b_1+) \cdot F_{\xi_2}(a_2) + F_{\xi_1}(a_1) \cdot F_{\xi_2}(a_2) = \\ = [F_{\xi_1}(b_1+) - F_{\xi_1}(a_1)][F_{\xi_2}(b_2+) - F_{\xi_2}(a_2)] = \mathbf{P}\{\xi_1 \in B_1\} \cdot \mathbf{P}\{\xi_2 \in B_2\}.$$

Тогда по теореме о продолжении вероятности с алгебры, порожденной прямоугольниками, на борелевскую σ-алгебру мы можем распространить результат на всю борелевскую σ-алгебру на $\mathbb{R}^2$. □

Следствие 2.2. *Пусть (ξ_1, ξ_2) — дискретный случайный вектор с распределением $\mathbf{P}\{\xi_1 = x_{1i}, \xi_2 = x_{2j}\} = p_{ij},\ \sum\limits_{i,j} p_{ij} = 1$. Случайные величины ξ_1 и ξ_2 независимы тогда и только тогда, если*

$$\mathbf{P}\{\xi_1 = x_{1i}, \xi_2 = x_{2j}\} = \mathbf{P}\{\xi_1 = x_{1i}\} \cdot \mathbf{P}\{\xi_2 = x_{2j}\}.$$

Следствие 2.3. *Пусть $p_{\xi_1,\ldots,\xi_n}(x_1,\ldots,x_n)$ — n-мерная функция плотности случайного вектора $(\xi_1,\ldots,\xi_n)$. Случайные величины $\xi_1,\xi_2,\ldots\xi_n$ независимы тогда и только тогда, если*

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

Пример 2.6. *В круг O_r радиуса r бросается точка. Координаты (x,y) этой точки в некоторой системе координат представляют собой реализацию случайного вектора (ξ_1,ξ_2). Проверим независимость координат этого вектора.*

Решение.

Вычислим вначале функцию распределения случайного вектора (ξ_1,ξ_2):

$$F_{\xi_1,\xi_2}(x_1,x_2) = \mathbf{P}\{\xi_1 < x_1, \xi_2 < x_2\}.$$

Центр круга совместим для удобства с центром координат. Заметим сразу, что если точка (x_1,x_2) не содержится в круге, то функция распределения будет зависеть только от одной координаты, или вообще от них не зависеть. Значит, вне круга функция плотности (как вторая производная по x_1 и x_2) будет равна 0.

Пусть точка (x_1,x_2) содержится в круге. Тогда, применяя геометрические вероятности, имеем:

$$F_{\xi_1,\xi_2}(x_1,x_2) = \mathbf{P}\{\xi_1 < x_1, \xi_2 < x_2\} = \frac{S(x_1,x_2)}{\pi r^2}.$$

Найдем теперь $S(x_1,x_2)$.

$$S(x_1,x_2) = \int\limits_{-r}^{-\sqrt{r^2-x_2^2}} du \int\limits_{-\sqrt{r^2-u^2}}^{\sqrt{r^2-u^2}} dv +$$

$$+ \int\limits_{-\sqrt{r^2-x_2^2}}^{x^1} du \int\limits_{-\sqrt{r^2-u^2}}^{x^2} dv = x_1x_2 + f_1(x_1) + f_2(x_2).$$

Тогда функция плотности будет равна:

$$p_{\xi_1,\xi_n}(x_1,x_2) = \frac{\partial^2 F_{\xi_1,\xi_2}(x_1,x_2)}{\partial x_1 \partial x_2} = \frac{1}{\pi r^2}, \quad (x_1,x_2) \in O_r.$$

Полученная функция является функцией плотности так называемого равномерного распределения в круге.

Отсюда легко видеть, что

$$p_{\xi_1}(x_1) = \int\limits_{-\infty}^{+\infty} p_{\xi_1,\xi_2}(x_1,x_2)dx_2 = \frac{1}{\pi r^2} \int\limits_{-\sqrt{r^2-x_1^2}}^{\sqrt{r^2-x_1^2}} dx_2 =$$

$$= \frac{2}{\pi r^2}\sqrt{r^2-x_1^2},\ x_1 \in [-r,r].$$

Точно так же

$$p_{\xi_2}(x_2) = \frac{2}{\pi r^2}\sqrt{r^2-x_2^2},\ x_2 \in [-r,r].$$

Ясно, что компоненты случайного вектора (ξ_1,ξ_2) в этом случае не являются независимыми. □

Обратимся к свойствам независимых случайных величин.

Теорема 2.7. *Пусть случайные величины ξ_1,ξ_2 независимы. Тогда*
1) Для любых борелевских функций $g_1 = g_1(x), g_2 = g_2(x)$ случайные величины $\eta_1 = g_1(\xi_1)$ и $\eta_2 = g_2(\xi_2)$ независимы.
2) $\mathbb{E}(\xi_1\xi_2) = \mathbb{E}\xi_1\mathbb{E}\xi_2$.
3) $\mathbb{D}(\xi_1+\xi_2) = \mathbb{D}(\xi_1)+\mathbb{D}(\xi_2)$.

Доказательство. 1) Пусть ξ_1,ξ_2 независимы, $B_1,B_2 \in \mathscr{B}(\mathbb{R})$. Тогда

$$\mathbf{P}\{\eta_1 \in B_1, \eta_2 \in B_2\} = \mathbf{P}\{g_1(\xi_1) \in B_1, g_2(\xi_2) \in B_2\} =$$
$$\mathbf{P}\{\xi_1 \in g_1^{-1}(B_1), \xi_2 \in g_2^{-1}(B_2)\} = \mathbf{P}\{\xi_1 \in g_1^{-1}(B_1)\} \cdot \mathbf{P}\{\xi_2 \in g_2^{-1}(B_2)\} =$$
$$= \mathbf{P}\{\eta_1 \in B_1\} \cdot P\{\eta_2 \in B_2\}.$$

2) Так как случайные величины ξ_1,ξ_2 независимы, то $p_{\xi_1,\xi_2}(x_1,x_2) = p_{\xi_1}(x_1) \cdot p_{\xi_2}(x_2)$.

Тогда

$$\mathbb{E}(\xi_1,\xi_2) = \int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty} x_1x_2p_{\xi_1,\xi_2}(x_1,x_2)dx_1dx_2 =$$

$$= \int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty} x_1x_2p_{\xi_1}(x_1) \cdot p_{\xi_2}(x_2)dx_1dx_2 =$$

$$= \int\limits_{-\infty}^{+\infty} x_1p_{\xi_1}(x_1)dx_1 \cdot \int\limits_{-\infty}^{+\infty} x_2p_{\xi_2}(x_2)dx_2 = E\xi_1 \cdot E\xi_2.$$

$$
\begin{aligned}
3)\mathbb{D}(\xi_1+\xi_2) = \mathbb{E}(\xi_1+\xi_2-\mathbb{E}(\xi_1+\xi_2))^2 &= \mathbb{E}[(\xi_1-\mathbb{E}\xi_1)+(\xi_2-\mathbb{E}\xi_2)]^2 = \\
= \mathbb{E}[(\xi_1-\mathbb{E}\xi_1)^2 + 2(\xi_1-\mathbb{E}\xi_1)(\xi_2-\mathbb{E}\xi_2) &+ (\xi_2-\mathbb{E}\xi_2)^2] = \\
= \mathbb{D}\xi_1+\mathbb{D}\xi_2+2\mathbb{E}(\xi_1-\mathbb{E}\xi_1)\cdot\mathbb{E}(\xi_2-\mathbb{E}\xi_2) &= \mathbb{D}\xi_1+\mathbb{D}\xi_2.
\end{aligned}
$$

□

Ясно, что эта теорема распространяется на произвольное число случайных величин.

Пример 2.7. *С точки зрения независимых случайных величин рассмотрим случайную величину ξ, имеющую биномиальное распределение с параметрами $(n,p) : \xi \sim B(n,p)$.*

Мы видели, что ξ — это число «успехов» в n независимых испытаниях Бернулли. Пусть ξ_i, $(i=\overline{1,n})$ — независимые случайные величины, имеющие распределение Бернулли с параметром $p>0$:

$$\mathbf{P}\{\xi_i=1\}=p, \mathbf{P}\{\xi_i=0\}=1-p\,(=q),\ (i=\overline{1,n}).$$

Тогда случайная величина ξ может быть представлена в виде

$$\xi=\xi_1+\xi_2+\cdots+\xi_n.$$

Выше мы видели, что $\mathbb{E}\xi=np$. Его можно было вычислить и так:

$$\mathbb{E}\xi=\mathbb{E}\sum_{i=1}^n\xi_i=\sum_{i=1}^n\mathbb{E}\xi_i=np.$$

Вычислим дисперсию $\mathbb{D}\xi$. По теореме 2.7

$$\mathbb{D}\xi=\mathbb{D}\sum_{i=1}^n\xi_i=\sum_{i=1}^n\mathbb{D}\xi_i=npq.$$

2.4.1 Понятие свёртки распределений

Представление случайных величин как суммы независимых случайных величин имеет распространение во многих задачах теории вероятностей и математической статистики. Поэтому остановимся подробнее на исследовании распределений сумм независимых случайных величин.

Пусть ξ_1,ξ_2 — независимые случайные величины с функциями распределения F_{ξ_1},F_{ξ_2} соответственно. Из независимости случайных величин ξ_1,ξ_2 следует, что $F_{\xi_1,\xi_2}(x_1,x_2)=F_{\xi_1}(x_1)\cdot F_{\xi_2}(x_2)$.

Рассмотрим

$$F_{\xi_1+\xi_2}(t) = \mathbf{P}\{\xi_1 + \xi_2 < t\} = \iint\limits_{\{x_1+x_2<t\}} dF_{\xi_1,\xi_2}(x_1, x_2).$$

Воспользовавшись теоремой Фубини, получим

$$F_{\xi_1+\xi_2}(t) = \iint\limits_{\{x_1+x_2<t\}} dF_{\xi_1}(x_1)dF_{\xi_2}(x_2) = \int\limits_{-\infty}^{+\infty} dF_{\xi_1}(x_1) \int\limits_{-\infty}^{t-x_1} dF_{\xi_2}(x_2) =$$
$$= \int\limits_{-\infty}^{+\infty} F_{\xi_2}(t - x_1)dF_{\xi_1}(x_1).$$

Итак,

$$F_{\xi_1+\xi_2}(t) = \int\limits_{-\infty}^{+\infty} F_{\xi_2}(t - x)dF_{\xi_1}(x),$$

и называется свёрткой функции распределения $F_{\xi_1}(x)$ и $F_{\xi_2}(x)$ и обозначается $F_{\xi_1} * F_{\xi_2}(t)$.

Отметим здесь, что если хотя бы одна из функций распределения имеет плотность, то и свёртка также её имеет.

Пусть, например,

$$F_{\xi_2}(x) = \int\limits_{-\infty}^{x} p_{\xi_2}(u)d(u).$$

Тогда

$$F_{\xi_1+\xi_2}(x) = \int\limits_{-\infty}^{+\infty} dF_{\xi_1}(t) \int\limits_{-\infty}^{x} p_{\xi_2}(u - t)du = \int\limits_{-\infty}^{x} du \int\limits_{-\infty}^{+\infty} p_{\xi_2}(u - t)dF_{\xi_1}(t).$$

Значит,

$$p_{\xi_1+\xi_2}(u) = \int\limits_{-\infty}^{+\infty} p_{\xi_2}(u - t)dF_{\xi_1}(t).$$

В частности, если оба распределения имеют плотность, то

$$p_{\xi_1+\xi_2}(u) = \int\limits_{-\infty}^{+\infty} p_{\xi_2}(u - t)p_{\xi_1}(t)dt.$$

Функция $p_{\xi_1+\xi_2}(u)$ называется свёрткой функций плотности $p_{\xi_1}(x)$ и $p_{\xi_2}(x)$ и обозначается $p_{\xi_1} * p_{\xi_2}(u)$.

Пример 2.8. *Пусть случайные величины ξ_1, ξ_2 независимы и равномерно распределены на отрезке $[0,1]$, то есть, они имеют одну и ту же функцию плотности:*

$$p(x) = \begin{cases} 1, & x \in [0,1] \\ 0, & x \notin [0,1] \end{cases}$$

Тогда функция плотности суммы случайных величин $\xi_1 + \xi_2$ равна

$$p_{\xi_1+\xi_2}(x) = \begin{cases} 0, & x \notin [0,2] \\ x, & x \in [0,1] \\ 2-x, & x \in [1,2] \end{cases}$$

Решение.

Воспользуемся выведенной ранее формулой для вычисления свёртки. При использовании этой формулы надо учесть, что подинтегральная функция будет отлична от нуля в том случае, если

$$0 \leqslant x \leqslant 1,\ 0 \leqslant x - t \leqslant 1.$$

Тогда

$$p_{\xi_1+\xi_2}(x) = \begin{cases} \int_0^x dt, & x \in [0,1] \\ \int_{x-1}^1 dt, & x \notin [1,2] \end{cases} = \begin{cases} x, & x \in [0,1] \\ 2-x, & x \in [1,2]. \end{cases}$$

□

Устойчивость некоторых распределений относительно свёртки

Определение 2.19. *Функция распределения $F(x)$ называется устойчивой, если для любых $a_1, a_2, b_1 > 0, b_2 > 0$ найдутся такие a и $b > 0$, такие что*

$$F\left(\frac{x-a_1}{b_1}\right) * F\left(\frac{x-a_2}{b_2}\right) = F\left(\frac{x-a}{b}\right).$$

Другими словами, распределение случайной величины ξ называется устойчивым, если случайные величины ξ и $\xi_1 + \xi_2$ имеют один и тот же тип (закон) распределения, где ξ_1, ξ_2 — независимые одинаково распределенные случайные величины.

Предложение 2.1. *Пусть* $\xi_1 \sim \mathscr{N}(a_1, \sigma_1^2)$, $\xi_2 \sim \mathscr{N}(a_2, \sigma_2^2)$. *Тогда*

$$\xi_1 + \xi_2 \sim \mathscr{N}(a_1 + a_2, \sigma_1^2 + \sigma_2^2).$$

Доказательство. Имеем

$$p_{\xi_1}(x) = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(x-a_1)^2}{2\sigma_1^2}}, \; p_{\xi_2}(x) = \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{(x-a_2)^2}{2\sigma_2^2}}.$$

Тогда

$$p_{\xi_1+\xi_2}(x) = \int\limits_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma_1} \cdot \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{(t-a_1)^2}{2\sigma_1^2}} \cdot e^{-\frac{(x-t-a_2)^2}{2\sigma_2^2}} dt =$$

$$\frac{1}{2\pi\sigma_1\sigma_2} \int\limits_{-\infty}^{+\infty} e^{-\frac{\sigma_2^2(t-a_1)^2 + \sigma_1^2(x-t-a_2)^2}{2\sigma_1^2\sigma_2^2}} dt = \cdots = \frac{1}{\sqrt{2\pi(\sigma_1^2+\sigma_2^2)}} e^{-\frac{(x-a_1-a_2)^2}{2(\sigma_1^2+\sigma_2^2)}}.$$

□

Предложение 2.2. *Пусть* $\xi_1 \sim B(n_1, p)$, $\xi_2 \sim B(n_2, p)$. *Тогда*

$$\xi_1 + \xi_2 \sim B(n_1 + n_2, p).$$

Доказательство. Достаточно вспомнить, что случайные величины с биномиальным распределением представляются в виде суммы независимых одинаково распределённых случайных величин с распределением Бернулли. □

2.4.2 Ковариация и её свойства

Пусть $(\Omega, \mathscr{F}, \mathbf{P})$ — некоторое вероятностное пространство, $\xi_1 = \xi_1(\omega)$, $\xi_2 = \xi_2(\omega)$ — случайные величины.

Определение 2.20. *Ковариацией случайных величин* ξ_1 *и* ξ_2 *называется число*

$$Cov(\xi_1, \xi_2) = \mathbb{E}[(\xi_1 - \mathbb{E}\xi_1)(\xi_2 - \mathbb{E}\xi_2)]$$

Точно также, как это было для дисперсии ξ, получим

$$Cov(\xi_1, \xi_2) = \mathbb{E}\xi_1\xi_2 - \mathbb{E}\xi_1 E\xi_2.$$

Очевидно, что

$$Cov(\xi, \xi) = \mathbb{D}\xi.$$

Свойства ковариации

1. $Cov(\xi_1, \xi_2) = Cov(\xi_2, \xi_1)$.
2. $Cov(c \cdot \xi_1, \xi_2) = c \cdot Cov(\xi_1, \xi_2)$, c — const.
3. Если случайные величины ξ_1, ξ_2 независимы, то $Cov(\xi_1, \xi_2) = 0$.
4. Если ξ_1, ξ_2 — произвольные случайные величины, то

$$\mathbb{D}(\xi_1 + \xi_2) = \mathbb{D}\xi_1 + \mathbb{D}\xi_2 + 2Cov(\xi_1, \xi_2).$$

Доказательства этих свойств очевидно из определения.

Пусть далее $(\xi_1, \xi_2, ..., \xi_n)$ — случайный вектор.
Рассмотрим

$$\lambda_{ij} = Cov(\xi_i, \xi_j), (i, j = 1, ..., n).$$

Матрица $\Lambda = (\lambda_{ij})$ называется ковариационной матрицей случайного вектора $(\xi_1, \xi_2, ..., \xi_n)$.

Теорема 2.8. *Для того, чтобы матрица $\Lambda = (\lambda_{ij})$ порядка $n \times n$ была ковариационной матрицей некоторого случайного вектора $(\xi_1, \xi_2, ..., \xi_n)$, необходимо и достаточно, чтобы она была симметричной и неотрицательно определенной.*

Доказательство. Пусть $(\xi_1, \xi_2, ..., \xi_n)$ — случайный вектор, (λ_{ij}) — его ковариационная матрица. Покажем, что она неотрицательно определена, то есть, для любых $c_1, c_2, ..., c_n$

$$\sum_{i,j} c_i c_j \lambda_{ij} \geqslant 0.$$

Действительно, обозначим

$$\eta_n = c_1\xi_1 + ... + c_n\xi_n.$$

Тогда

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

$$= \mathbb{E}\left[\sum_{i=1}^{n} c_i(\xi_i - \mathbb{E}\xi_i)\right]^2 = \mathbb{E}\left[\sum_{i,j} c_i c_j (\xi_i - \mathbb{E}\xi_i)(\xi_j - \mathbb{E}\xi_j)\right] =$$

$$= \sum_{i,j} c_i c_j \mathbb{E}(\xi_i - \mathbb{E}\xi_i)(\xi_j - \mathbb{E}\xi_j) = \sum_{i,j} c_i c_j Cov(\xi_i, \xi_j) = \sum_{i,j} c_i c_j \lambda_{ij}.$$

Так как в левой части равенства стоит дисперсия, то для любых $c_1, ..., c_n$ имеем

$$\sum_{i,j} \lambda_{ij} c_i c_j \geqslant 0.$$

Значит, матрица $\Lambda = (\lambda_{ij})$ является неотрицательно определенной. Симметричность матрицы Λ следует из свойства 1 ковариации.

Обратно, пусть $\Lambda = (\lambda_{ij})$ — симметричная, неотрицательно определенная матрица. Из теории матриц известно, что тогда существует такая ортогональная матрица Q (то есть, $QQ^* = E$, где E — единичная матрица), что

$$Q^* \Lambda Q = D,$$

где

$$D = \begin{pmatrix} d_1 & ... & 0 \\ ... & ... & ... \\ 0 & ... & d_n \end{pmatrix}$$

— диагональная матрица с неотрицательными элементами $d_1, d_2, ..., d_n$. Значит,

$$\Lambda = QDQ^* = (QB)(B^*Q^*) = (QB)(QB)^*,$$

где B — диагональная матрица с элементами $b_i = \sqrt{d_i}$ $(i = 1, ..., n)$.

Положим $A = QB$,. Тогда $\Lambda = AA^*$.

Пусть далее $\eta_1, \eta_2, ..., \eta_n$ — последовательность независимых одинаково распределённых случайных величин с распределением $\mathcal{N}(0, 1)$. Тогда случайный вектор

$$\xi = (\xi_i) = A(\eta_i) = A\eta$$

и есть искомый вектор. Действительно,

$$\mathbb{E}\xi\xi^* = \mathbb{E}(A\eta)(A\eta)^* = A\mathbb{E}(\eta\eta^*)A^* = AEA^* = AA^* = \Lambda.$$

□

Так как матрица $\Lambda = (\lambda_{ij})$ неотрицательно определенна, то по критерию Сильвестра все её главные миноры неотрицательны. Значит, при $n = 2$

$$\Delta_2 = \begin{vmatrix} D\xi_1 & Cov(\xi_1, \xi_2) \\ Cov(\xi_2, \xi_1) & D\xi_2 \end{vmatrix} = D\xi_1 D\xi_2 - Cov^2(\xi_1, \xi_2) \geqslant 0.$$

Следовательно,

$$|Cov(\xi_1, \xi_2)| \leqslant \sqrt{D\xi_1 D\xi_2}.$$

Ясно, что если ξ_1, ξ_2 независимы, то $Cov(\xi_1, \xi_2) = 0$. В случае, если $Cov(\xi_1, \xi_2) \neq 0$, то случайные величины ξ_1, ξ_2 зависимы. В качестве количественной характеристики такой зависимости случайных величин ξ_1, ξ_2 используется коэффициент корреляции $\rho(\xi_1, \xi_2)$:

$$\rho(\xi_1, \xi_2) = \frac{Cov(\xi_1, \xi_2)}{\sqrt{D\xi_1 D\xi_2}}.$$

Свойства коэффициента корреляции

1. $|\rho(\xi_1, \xi_2)| \leqslant 1$.
2. Если ξ_1, ξ_2 независимы, то $\rho(\xi_1, \xi_2) = 0$.
3. $|\rho(\xi_1, \xi_2)| = 1$ тогда и только тогда, если существуют такие $a \neq 0$ и b, что с вероятностью единица

$$\xi_2 = a\xi_1 + b.$$

Свойства $1, 2$ очевидны из определения.

Докажем свойство 3. Пусть $\xi_2 = a\xi_1 + b$ с вероятностью единица. Положим $\mathbb{E}\xi_1 = m,\ \mathbb{D}\xi_1 = \sigma^2$. Тогда

$$\mathbb{E}\xi_2 = \mathbb{E}(a\xi_1 + b) = am + b,\ \mathbb{D}\xi_2 = \mathbb{D}(a\xi_1 + b) = a^2\mathbb{D}\xi_1 = a^2\sigma^2,$$

$$Cov(\xi_1, \xi_2) = \mathbb{E}[(\xi_1 - m)(\xi_2 - (am + b))] =$$
$$= \mathbb{E}[(\xi_1 - m)(a\xi_1 + b - am - b)] = a\mathbb{E}(\xi_1 - m)^2 = a\mathbb{D}\xi_1 = a\sigma^2.$$

Значит,

$$\rho(\xi_1, \xi_2) = \frac{Cov(\xi_1, \xi_2)}{\sqrt{\sigma^2 a^2 \sigma^2}} = \frac{a\sigma^2}{|a|\sigma^2} = \frac{a}{|a|} = sign(a),$$

то есть, $|\rho(\xi_1, \xi_2)| = 1$.

Обратно, пусть $|\rho(\xi_1, \xi_2)| = 1$, например, $\rho(\xi_1, \xi_2) = 1$. Тогда

$$\mathbb{D}\left(\frac{\xi_1 - \mathbb{E}\xi_1}{\sqrt{\mathbb{D}\xi_1}} - \frac{\xi_2 - \mathbb{D}\xi_2}{\sqrt{\mathbb{D}\xi_2}}\right) = \mathbb{D}\left(\frac{\xi_1 - \mathbb{E}\xi_1}{\sqrt{\mathbb{D}\xi_1}}\right) + \mathbb{D}\left(\frac{\xi_2 - \mathbb{E}\xi_2}{\sqrt{\mathbb{D}\xi_2}}\right) -$$
$$-2\frac{1}{\sqrt{\mathbb{D}\xi_1\mathbb{D}\xi_2}}Cov(\xi_1, \xi_2) = \frac{1}{\mathbb{D}\xi_1}\mathbb{D}\xi_1 + \frac{1}{\mathbb{D}\xi_2}\mathbb{D}\xi_2 - 2 = 0.$$

Это означает (по свойству дисперсии), что

$$\frac{\xi_1 - \mathbb{E}\xi_1}{\sqrt{\mathbb{D}\xi_1}} - \frac{\xi_2 - \mathbb{E}\xi_2}{\sqrt{\mathbb{D}\xi_2}} = c$$

с вероятностью единица, где c — некоторая константа.

Отсюда следует, что ξ_1, ξ_2 линейно зависимы, a и b несложно вычислить из последнего равенства.

Итак, коэффициент корреляции выражает «тесноту» линейной зависимости случайных величин.

Отметим здесь, что если $\rho(\xi_1, \xi_2) = 0$ (некоррелированность), то это ещё не означает, что ξ_1, ξ_2 являются независимыми.

Это показывает следующий пример.

Пример 2.9. *Пусть случайный вектор (ξ_1, ξ_2) равномерно распределён в круге O_r радиуса r. Имеем:*

$$p_{\xi_1\xi_2}(x_1, x_2) = \begin{cases} \frac{1}{\pi r^2}, & (x_1, x_2) \in Q_r, \\ 0, & (x_1, x_2) \notin Q_r. \end{cases}$$

Мы видели, что

$$p_{\xi_1}(x_1) = \frac{2}{\pi r^2}\sqrt{r^2 - x_1^2},\ p_{\xi_2}(x_2) = \frac{2}{\pi r^2}\sqrt{r^2 - x_2^2}.$$

Тогда для $(x_1, x_2) \in O_r$

$$p_{\xi_1,\xi_2}(x_1, x_2) \neq p_{\xi_1}(x_1)p_{\xi_2}(x_2),$$

то есть, ξ_1, ξ_2 являются зависимыми.

Вычислим $Cov(\xi_1, \xi_2)$.

$$Cov(\xi_1, \xi_2) = \mathbb{E}[(\xi_1 - \mathbb{E}\xi_1)(\xi_2 - \mathbb{E}\xi_2)] = \mathbb{E}(\xi_1\xi_2) =$$
$$= \iint_{O_r} \frac{1}{\pi r^2} x_1 x_2 \, dx_1 dx_2 = \frac{1}{\pi r^2}\int_{-r}^{r} x_1 dx_1 \int_{-\sqrt{r^2-x_1^2}}^{\sqrt{r^2-x_1^2}} x_2 dx_2 = 0.$$

Следовательно, $\rho(\xi_1, \xi_2) = 0$, то есть, случайные величины ξ_1, ξ_2 некоррелированы.

Пусть далее $(\xi_1, \xi_2, ..., \xi_n)$ — случайный вектор. Обозначим

$$\rho_{ij} = \rho(\xi_i, \xi_j) = \frac{Cov(\xi_i, \xi_j)}{\sqrt{\mathbb{D}\xi_i \mathbb{D}\xi_j}},\ (i, j = 1, ..., n).$$

Матрица $R = (\rho_{ij})$ называется корреляционной матрицей случайного вектора $(\xi_1, \xi_2, ..., \xi_n)$. Ясно, что матрица R также симметрична и неотрицательно определена.

Рассмотрим далее положительно определенную симметричную матрицу A. Тогда её определитель

$$|A| = \det A > 0,$$

следовательно, определена обратная матрица $A^{-1} = (a^*_{ij})$.

Определение 2.21. *Функция*

$$p_n(x_1, ...x_n) = \frac{|A|^{-\frac{1}{2}}}{(2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2}\sum a^*_{ij}(x_i - m_i)(x_j - m_j)},$$

где $m_i \in \mathbb{R}$, $(i = 1, ..., n)$, *называется функцией плотности* n*-мерного нормального распределения с вектором средних значений* $\overline{m} = (m_1, ..., m_n)$ *и ковариационной матрицей* A.

В векторной форме её можно записать в следующей виде: для $\overline{x} = (x_1, ..., x_n)$

$$p_n(\overline{x}) = \frac{|A|^{-\frac{1}{2}}}{(2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2}(\overline{x} - \overline{m})^* A^{-1}(\overline{x} - \overline{m})}.$$

В частном случае, при $n = 2$ функция плотности нормального распределения приводится к следующему виду:

$$p_{\xi_1,\xi_2}(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\{-\frac{1}{2(1-\rho^2)}[\frac{(x_1 - m_1)^2}{\sigma_1^2} - \\ -2\rho\frac{(x_1 - m_1)(x_2 - m_2)}{\sigma_1\sigma_2} + \frac{(x_2 - m_2)^2}{\sigma_2^2}]\}.$$

Простой подсчёт раскрывает смысл всех входящих в последнее равенство параметров:

$$m_1 = E\xi_1,\ m_2 = E\xi_2,$$

$$\sigma_1^2 = D\xi_1,\ \sigma_2^2 = D\xi_2,\ \rho = \rho(\xi_1, \xi_2).$$

Пусть $\rho = \rho(\xi_1, \xi_2) = 0$. В этом случае функция плотности $p_{\xi_1,\xi_2}(x_1, x_2)$ случайного вектора (ξ_1, ξ_2) будет равна

$$p_{\xi_1,\xi_2}(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2} e^{-\frac{(x_1 - m_1)^2}{2\sigma_1^2}} e^{-\frac{(x_2 - m_2)^2}{2\sigma_2^2}}.$$

Тогда

$$p_{\xi_1}(x_1) = \int_{-\infty}^{+\infty} p_{\xi_1,\xi_2}(x_1, x_2)dx_2 = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(x_1 - m_1)^2}{2\sigma_1^2}},$$

$$p_{\xi_2}(x_2) = \int_{-\infty}^{+\infty} p_{\xi_1,\xi_2}(x_1,x_2)dx_1 = \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{(x_2-m_2)^2}{2\sigma_2^2}},$$

то есть,

$$p_{\xi_1\xi_2}(x_1,x_2) = p_{\xi_1}(x_1)p_{\xi_2}(x_2).$$

Значит, случайные величины ξ_1, ξ_2 независимы.

Итак, мы доказали следующую теорему.

Теорема 2.9. *Случайные величины, имеющие совместное нормальное распределение, независимы тогда и только тогда, когда они некоррелированы.*

2.4.3 Полиномиальное распределение. Сходимость к многомерному нормальному распределению

Обобщая биномиальную схему, предположим, что пространство элементарных исходов Ω имеет следующую структуру:

$$\Omega = \{\omega : \omega = \{a_1, \dots, a_n\}, a_i = b_j,\ i = 1, \dots n,\ j = 1, \dots, r\}$$

где $b_1, \dots, b_r$ — заданные числа. Пусть $\nu_j(\omega)$ — число элементов в последовательности $\omega = (a_1, \dots, a_n)$ равных b_j.

Положим

$$p(\omega) = p_1^{\nu_1(\omega)} p_2^{\nu_2(\omega)} \cdots p_r^{\nu_r(\omega)},$$

где $p_j \geqslant 0,\ (j = 1, \dots, r)$, и $\sum_{j=1}^r p_j = 1$.

Заметим, что при этом

$$\sum_{\omega\in\Omega} p(\omega) = \sum_{\{n_1+\cdots+n_r=1\}} C_n(n_1, \dots, n_r) p_1^{n_1} \cdots p_r^{n_r},$$

где $C_n(n_1, \dots, n_r)$ — число таких упорядоченных последовательностей $(a_1, \dots, a_r)$, где элемент b_1 встречается n_1 раз, … , элемент b_r встречается n_r раз. Легко видеть, что

$$C_n(n_1, \dots, n_r) = C_n^{n_1} C_{n-n_1}^{n_2} \cdots C_{n-(n_1+n_2+\cdots+n_{r-1})}^{n_r} =$$

$$= \frac{n!}{n_1!(n-n_1)!} \cdot \frac{(n-n_1)!}{n_2!(n-n_1-n_2)!} \cdots 1 = \frac{n!}{n_1! \cdots n_r!}.$$

Значит,

$$\sum_{\omega\in\Omega} p(\omega) = \sum_{\{n_j\geqslant 0,\, n_1+\cdots+n_r=1\}} \frac{n!}{n_1! \cdots n_r!} p_1^{n_1} \cdots p_r^{n_r} =$$

$$= (p_1 + \cdots + p_r)^n = 1.$$

Последнее равенство — это полиномиальная формула Ньютона.

Пусть $A_{n_1,\dots,n_r} = \{\omega : \nu_1(\omega) = n_1, \dots, \nu_r(\omega) = n_r\}$. Тогда

$$\mathbf{P}(A_{n_1,\dots,n_r}) = \frac{n!}{n_1!\cdots n_r!} p_1^{n_1}\cdots p_r^{n_r}.$$

Набор вероятностей $(\mathbf{P}(A_{n_1,\dots,n_r}))$ называется полиномиальным распределением.

Интерпретировать это распределение можно следующим образом. Пусть производится n независимых однотипных опытов, в любом из которых может произойти событие A_1 с вероятностью p_1, $\dots$, A_r — с вероятностью p_r, такие, что $p_1 + \dots + p_r = 1$. Тогда $A_{n_1,\dots,n_r}$ — это событие, состоящее в том, что событие A_1 произошло n_1 раз, $\dots$, A_r произошло n_r раз.

Результат опыта есть $r-$мерный случайный вектор $\xi = (\xi_1, \xi_2, \dots, \xi_r)$, где $\xi_j \sim B(n, p_j),\ (j = 1, \dots, r)$. Это легко видеть, если результат отдельного испытания представить в виде: $(0, \dots, 0, 1, 0, \dots, 0)$, где на j-том месте стоит единица (произошло событие A_j, $j = 1, \dots, r$).

Тогда

$$\mathbb{E}\xi_j = np_j, (j = 1, \dots, r),\ \ \mathbb{D}\xi_j = np_jq_j,$$

где $q_j = 1 - p_j, (j = 1, \dots, r)$.

Посчитаем ковариационную матрицу случайного вектора ξ. Для $i \neq j$

$$Cov(\xi_i, \xi_j) = \mathbb{E}\xi_i\xi_j - \mathbb{E}\xi_i\mathbb{E}\xi_j.$$

Учитывая, что $\xi_j = \sum_{k=1}^{n} \xi_{jk}\,(j = 1, \dots, r)$, (здесь k — это номер испытания), ξ_{jk} — бернулевские случайные величины с параметром p_j, имеем

$$Cov(\xi_i, \xi_j) = \mathbb{E}\left(\sum_{k=1}^{n}\xi_{ik}\right)\left(\sum_{l=1}^{n}\xi_{jl}\right) - n^2p_ip_j = \sum_{k,l}\mathbb{E}\xi_{ik}\xi_{jl} - n^2p_ip_j =$$
$$= \sum_{k=l}\mathbb{E}\xi_{ik}\xi_{jl} + \sum_{k\neq l}\mathbb{E}\xi_{ik}\xi_{jl} - n^2p_ip_j.$$

Поскольку при k-ом испытании только одна из случайных величин $\xi_{1k}, \dots, \xi_{rk}$ равна единице, то $\xi_{jk}\xi_{jk} = 0$, и, значит, $\mathbb{E}\xi_{ik}\xi_{jk} = 0$. Тогда и первое слагаемое в последнем равенстве равно нулю.

Итак,

$$Cov(\xi_i, \xi_j) = n(n-1)p_ip_j - n^2p_ip_j = -np_ip_j, i \neq j.$$

Знак «минус» здесь естественным образом следует из условия связи: $\xi_1+...+\xi_r=n$. Тогда коэффициент корреляции для случайных величин ξ_i,ξ_j равен

$$\rho_{ij}=\rho(\xi_i,\xi_j)=\frac{-np_ip_j}{\sqrt{np_iq_inp_jq_j}}=-\sqrt{\frac{p_ip_j}{q_iq_j}}.$$

Матрица

$$R=(\rho_{ij})$$

называется корреляционной матрицей полиномиального распределения.

Сформулируем теорему типа теоремы Муавра-Лапласа для полиномиального распределения.

Теорема 2.10. *Пусть $\xi=(\xi_1,\dots,\xi_{r+1})$ — $(r+1)$-мерный случайный вектор, имеющий полиномиальное распределение с параметром $(r+1,n,\bar{p})$, где $\bar{p}=(p_1,\dots,p_r)$. Тогда при $n\to\infty$ для любых $a_1,b_1,a_2,b_2,\dots\dots,a_r,b_r$*

$$\lim_{n\to\infty}\mathbf{P}\left\{a_1\leqslant\frac{\xi_1-np_1}{\sqrt{np_1q_1}}\leqslant b_1,\dots,a_r\leqslant\frac{\xi_r-np_r}{\sqrt{np_rq_r}}\leqslant b_r\right\}=$$

$$=\frac{|R|^{-\frac{1}{2}}}{(2\pi)^{\frac{r}{2}}}\int_{a_1}^{b_1}\cdots\int_{a_r}^{b_r}e^{-\frac{1}{2}x^*R^{-1}x}dx_1,\dots,dx_r,$$

где $x=(x_1,\dots,x_r)$, x^* — вектор-столбец, $R=(\rho_{ij})$ — корреляционная матрица определенная выше.

Вернемся к двумерному случаю нормального распределения. Рассмотрим эллипсы

$$\frac{1}{2(1-\rho^2)}\left[\frac{(x_1-m_1)^2}{\sigma_1^2}-2\rho\frac{(x_1-m_1)(x_2-m_2)}{\sigma_1\sigma_2}+\frac{(x_2-m_2)^2}{\sigma_2^2}\right]=C^2.$$

Известно, что они суть проекции сечений поверхности $z=\rho_{\xi_1\xi_2}(x_1,x_2)$ плоскостями, параллельными плоскости X_1OX_2. Они характеризуют поверхность $z=\rho_{\xi_1\xi_2}(x_1,x_2)$ достаточно хорошо: если $\rho=0$ и $\sigma_1=\sigma_2$, то это окружности; при $\rho\to\pm1$ они «вытягиваются» вдоль координат.

В случае $=2$ рассматриваемый эллипс представляет из себя так называемый эллипс рассеяния: вероятность попадания значений (ξ_1,ξ_2) внутрь этого эллипса очень большая $\approx 0,98$. По его виду можно судить о характере зависимости между случайными величинами ξ_1,ξ_2.

Двумерное нормальное распределение играет важную роль в теории стрельб: распределение координаты (x_1,x_2) попадания в цель при стрельбе из закрепленного ствола хорошо согласуется с двумерным нормальным распределением.

2.5 Законы больших чисел Чебышёва и Бернулли

Пусть $(\Omega, \mathscr{F}, \mathbf{P})$ — некоторое вероятностное пространство, $\xi_1, \xi_2 \ldots \xi_n, \ldots$ — произвольные случайные величины.

Теорема 2.11. *Если случайные величины $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ попарно независимы и $\lim\limits_{n\to\infty} \frac{1}{n^2}\sum\limits_{k=1}^{n} \mathbb{D}\xi_k = 0$, то для любого $\varepsilon > 0$*

$$\lim_{n\to\infty} \mathbf{P}\left\{\left|\frac{\xi_1 + \cdots + \xi_n}{n} - \frac{\mathbb{E}\xi_1 + \cdots + \mathbb{E}\xi_n}{n}\right| < \varepsilon\right\} = 1.$$

Доказательство. Обозначим $\eta_n = \frac{\xi_1 + \cdots + \xi_n}{n}$. Тогда утверждение теоремы можно записать так:

$$\lim_{n\to\infty} \mathbf{P}\left\{|\eta_n - \mathbb{E}\eta_n| \geqslant \varepsilon\right\} = 0.$$

Так как $\xi_1, \xi_2 \ldots \xi_n$ попарно независимы, то $\mathbb{D}\eta_n = \frac{1}{n^2}\sum\limits_{k=1}^{n} \mathbb{D}\xi_k$.

По неравенству Чебышёва имеем:

$$\mathbf{P}\left\{|\eta_n - \mathbb{E}\eta_n| \geqslant \varepsilon\right\} \leqslant \frac{\mathbb{D}\eta_n}{\varepsilon^2} = \frac{1}{n^2\varepsilon^2}\sum_{k=1}^{n} \mathbb{D}\xi_k \to 0, (n \to \infty).$$

□

Теорема 2.12. *Если случайные величины $\xi_1, \xi_2 \ldots \xi_n, \ldots$ попарно независимы и $\mathbb{D}\xi_k \leqslant C, (k = 1, 2 \ldots)$, где C — некоторая константа, то для любого $\varepsilon > 0$*

$$\lim_{n\to\infty} \mathbf{P}\left\{\left|\frac{\xi_1 + \cdots + \xi_n}{n} - \frac{\mathbb{E}\xi_1 + \cdots + \mathbb{E}\xi_n}{n}\right| < \varepsilon\right\} = 1.$$

Доказательство. Очевидно следует из теоремы 2.11. □

Теорема 2.13 (ЗБЧ Бернулли)**.** *Пусть μ_n — число «успехов» в n независимых испытаниях Бернулли, p — вероятность «успеха» в каждом отдельном испытании. Тогда для любого $\varepsilon > 0$*

$$\lim_{n\to\infty} \mathbf{P}\left\{\left|\frac{\mu_n}{n} - p\right| < \varepsilon\right\} = 1.$$

Доказательство. Представим случайную величину $\mu_n = \xi_1 + \cdots + \xi_n$, где ξ_i — независимые одинаково распределенные случайные величины с распределением Бернулли: $\mathbf{P}\{\xi_i = 1\} = p$, $\mathbf{P}\{\xi_i = 0\} = q$, $(i = 1, \ldots, n)$.

Очевидно, $\mathbb{E}\xi_i = p$ и $\mathbb{D}\xi_i = pq < 1$. Тогда утверждение теоремы следует из теоремы 2.12. □

Приведём пример, показывающий, что в некоторых задачах, связанных с планированием эксперимента, доказанные выше теоремы играют некоторую роль.

Пример 2.10 (Оценка ошибки приближённого значения неизвестной величины при неоднократных измерениях)**.** *Пусть производятся n независимых измерений некоторой неизвестной величины a. Ошибки измерения $\delta_1, \delta_2, \ldots, \delta_n$ будем считать независимыми случайными величинами и положим для всех i: $\mathbb{E}\delta_i = 0$ (отсутствие систематической ошибки) и $\mathbb{D}\delta_i = b^2$ (точность прибора), $i = 1, \ldots, n$.*

За значение ошибки измеримой величины a естественно взять случайную величину $\eta_n = \frac{\delta_1 + \cdots + \delta_n}{n}$. При этом $\mathbb{E}\eta_n = 0$, $\mathbb{D}\eta_n = \frac{1}{n^2}\sum\limits_{i=1}^{n}\mathbb{D}\delta_i = \frac{b^2}{n}$.

Предположим, что нам необходимо, чтобы ошибка измерения не превосходила величины Δ с достаточно большой вероятностью, например,

$$\mathbf{P}\left\{|\eta_n| < \Delta\right\} > 0,99,$$

или, что то же самое,

$$\mathbf{P}\left\{|\eta_n| \geqslant \Delta\right\} \leqslant 0,01.$$

Каково число независимых измерений n, чтобы удовлетворить этому условию?

Из неравенства Чебышёва

$$\mathbf{P}\left\{|\eta_n| \geqslant \Delta\right\} \leqslant \frac{b^2}{n\Delta^2}.$$

Значит, наше условие будет возможно, если $\frac{b^2}{n\Delta^2} \leqslant 0,01$. Тогда

$$n \geqslant 100\frac{b^2}{\Delta^2}.$$

Таким образом, мы оценили число измерений, необходимых для получения нужной точности.

Нетрудно видеть, что полученная оценка является достаточно грубой, она основана только на неравенстве Чебышёва. Несомненно, есть более точные оценки.

2.6 Сходимость последовательностей

Пусть $(\Omega, \mathscr{F}, \mathbf{P})$ — вероятностное пространство, $\xi, \xi_1, \xi_2, \ldots, \xi_n, \ldots$ — случайные величины.

Определение 2.22. *Последовательность случайных величин $\xi_1, \xi_2, \ldots$ сходится к случайной величине ξ по вероятности, если для любого $\varepsilon > 0$*

$$\lim_{n\to\infty} \mathbf{P}\left\{|\xi_n - \xi| < \varepsilon\right\} = 1.$$

Этот вид сходимости нам уже встречался в законах больших чисел. Будем обозначать эту сходимость так:

$$\xi_n \xrightarrow{\mathbf{P}} \xi.$$

Определение 2.23. *Последовательность случайных величин $\xi_1, \xi_2, \ldots$ сходится почти наверное к случайной величине ξ (или с вероятностью единица), если*

$$\mathbf{P}\{\omega : \xi_n(\omega) \not\to \xi(\omega), n \to \infty\} = 0.$$

Эту сходимость будем обозначать

$$\xi_n \xrightarrow{\text{п.н.}} \xi.$$

Предложение 2.3. *Сходимость почти наверное влечёт сходимость по вероятности.*

Доказательство. Пусть последовательность (ξ_n) сходится к ξ почти наверное. Значит, для любого $\varepsilon > 0$, найдётся такое $N = N(\omega)$, что имеет место неравенство

$$\sup_{n\geqslant N} |\xi_n - \xi| < \varepsilon.$$

Тогда:

$$\bigcup_N \bigcap_{n\geqslant N} \left\{|\xi_n - \xi| < \varepsilon\right\} = \Omega - \mathscr{N},$$

где $\mathbf{P}(\mathscr{N}) = 0$. Следовательно,

$$\mathbf{P}(\mathscr{N}) = \mathbf{P}\left(\bigcap_N \bigcup_{n\geqslant N} \left\{|\xi_n - \xi| \geqslant \varepsilon\right\}\right) = 0.$$

Это означает, что

$$\lim_{N\to\infty} \mathbf{P}\left(\bigcup_{n\geqslant N}\{|\xi_n-\xi|\geqslant\varepsilon\}\right)=0.$$

Отсюда и

$$\lim_{n\to\infty}\mathbf{P}\{|\xi_n-\xi|\geqslant\varepsilon\}=0,$$

то есть, последовательность (ξ_n) сходится к ξ по вероятности. □

Итак, сходимость почти наверное влечёт сходимость по вероятности. Обратное утверждение неверно. Это показывает следующий пример.

Пример 2.11. *Пусть* $\Omega=[0,1), \mathscr{F}=\mathscr{B}([0,1))$, $\mathbf{P}$ *— мера Лебега на* $\mathscr{B}([0,1))$. *Определим для всех натуральных* n *случайные величины* $\xi_n^{(k)}$, $k=1,\dots,n$, *на полуинтервале* $[0,1)$ *следующим образом:*

$$\xi_n^{(k)}(\omega)=\begin{cases}1, & \frac{k-1}{n}\leqslant\omega<\frac{k}{n},\\ 0, & \text{для всех остальных } \omega\in[0,1).\end{cases}$$

Перенумеруем эти случайные величины следующим образом:

$$\xi_1^{(1)},\xi_2^{(1)},\xi_2^{(2)},\xi_3^{(1)},\xi_3^{(2)},\xi_3^{(3)},\dots,$$

и обозначим полученную последовательность $\xi_1',\xi_2',\dots,\xi_n',\dots$.

Для любого $\varepsilon>0$

$$\lim_{n\to\infty}\mathbf{P}\left\{\left|\xi_n'-0\right|\geqslant\varepsilon\right\}=\lim_{n\to\infty}\frac{1}{n(k)}=0.$$

Но в то же время последовательность (ξ_n') *не сходится к нулю ни в одной точке.*

Действительно, возьмём произвольное $\omega\in(0,1]$. *Зададим* $\varepsilon=\frac{1}{2}$. *Тогда для любого* N' *существует такое* $n'\geqslant N'$, *что* $\omega\in\left[\frac{k-1}{n'},\frac{k}{n'}\right)$, *то есть,* $\xi_{n'}'(\omega)=1$. *Значит,* $|\xi_{n'}'(\omega)|=1>\frac{1}{2}$.

Таким образом последовательность (ξ_n') *не сходится к нулю в точке* $\omega\in[0,1)$. *В силу произвольности* ω *она не сходится к нулю ни в одной точке* $\omega\in[0,1)$.

Определение 2.24. *Пусть* $\xi_1,\xi_2,\dots,\xi_n,\dots$ *— последовательность случайных величин,* $F_{\xi_1},F_{\xi_2},\dots,F_{\xi_n},\dots$ *— соответствующая последовательность функций распределения. Последовательность случайных величин* $\xi_1,\xi_2,\dots,\xi_n,\dots$ *сходится к случайной величине* ξ *по распределению, если* $F_{\xi_n}(x)\to F_\xi(x),(n\to\infty)$ *в каждой точке непрерывности функции* $F_\xi(x)$.

Сходимость последовательности функций распределения $\{F_{\xi_n}(x)\}$ при этом называется слабой. Обозначается эта сходимость так: для случайных величин — $\xi_n \xrightarrow{d} \xi$, для соответствующих функций распределения — $F_{\xi_n} \xrightarrow{w} F_\xi$.

Теорема 2.14 (Критерий слабой сходимости, Хелли)**.** *Последовательность случайных величин $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ сходится к случайной величине ξ по распределению тогда и только тогда, если для любой непрерывной ограниченной функции $g = g(x), x \in \mathbb{R}$,*

$$\lim_{n\to\infty} \int\limits_{-\infty}^{+\infty} g(x) dF_{\xi_n}(x) = \int\limits_{-\infty}^{+\infty} g(x) dF_\xi(x).$$

или, что то же самое,

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

Доказательство. Пусть выполнено условие теоремы, и x — точка непрерывности $F_\xi(x)$. Это значит, что для любого $\varepsilon > 0$ существует такое $\delta = \delta(\varepsilon, x)$, что для всех $y \in U_\delta(x)$ (то есть, $|y - x| < \delta$)

$$|F_\xi(y) - F_\xi(x)| < \varepsilon.$$

Покажем, что в этой точке $F_{\xi_n}(x) \to F_\xi(x)$ $(n \to \infty)$. Возьмём две непрерывные ограниченные функции, полагая,

$$f_1(t) = \begin{cases} 1, & t \leqslant x, \\ -\frac{t-x}{\delta} + 1, & x < t \leqslant x + \delta, \\ 0, & t > x + \delta. \end{cases}$$

и

$$f_2(t) = \begin{cases} 1, & t \leqslant x - \delta, \\ -\frac{t-x}{\delta}, & x - \delta < t \leqslant x, \\ 0, & t > x, \end{cases}$$

При этом $f_1(t) \leqslant I_{(-\infty;x+\delta)}(t)$, $f_2(t) \geqslant I_{(-\infty;x-\delta)}(t)$.

Рассмотрим

$$F_{\xi_n}(x) = \int\limits_{-\infty}^{x} dF_{\xi_n}(t) = \int\limits_{-\infty}^{+\infty} I_{(-\infty;x)}(t) dF_{\xi_n}(t) \leqslant \int\limits_{-\infty}^{+\infty} f_1(t) dF_{\xi_n}(t).$$

Из сходимости интегралов от непрерывных ограниченных функций следует, что

$$\int\limits_{-\infty}^{+\infty} f_1(t)dF_{\xi_n}(t) \to \int\limits_{-\infty}^{+\infty} f_1(t)dF_\xi(t) \leqslant \int\limits_{-\infty}^{+\infty} I_{(-\infty;x+\delta)}(t)dF_\xi(t) = \\ = \int\limits_{-\infty}^{x+\delta} dF_\xi(t) = F_\xi(x+\delta) < F_\xi(x) + \varepsilon.$$

Значит, $F_{\xi_n}(x) < F_\xi(x) + \varepsilon$. С другой стороны

$$F_{\xi_n}(x) = \int\limits_{-\infty}^{x} dF_{\xi_n}(t) = \int\limits_{-\infty}^{+\infty} I_{(-\infty;x)}(t)dF_{\xi_n}(t) \geqslant \\ \geqslant \int\limits_{-\infty}^{+\infty} f_2(t)dF_{\xi_n}(t) \to \int\limits_{-\infty}^{+\infty} f_2(t)dF_\xi(t) \geqslant \int\limits_{-\infty}^{+\infty} I_{(-\infty,x-\delta)}(t)dF_\xi(t) = \\ = \int\limits_{-\infty}^{x-\delta} dF_\xi(t) = F_\xi(x-\delta) > F_\xi(x) - \varepsilon,$$

то есть, $F_{\xi_n}(x) > F_\xi(x) - \varepsilon$. Объединяя два полученных неравенства, получим $|F_{\xi_n}(x) - F_\xi(x)| < \varepsilon$. Это значит, что $F_{\xi_n}(x) \to F_\xi(x)$ в точке x.

Обратно, пусть $F_n(x) \to F(x)$ в точках непрерывности функции $F(x)$. Возьмём произвольную непрерывную ограниченную функцию $f = f(x)$. Вначале докажем равенство

$$\lim_{n\to\infty} \int\limits_a^b f(x)dF_n(x) = \int\limits_a^b f(x)dF(x),$$

где a и b — точки непрерывности функции $F(x)$.

Так как функция $f(x)$ непрерывна на отрезке $[a,b]$, то она равномерно непрерывна на этом отрезке. Значит, отрезок $[a,b]$ можно разбить точками

$$a = x_0, x_1, \ldots, x_{N-1}, x_N = b,$$

так, что $|f(x) - f(x_k)| < \varepsilon$ для всех $x \in [x_k, x_{k+1}]$ для всех k и любого наперёд заданного $\varepsilon > 0$. Кроме того, эти точки можно выбрать так, чтобы они были точками непрерывности функции $F(x)$.

Обозначим $M = \max\limits_{a \leqslant x \leqslant b} |f(x)|$. Рассмотрим

$$\left| \int\limits_a^b f(x)dF(x) - \int\limits_a^b f(x)dF_n(x) \right| \leqslant \left| \int\limits_a^b f(x)dF(x) - \int\limits_a^b f(x_k)dF(x) \right| +$$
$$+ \left| \int\limits_a^b f(x_k)dF(x) - \int\limits_a^b f(x_k)dF_n(x) \right| + \left| \int\limits_a^b f(x_k)dF_n(x) - \int\limits_a^b f(x)dF_n(x) \right|.$$

Оценим каждое из слагаемых в последнем неравенстве. Очевидно, что первое слагаемое оценивается

$$\left| \int\limits_a^b f(x)dF(x) - \int\limits_a^b f(x_k)dF(x) \right| \leqslant \int\limits_a^b |f(x) - f(x_k)|dF(x) <$$
$$< \frac{\varepsilon}{8} \int\limits_a^b dF(x) = \frac{\varepsilon}{8}(F(b) - F(a)) < \frac{\varepsilon}{8}.$$

Третье слагаемое оценивается точно также:

$$\left| \int\limits_a^b f(x_k)dF_n(x) - \int\limits_a^b f(x)dF_n(x) \right| < \frac{\varepsilon}{8}(F_n(b) - F_n(a)) < \frac{\varepsilon}{8}.$$

Оценим второе слагаемое.

$$\left| \int\limits_a^b f(x_k)dF(x) - \int\limits_a^b f(x_k)dF_n(x) \right| =$$
$$= \left| \sum_{k=0}^{N-1} f(x_k)[F(x_{k+1}) - F(x_k)] - \sum_{k=0}^{N-1} f(x_k)[F_n(x_{k+1}) - F_n(x_k)] \right| =$$
$$= \left| \sum_{k=0}^{N-1} f(x_k)[F(x_{k+1}) - F_n(x_{k+1})] - \sum_{k=0}^{N-1} f(x_k)[F(x_k) - F_n(x_k)] \right|.$$

Так как $F_n(x) \xrightarrow{w} F(x)$ в точках непрерывности функции $F(x)$, то можно выбрать достаточно большое n таким образом, чтобы для всех k

$$|F(x_k) - F_n(x_k)| < \frac{\varepsilon}{4MN}.$$

Тогда

$$\left|\int\limits_a^b f(x_k)dF(x) - \int\limits_a^b f(x_k)dF_n(x)\right| < \frac{\varepsilon}{4}.$$

Значит,

$$\left|\int\limits_a^b f(x)dF(x) - \int\limits_a^b f(x)dF_n(x)\right| < \frac{\varepsilon}{2}.$$

Оценим теперь разность интегралов на всей числовой прямой.

$$\left|\int\limits_{-\infty}^{+\infty} f(x)dF(x) - \int\limits_{-\infty}^{+\infty} f(x)dF_n(x)\right| \leqslant I_1 + I_2 + I_3,$$

где

$$I_1 = \left|\int\limits_{-\infty}^{a} f(x)dF(x) - \int\limits_{-\infty}^{a} f(x)dF_n(x)\right|,$$

$$I_2 = \left|\int\limits_a^b f(x)dF(x) - \int\limits_a^a f(x)dF_n(x)\right|,$$

$$I_3 = \left|\int\limits_b^{+\infty} f(x)dF(x) - \int\limits_b^{+\infty} f(x)dF_n(x)\right|.$$

Очевидно, что $I_1 \leqslant M|F(a) - F_n(a)|$. Так как $\lim\limits_{a\to-\infty} F(a) = 0$ и $\lim\limits_{n\to\infty} F_n(a) = F(a)$, то можно подобрать a и n такими, чтобы $I_1 < \frac{\varepsilon}{4}$.

Точно также $I_3\,|M[1 - F(b)] - M[1 - F_n(b)|$ и, подбирая, n и b, можно сделать $I_3 < \frac{\varepsilon}{4}$.

Мы уже показали, что $I_2 < \frac{\varepsilon}{2}$.

Таким образом, при достаточно больших n

$$\left|\int\limits_{-\infty}^{+\infty} f(x)dF(x) - \int\limits_{-\infty}^{+\infty} f(x)dF_n(x)\right| < \varepsilon.$$

□

Предложение 2.4. *Пусть* $\xi_n \xrightarrow{\mathbf{P}} \xi$. *Тогда* $\xi_n \xrightarrow{d} \xi$.

Пусть $f(x)$ — непрерывная ограниченная функция, $|f(x)| \leqslant$.
Зададим произвольное $\varepsilon > 0$ и выберем N так, чтобы

$$\mathbf{P}\{|\xi| > N\} \leqslant \frac{\varepsilon}{8C}.$$

Выберем также $\delta > 0$ таким, что для всех $x, |x| \leqslant N$, и $|x-y|$

$$|f(x) - f(y)| < \frac{\varepsilon}{4}.$$

Это всегда можно сделать, поскольку функция $f(x)$ равномерно непрерывна на отрезке $[-N, N]$. Тогда

$$\int\limits_{-\infty}^{+\infty} f(x)dF_n(x) - \int\limits_{-\infty}^{+\infty} f(x)dF(x) = \int\limits_{\{|\xi_n-\xi|\leqslant\delta;|\xi|\leqslant N\}} f(x)d[F_n(x) - F(x)] +$$
$$+ \int\limits_{\{|\xi_n-\xi|\leqslant\delta;|\xi|>N\}} f(x)d[F_n(x) - F(x)] + \int\limits_{\{|\xi_n-\xi|>\delta\}} f(x)d[F_n(x) - F(x)].$$

Или в терминах математических ожиданий ожиданий:

$$\mathbb{E}(|f(\xi_n) - f(\xi)| \leqslant \mathbb{E}(|f(\xi_n) - f(\xi)|; |\xi_n - \xi| \leqslant \delta, |\xi| \leqslant N)+$$
$$+\mathbb{E}(|f(\xi_n) - f(\xi)|; |\xi_n - \xi| \leqslant \delta, |\xi| > N) + \mathbb{E}(|f(\xi_n) - f(\xi)|; |\xi_n - \xi| > \delta) <$$
$$< \frac{\varepsilon}{4} + \frac{\varepsilon}{4} + 2C\mathbf{P}\{|\xi_n - \xi| > \delta\} = \varepsilon + 2C\mathbf{P}\{|\xi_n - \xi| > \delta\}.$$

Но, из сходимости по вероятности случайных величин можно подобрать такое n_0, что для всех $n > n_0$ $\mathbf{P}\{|\xi_n - \xi| > \delta\} < \frac{\varepsilon}{2}$. Значит, для достаточно больших n

$$\mathbb{E}|f(\xi_n) - f(\xi)| < \varepsilon,$$

то есть, $\xi_n \xrightarrow{d}$, $(n \to \infty)$.

Обратное утверждение неверно. Это показывает следующий пример.

Пример 2.12. *Пусть* $\Omega = [0,1]$, $\mathscr{F} = \mathscr{B}([0,1])$, $\mathbf{P}$ *— мера Лебега на* $\mathscr{B}([0,1])$.

Положим

$$\xi_n(\omega) = \begin{cases} (-1)^n, & \omega \in [0, \frac{1}{2}], \\ (-1)^{n-1}, & \omega \in (\frac{1}{2}, 1]. \end{cases}$$

Тогда все случайные величины имеют одно и то же распределение:

$$\mathbf{P}\{\xi_n = 1\} = \mathbf{P}\{\xi_n = -1\} = \frac{1}{2}.$$

Значит, соответствующие функции распределения при $n \to \infty$ сходятся к той же функции распределения.

Но случайные величины ξ_n не сходятся к случайной величине ξ по вероятности, какая бы ни была случайная величина ξ, поскольку последовательность $\{(-1)^n\}$ не имеет предела.

Теорема 2.15. *Если последовательность случайных величин (ξ_n) сходится к вырожденной случайной величине по распределению, то она сходится к ней по вероятности, то есть*

$$\xi_n \xrightarrow{d} C \Rightarrow \xi_n \xrightarrow{\mathbf{P}} C.$$

Доказательство. Рассмотрим функцию распределения вырожденной случайной величины $\xi = \xi(\omega)$:

$$F(x) = \begin{cases} 0, & x \leqslant C, \\ 1, & x > C. \end{cases}$$

Она имеет единственную точку разрыва $x = C$. В каждой точке $x \neq C$ по определению имеем: $F_n(x) \to F(x)$, $(n \to \infty)$, где $F_n(x)$ — функция распределения случайной величины ξ_n, $(n = 1, 2, \dots)$. Зададим произвольное $\varepsilon > 0$ и рассмотрим

$$\mathbf{P}\{|\xi_n - C| < \varepsilon\} = \mathbf{P}\{-\varepsilon < \xi_n - C < \varepsilon\} = \mathbf{P}\{C - \varepsilon < \xi_n < C + \varepsilon\}.$$

Поскольку точки $C - \varepsilon$ и $C + \varepsilon$ — точки непрерывности функции $F(x)$, то

$$\mathbf{P}\{|\xi_n - C| < \varepsilon\} = F_n(C + \varepsilon) - F_n(C - \varepsilon) \longrightarrow F(C + \varepsilon) - F(C - \varepsilon) = 1.$$

□

Теорема 2.16. *Если последовательность функций распределения $F_n(x)$ сходится слабо к функции распределения $F(x)$, и функция $F(x)$ непрерывна всюду, то последовательность $F_n(x)$ сходится равномерно к функции $F(x)$:*

$$F_n(x) \xrightarrow{w} F(x) \Rightarrow F_n(x) \rightrightarrows F(x), \ (n \to \infty), \ \ x \in \mathbb{R}.$$

Доказательство. Поскольку функция $F(x)$ монотонна и ограничена, то для любого $\varepsilon > 0$ мы можем подобрать конечное число точек $x_1, x_2, \dots, x_N$ так, что на каждом интервале

$$(-\infty, x_1), \ [x_1, x_2), \ [x_2, x_3), \ \dots, \ [x_N, +\infty)$$

приращение функции $F(x)$ будет меньше $\varepsilon/5$. Пусть $x \in [x_k, x_{k+1})$. Рассмотрим

$$|F_n(x) - F(x)| = |(F_n(x) - F_n(x_k)) + (F_n(x_k) - F(x_k)) + (F(x_k) - F(x))| \leqslant \\ \leqslant |F_n(x) - F_n(x_k)| + |F_n(x_k) - F(x_k)| + |F(x_k) - F(x)|.$$

Оценим каждое из слагаемых в последнем неравенстве.

$$|F_n(x) - F_n(x_k)| \leqslant |F_n(x_{k+1}) - F_n(x_k)| \leqslant \\ \leqslant |F_n(x_{k+1}) - F(x_{k+1})| + |F(x_k) - F_n(x_k)| + |F(x_{k+1}) - F(x_k)|$$

Ясно, что

$$|F(x_k) - F(x)| \leqslant |F(x_{k+1}) - F(x_k)|.$$

Отсюда при $n > n_0(k)$

$$|F_n(x) - F(x)| \leqslant \\ |F_n(x_{k+1}) - F(x_{k+1})| + 2|F_n(x_k) - F(x_k)| + 2|F(x_{k+1}) - F(x_k)| < \varepsilon.$$

Положим

$$n_0 = \max_{0 \leqslant k \leqslant N} n_0(k).$$

Тогда при $n > n_0$

$$|F_n(x) - F(x)| < \varepsilon.$$

□

2.7 Характеристические функции

Введем предварительно понятие комплекснозначной случайной величины. Пусть ξ_1, ξ_2 — некоторые случайные величины. Тогда

$$\xi(\omega) = \xi_1(\omega) + i\xi_2(\omega)$$

называется комплекснозначной случайной величиной. При этом случайная величина ξ_1 называется вещественной частью ξ, а случайная величина ξ_2 — мнимой частью ξ. Естественно положить

$$\mathbb{E}\xi = \mathbb{E}\xi_1 + i\mathbb{E}\xi_2.$$

Определение 2.25. *Характеристической функцией вещественнозначной случайной величины ξ называется комплекснозначная (вообще говоря) функция*

$$\varphi_\xi(t) = \mathbb{E}e^{it\xi} = \int\limits_{-\infty}^{+\infty} e^{itx} dF_\xi(x), \;\; t \in \mathbb{R}.$$

Если функция распределения имеет плотность $p_\xi(x)$, то характеристическая функция равна

$$\varphi_\xi(t) = \int_{-\infty}^{+\infty} e^{itx} p_\xi(x) dx$$

и называется преобразованием Фурье функции $p_\xi(x)$.

2.7.1 Свойства характеристических функций

1. Для любой случайной величины ξ

$$|\varphi_\xi(t)| \leqslant 1, \;\; \varphi_\xi(0) = 1.$$

2. Если $\eta = a\xi + b$, то

$$\varphi_\eta(t) = e^{itb}\varphi_\xi(ta).$$

3. Если $\xi_1, \; \xi_2$ — независимые случайные величины, то

$$\varphi_{\xi_1+\xi_2}(t) = \varphi_{\xi_1}(t)\varphi_{\xi_2}(t).$$

4. Характеристическая функция $\varphi_\xi(t)$ равномерно непрерывна на $\mathbb{R}$.
5. Если существует k-ый абсолютный момент случайной величины ξ ($\mathbb{E}|\xi^k| < \infty$), $k \geqslant 1$, то существует непрерывная k-ая производная функции $\varphi_\xi(t)$ и

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

6. $\overline{\varphi}_\xi(t) = \varphi_\xi(-t) = \varphi_{-\xi}(t)$.

Доказательство. Докажем приведённые свойства.

1. Заметим сперва, что

$$|\varphi_\xi(t)| = \left| \int\limits_{-\infty}^{+\infty} e^{itx} dF_\xi(x) \right| \leqslant \int\limits_{-\infty}^{+\infty} |e^{itx}| dF_\xi(x) \leqslant \int\limits_{-\infty}^{+\infty} dF_\xi(x) = 1,$$

так как $|e^{itx}| \leqslant 1$. Ясно, что $\varphi_\xi(0) = 1$.

2. Действительно,

$$\varphi_{a\xi+b}(t) = \mathbb{E}e^{it(a\xi+b)} = \mathbb{E}e^{ita\xi}e^{itb} = e^{itb}\mathbb{E}e^{i(ta)\xi} = e^{itb}\varphi_\xi(at).$$

3. Это свойство следует из свойств математического ожидания:

$$\varphi_{\xi_1+\xi_2}(t) = \mathbb{E}e^{it(\xi_1+\xi_2)} = \mathbb{E}e^{it\xi_1}e^{it\xi_2} = \mathbb{E}e^{it\xi_1}\mathbb{E}e^{it\xi_2} = \varphi_{\xi_1}(t)\varphi_{\xi_2}(t).$$

Свойство распространяется на любое конечное число слагаемых.

4. Возьмём произвольную точку $t \in \mathbb{R}$ и зададим произвольное $\varepsilon > 0$. Рассмотрим

$$|\varphi_\xi(t+h) - \varphi_\xi(t)| = \left| \int\limits_{-\infty}^{\infty} (e^{i(t+h)x} - e^{itx}) dF_\xi(x) \right| =$$

$$= \left| \int\limits_{-\infty}^{\infty} e^{itx}(e^{ihx} - 1) dF_\xi(x) \right| \leqslant \int |e^{ihx} - 1| dF_\xi(x) =$$

$$= \int\limits_{|x|<N} |e^{ihx} - 1| dF_\xi(x) + \int\limits_{|x|\geqslant N} |e^{ihx} - 1| dF_\xi(x) = I_1 + I_2.$$

Выберем N так, чтобы $I_2 < \varepsilon/2$, затем выберем h так, чтобы $I_1 < \varepsilon/2$. Тогда для для всех $t \in \mathbb{R}$

$$|\varphi_\xi(t+h) - \varphi_\xi(t)| < \varepsilon.$$

5. Так как
$$|\int ixe^{itx}dF_\xi(x)| \leqslant \int |x|dF_\xi(x) = \mathbb{E}|\xi| < \infty,$$
то интеграл $\int ixe^{itx}dF_\xi(x)$ сходится равномерно по t. Поэтому можно дифференцировать под знаком интеграла:
$$\varphi'_\xi(t) = i\int xe^{itx}dF_\xi(x),\ \varphi'(0) = i\mathbb{E}\xi.$$
Дальнейшее доказательство идёт по индукции
$$\varphi_\xi^{(k)}(t) = i^k \int x^k e^{itx}dF_\xi(x), \text{ и } \varphi_\xi^{(k)}(0) = i^k\mathbb{E}\xi^k.$$

6. Докажем последнее свойство.

$$\overline{\varphi}_\xi(t) = \overline{\mathbb{E}(e^{it\xi})} = \mathbb{E}(\overline{e^{it\xi}}) = \mathbb{E}\overline{(\cos(t\xi) + i\sin(t\xi))} = \mathbb{E}(\cos(t\xi) - i\sin(t\xi)) = \\ = \mathbb{E}(\cos(-t\xi) + i\sin(-t\xi)) = \mathbb{E}e^{-it\xi} = \varphi_\xi(-t) = \varphi_{-\xi}(t).$$

Отсюда следует, что если случайная величина ξ имеет симметричное распределение, то её характеристическая функция вещественна. □

2.7.2 Примеры вычисления характеристических функций

1. Вырожденное распределение:
$$\mathbf{P}\{\xi = a\} = 1,\ a - \text{const}.$$
Тогда
$$\varphi_\xi(t) = \mathbb{E}e^{it\xi} = e^{ita}.$$

2. Распределение Бернулли с параметром $p > 0$:
$$\mathbf{P}\{\xi = 1\} = p,\ \mathbf{P}\{\xi = 0\} = 1 - p.$$
В этом случае
$$\varphi_\xi(t) = \mathbb{E}(e^{it\xi}) = pe^{it} + (1-p)e^{it} = pe^{it} + (1-p) = p(e^{it} - 1) + 1.$$

3. Биномиальное распределение с параметрами n и p:

Здесь для вычисления характеристической функции используем свойство 3 характеристических функций и представление
$$\xi = \xi_1 + \xi_2 + \cdots + \xi_n,$$

где ξ_i, $(i = 1, \dots, n)$, независимы и имеют распределение Бернулли с параметром p.
Тогда

$$\varphi_\xi(t) = \prod_{i=1}^{n} \varphi_{\xi_i}(t) = (1 + p(e^{it} - 1))^n.$$

4. Распределение Пуассона с параметром $\lambda > 0$:

$$\mathbf{P}\{\xi = k\} = \frac{\lambda^k}{k!} e^{-\lambda},\ k = 0, 1, 2, \dots$$

Тогда

$$\varphi_\xi(t) = \mathbb{E} e^{it\xi} = \sum_{k=0}^{\infty} e^{itk} \frac{e^{-\lambda}\lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda e^{it})^k}{k!} = e^{-\lambda} e^{\lambda e^{it}} = \exp[\lambda(e^{it} - 1)].$$

5. Равномерное распределение на отрезке $[0, 1]$:

$$p_\xi(x) = \begin{cases} 1, & \text{x} \in [0, 1]; \\ 0, & \text{x} \notin [0, 1]. \end{cases}$$

Здесь

$$\varphi_\xi(t) = \int_0^1 e^{itx} dx = \frac{e^{itx}}{it}\Big|_0^1 = \frac{e^{it} - 1}{it}.$$

6. Нормальное распределение $N(0, 1)$:

$$p_\xi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2},\ x \in \mathbb{R}.$$

Тогда

$$\varphi_\xi(t) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{+\infty} e^{itx} e^{-x^2/2} dx = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{+\infty} \cos(tx) e^{-x^2/2} dx +$$

$$+ \frac{1}{\sqrt{2\pi}} i \int\limits_{-\infty}^{+\infty} \sin(tx) e^{-x^2/2} dx = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{+\infty} \cos(tx) e^{-x^2/2} dx.$$

Итак,

$$\varphi_\xi(t) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{+\infty} \cos(tx) e^{-x^2/2} dx.$$

Так как последний интеграл сходится равномерно по t, то его можно дифференцировать под знаком интеграла:

$$\varphi'_\xi(t) = -\frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{+\infty} x\sin(tx)e^{-x^2/2}dx = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{+\infty} \sin(tx)d(e^{-x^2/2}) =$$

$$= \frac{1}{\sqrt{2\pi}} \sin(tx)e^{-x^2/2}\Big|_{-\infty}^{+\infty} - \frac{1}{\sqrt{2\pi}} t \int\limits_{-\infty}^{+\infty} \cos(tx)e^{-x^2/2}dx = -t\varphi_\xi(t).$$

Таким образом, для решения задачи мы получили дифференциальное уравнение:

$$\varphi'_\xi(t) = -t\varphi_\xi(t).$$

Решая последнее уравнение, получаем $\varphi_\xi(t) = Ce^{-t^2/2}$. Из условия $\varphi_\xi(0) = 1$ получим, что $C = 1$. Тогда

$$\varphi_\xi(t) = e^{-t^2/2}.$$

2.7.3 Теоремы о характеристических функциях

Теорема 2.17 (Формула обращения П. Леви)**.** *Пусть $F(x)$ — функция распределения случайной величины ξ, $\varphi(t)$ — её характеристическая функция. Тогда для любых точек непрерывности x, y функции $F(x)$ выполняется равенство*

$$F(y) - F(x) = \frac{1}{2\pi} \lim_{A\to\infty} \int\limits_{-A}^{A} \frac{e^{-itx} - e^{-ity}}{it} \varphi(t)dt.$$

Доказательство. Для $y > x$ рассмотрим

$$I_A = \frac{1}{2\pi}\int\limits_{-A}^{A} \frac{e^{-itx} - e^{-ity}}{it}\varphi(t)dt =$$

$$= \frac{1}{2\pi}\int\limits_{-A}^{A}\Big[\int\limits_{-\infty}^{+\infty} \frac{e^{-itx} - e^{-ity}}{it} e^{itu} dF(u)\Big]dt =$$

$$= \frac{1}{2\pi}\int\limits_{-A}^{A}\Big[\int\limits_{-\infty}^{+\infty} \frac{e^{it(u-x)} - e^{it(u-y)}}{it} dF(u)\Big]dt =$$

$$= \frac{1}{2\pi}\int\limits_{-\infty}^{+\infty}\Big[\int\limits_{-A}^{A} \frac{e^{it(u-x)} - e^{it(u-y)}}{it} dt\Big]dF(u).$$

Внутренний интеграл разобьём на две части и произведём во втором интеграле замену $t \to -t$. Перегруппировывая слагаемые, получим:

$$I_A = \frac{1}{2\pi}\int\limits_{-\infty}^{+\infty}\Big[\int\limits_{0}^{A} \frac{e^{it(u-x)} - e^{-it(u-x)} - e^{it(u-y)} + e^{-it(u-y)}}{it} dt\Big]dF(u) =$$

$$= \frac{1}{\pi}\int\limits_{-\infty}^{+\infty}\Big[\int\limits_{0}^{A} \frac{\sin(t(u-x)) - \sin(t(u-y))}{t} dt\Big]dF(u) =$$

$$= \frac{1}{\pi}\int\limits_{-\infty}^{+\infty}\Big[\int\limits_{A(u-y)}^{A(u-x)} \frac{\sin t}{t} dt\Big]dF(u).$$

Поскольку

$$\lim_{A,B\to\infty} \frac{1}{\pi}\int\limits_{-B}^{A} \frac{\sin t}{t} dt = 1,$$

и интеграл под знаком предела равномерно ограничен по A и B, следовательно, равномерно ограничен и интеграл

$$S_A(u) = \frac{1}{\pi}\int\limits_{A(u-y)}^{A(u-x)} \frac{\sin t}{t} dt.$$

Значит,

$$\lim_{A\to\infty} I_A = \lim_{A\to\infty} \int\limits_{-\infty}^{+\infty} S_A(u)dF(u) = \int\limits_{-\infty}^{+\infty} \lim_{A\to\infty} S_A(u)dF(u).$$

Представим

$$I_A = \int\limits_{-\infty}^{x} S_A(u)dF(u) + \int\limits_{x}^{y} S_A(u)dF(u) + \int\limits_{y}^{+\infty} S_A(u)dF(u).$$

Заметим сразу, что для $u < x$ и $u > y$:

$$\lim_{A\to\infty} S_A(u) = 0,$$

то есть первый и третий интегралы сходятся к нулю при $A \to \infty$.

Далее

$$\lim_{A\to\infty} S_A(u) = \begin{cases} 0, & u < x, u > y; \\ 1, & u \in (x, y); \\ \frac{1}{2}, & u = x, \quad u = y. \end{cases}$$

Так как x и y — точки непрерывности $F(x)$, то

$$\int\limits_{x}^{y} \lim_{A\to\infty} S_A(u)dF(u) = \int\limits_{x}^{y} dF(u) = F(y) - F(x).$$

□

Следствие 2.4. *Характеристическая функция случайной величины однозначно определяет её функцию распределения.*

Теорема 2.18 (Теорема непрерывности). *Пусть $(F_n(x))_{n=1}^{\infty}$ — последовательность функций распределения, $(\varphi_n(t))_{n=1}^{\infty}$ — последовательность соответствующих характеристических функций. Тогда*
1) Если $F_n(x) \xrightarrow[n\to\infty]{w} F(x)$ и $F(x)$ есть функция распределения, то $\varphi_n(t) \to \varphi(t)$ $(n \to \infty)$, причём $\varphi(t)$ — характеристическая функция распределения $F(x)$.
2) Если $\varphi_n(t) \to \varphi(t)$ $(n \to \infty)$ для всех t и $\varphi(t)$ непрерывна в точке $t = 0$, то $\varphi(t)$ есть характеристическая функция некоторого распределения $F(x)$, причём $F_n(x) \xrightarrow[n\to\infty]{w} F(x)$.

Прежде чем доказывать эту теорему, докажем два вспомогательных утверждения.

Лемма 2.1. *Для того, чтобы последовательность функций распределения $F_n(x)$ слабо сходилась к неубывающей функции $F(x)$, достаточно, чтобы $F_n(x) \to F(x)$ при $n \to \infty$ на каком-нибудь всюду плотном множестве D вещественных чисел.*

Доказательство. Пусть x — точка непрерывности функции $F(x)$. Для любых $x', x'' \in D: \ x' \leqslant x \leqslant x''$ выполнено $F_n(x') \leqslant F_n(x) \leqslant F_n(x'')$. Следовательно,

$$\lim_{n\to\infty} F_n(x') \leqslant \varliminf_{n\to\infty} F_n(x) \leqslant \varlimsup_{n\to\infty} F_n(x) \leqslant \lim_{n\to\infty} F_n(x'').$$

Значит,

$$F(x') \leqslant \varliminf_{n\to\infty} F_n(x) \leqslant \varlimsup_{n\to\infty} F_n(x) \leqslant F(x'').$$

Полагая теперь $x' \uparrow x$ и $x'' \downarrow x$ по множеству D, и учитывая, что x — точка непрерывности $F(x)$, получим:

$$\lim_{n\to\infty} F_n(x) = F(x).$$

□

Лемма 2.2. *Произвольная последовательность функций распределения $(F_n(x))_{n=1}^{\infty}$ содержит подпоследовательность $\{F_{nn}(x)\}$, слабо сходящуюся к некоторой неубывающей функции $F(x)$.*

Доказательство. Пусть $D = \{x'_n\}$ — произвольное счётное, всюду плотное множество вещественных чисел. Возьмём точку x'_1 и рассмотрим последовательность $\{F_n(x'_1)\}$. Она ограничена, а, значит, содержит сходящуюся подпоследовательность $\{F_{1n}(x'_1)\}$ и

$$F_{1n}(x'_1) \xrightarrow[n\to\infty]{} F(x'_1).$$

Рассмотрим далее подпоследовательность $\{F_{1n}(x'_2)\}$. Точно также, как и ранее,

$$F_{2n}(x'_2) \to F(x'_2).$$

При этом $\lim\limits_{n\to\infty} F_{2n}(x'_1) = F(x'_1)$.

Продолжая этот процесс, получим k подпоследовательностей

$$\{F_{1n}(x'_1)\}, \{F_{2n}(x'_2)\}, \ldots, \{F_{kn}(x'_k)\},$$

таких, что $\lim\limits_{n\to\infty} F_{kn}(x'_i) = F(x'_i), (i = \overline{1,k})$.

Рассмотрим диагональную последовательность $\{F_{nn}(x'_k)\}$. Тогда для любых $x'_k \in D$ только $k-1$ первых членов последовательности $\{F_{nk}(x'_k)\}$ могут не принадлежать последовательности $\{F_{nn}(x'_k)\}$. Поэтому

$$\lim_{n\to\infty} F_{nn}(x'_k) = F(x'_k).$$

Причём очевидно, что $F(x)$ задана на множестве D и не убывает. Продолжим её до неубывающей функции на всей числовой прямой. Теперь мы находимся в условиях доказанной предыдущей леммы. □

Перейдём к доказательству теоремы непрерывности.

Доказательство. Часть 1) следует из критерия слабой сходимости функций распределения для функций $Re(e^{itx})$ и $Im(e^{itx})$, которые ограничены и непрерывны.
Докажем часть 2). Из леммы 2.2 сразу следует, что последовательность $\{F_n(x)\}_{n=1}^{\infty}$ содержит такую подпоследовательность $\{F_{nk}(x)\}_{k=1}^{\infty}$, что

$$F_{nk}(x) \xrightarrow[k\to\infty]{w} F(x).$$

Покажем, что $F(x)$ является функцией распределения.

Она не убывает и ограничена. Предположим противное, то есть, что $F(x)$ не является функцией распределения. Положим

$$\delta = VarF(x) = \sup_x F(x) - \inf_x F(x) < 1.$$

Возьмём $\varepsilon < 1 - \delta$. Заметим, что $\varphi(t)$ интегрируема на любом конечном отрезке $[-\tau, \tau]$ как предел интегрируемых функций. Так как $\varphi(0) = \lim\limits_{n\to\infty} \varphi_n(0) = 1$, то в силу непрерывности функции φ для заданного $\varepsilon > 0$ существует $\tau > 0$:

$$\frac{1}{2\tau}\left|\int_{-\tau}^{\tau} \varphi(t)dt\right| > 1 - \frac{\varepsilon}{2} > \delta + \frac{\varepsilon}{2}. \qquad (*)$$

Оценим этот же интеграл сверху. Имеем

$$\int_{-\tau}^{\tau} \varphi_{n_k}(t)dt = \int_{-\infty}^{+\infty}\int_{-\tau}^{\tau} e^{itx}dtdF_{n_k}(x) =$$

$$= \int_{\{|x|<A\}}\int_{-\tau}^{\tau} e^{itx}dtdF_{n_k}(x) + \int_{\{|x|\geqslant A\}}\int_{-\tau}^{\tau} e^{itx}dtdF_{n_k}(x).$$

Тогда

$$\left|\int\limits_{-\tau}^{\tau}\varphi_{n_k}(t)dt\right| \leqslant \left|\int\limits_{\{|x|<A\}}\int\limits_{-\tau}^{\tau}e^{itx}dtdF_{n_k}(x)\right| + \left|\int\limits_{\{|x|\geqslant A\}}\int\limits_{-\tau}^{\tau}e^{itx}dtdF_{n_k}(x)\right|.$$

Для достаточно больших k имеем:

$$\delta_k = F_{n_k}(A) - F_{n_k}(-A) < \delta + \frac{\varepsilon}{4}.$$

Так как $|\int\limits_{-\tau}^{\tau} e^{itx}dt| \leqslant 2\tau$, то первое слагаемое оценивается числом $2\tau\delta_k$. Далее

$$\left|\int\limits_{-\tau}^{\tau}e^{itx}dt\right| = \left|\frac{e^{itx}}{ix}\right|\Big|_{-\tau}^{\tau} = \left|\frac{e^{i\tau x} - e^{-i\tau x}}{ix}\right| \leqslant \frac{2}{x}.$$

Значит, второе слагаемое оценивается сверху числом $\frac{2}{A}$. Поэтому выбирая $A > \frac{4}{\tau\varepsilon}$, получаем

$$\frac{1}{2\tau}\left|\int\limits_{-\tau}^{\tau}\varphi_{n_k}(t)dt\right| < \delta + \frac{\varepsilon}{4} + \frac{\varepsilon}{4} = \delta + \frac{\varepsilon}{2}.$$

Переходя к пределу при $n \to \infty$, получим

$$\frac{1}{2\tau}\left|\int\limits_{-\tau}^{\tau}\varphi(t)dt\right| \leqslant \delta + \frac{\varepsilon}{2}.$$

Но это противоречит $(*)$. Значит, $F(x)$ есть функция распределения. Всё остальное следует из части 1). $\square$

2.8 Сходимость рядов

Будем предполагать, что $\xi_1, \xi_2, \dots$ — последовательность независимых случайных величин, $S_n = \xi_1 + \ldots + \xi_n$ и A — множество тех элементарных исходов ω, где ряд $\sum \xi_n(\omega)$ сходится к конечному пределу. Цель настоящего параграфа — дать критерии, позволяющие определять, сходится или расходится ряд из независимых случайных величин.

Теорема 2.19 (Колмогоров, Хинчин). *а) Пусть $\mathbb{E}\xi_n = 0, n \geqslant 1$. Тогда, если*

$$\sum \mathbb{E}\xi_n^2 < \infty, \tag{2.1}$$

то ряд $\sum \xi_n$ сходится с вероятностью единица.

b) Если к тому же случайные величины $\xi_n, n \geqslant 1$, равномерно ограничены ($\mathbf{P}\,|\xi_n| \leqslant c = 1$ для некоторого $\leqslant c < \infty$), то верно и обратное: из сходимости ряда $\sum \xi_n$ с вероятностью единица следует условие 2.1.

Доказательство этой теоремы существенно опирается на следующие неравенства.

Неравенства Колмогорова

а) Пусть $\xi_1, \xi_2, \dots, \xi_n$ — независимые случайные величины с $\mathbb{E}\xi_i = 0$, $\mathbb{E}\xi_i^2 < \infty$, $1 \leqslant i \leqslant n$. Тогда для всякого $\varepsilon > 0$

$$\mathbf{P}\left\{\max_{1\leqslant k\leqslant n} |S_k| \geqslant \varepsilon\right\} \leqslant \frac{\mathbb{E}S_n^2}{\varepsilon^2}$$

b) Если к тому же $\mathbf{P}\left\{|\xi_i| \leqslant c\right\} = 1, 1 \leqslant i \leqslant n$, то

$$\mathbf{P}\left\{\max_{1\leqslant k\leqslant n} |S_k| \geqslant \varepsilon\right\} \geqslant 1 - \frac{(c+\varepsilon)^2}{\mathbb{E}S_n^2}$$

Доказательство. а) Обозначим

$$A = \left\{\max_{1\leqslant k\leqslant n} |S_k| \geqslant \varepsilon\right\}, \; A_k = \left\{|S_i| < \varepsilon, i = 1, \dots, k-1, |S_k| \geqslant \varepsilon\right\}, \quad 1 \leqslant k \leqslant n.$$

Тогда $A = \sum_{k=1}^{\infty} A_k$ и

$$\mathbb{E}S_n^2 \geqslant \mathbb{E}S_n^2 I_A = \sum_{k=1}^{\infty} \mathbb{E}S_n^2 I_{A_k}$$

Но

$$\mathbb{E}S_n^2 I_{A_k} = \mathbb{E}\left(S_k + (\xi_{k+1} + \ldots + \xi_n)\right)^2 I_{A_k} =$$
$$= \mathbb{E}S_k^2 I_{A_k} + 2\mathbb{E}S_k\left(\xi_{k+1} + \ldots + \xi_n\right) I_{A_k} + \mathbb{E}\left(\xi_{k+1} + \ldots + \xi_n\right)^2 I_{A_k} \geqslant \mathbb{E}S_k^2 I_{A_k}$$

поскольку

$$\mathbb{E}S_k\left(\xi_{k+1} + \ldots + \xi_n\right) I_{A_k} = \mathbb{E}S_k I_{A_k} \cdot \mathbb{E}\left(\xi_{k+1} + \ldots + \xi_n\right) = 0$$

в силу предположенной независимости и условий $\mathbb{E}\xi_i = 0, 1 \leqslant i \leqslant n$. Поэтому

$$\mathbb{E}S_n^2 \geqslant \sum \mathbb{E}S_k^2 I_{A_k} \geqslant \varepsilon^2 \sum \mathbf{P}\left(A_k\right) = \varepsilon^2 \mathbf{P}(A)$$

что и доказывает первое неравенство. Для доказательства второго неравенства заметим, что

$$\mathbb{E}S_n^2 I_A = \mathbb{E}S_n^2 - \mathbb{E}S_n^2 I_{\bar{A}} \geqslant \mathbb{E}S_n^2 - \varepsilon^2 \mathbf{P}(\bar{A}) = \mathrm{E}S_n^2 - \varepsilon^2 + \varepsilon^2 \mathbf{P}(A)$$

С другой стороны, на множестве A_k

$$|S_{k-1}| \leqslant \varepsilon, \quad |S_k| \leqslant |S_{k-1}| + |\xi_k| \leqslant \varepsilon + c$$

и, значит,

$$\mathbb{E}S_n^2 I_A = \sum_k \mathbb{E}S_k^2 I_{A_k} + \sum_k \mathbb{E}\left(I_{A_k}\left(S_n - S_k\right)^2\right) \leqslant$$
$$\leqslant (\varepsilon + c)^2 \sum_k \mathbf{P}\left(A_k\right) + \sum_{k=1}^{n} \mathbf{P}\left(A_k\right) \sum_{j=k+1}^{n} \mathbb{E}\xi_j^2 \leqslant$$
$$\leqslant \mathbf{P}(A)\left[(\varepsilon + c)^2 + \sum_{j=1}^{n} \mathbb{E}\xi_j^2\right] = \mathbf{P}(A)\left[(\varepsilon + c)^2 + \mathbb{E}S_n^2\right]$$

Из полученных неравенств находим, что

$$\mathbf{P}(A) \geqslant \frac{\mathbb{E}S_n^2 - \varepsilon^2}{(\varepsilon + c)^2 + \mathbb{E}S_n^2 - \varepsilon^2} = 1 - \frac{(\varepsilon + c)^2}{(\varepsilon + c)^2 + \mathbb{E}S_n^2 - \varepsilon^2} \geqslant 1 - \frac{(\varepsilon + c)^2}{\mathbb{E}S_n^2}$$

Неравенство b) доказано. □

Доказательство. Докажем теперь теорему.

a) Сходимость последовательности случайных величин почти наверное эквивалентна фундаментальности этой последовательности почти наверное (оставим это утверждение без доказательства, оно достаточно просто). Но последовательность $(S_n)_{n\geqslant 1}$ фундаментальна ($\mathbf{P}$-п. н.) в том и только том случае, когда

$$\mathbf{P}\left\{\sup_{k\geqslant 1}|S_{n+k}-S_n|\geqslant\varepsilon\right\}\to 0,\quad n\to\infty.$$

В силу неравенства Колмогорова a)

$$\mathbf{P}\left\{\sup_{k\geqslant 1}|S_{n+k}-S_n|\geqslant\varepsilon\right\}=\lim_{N\to\infty}\mathbf{P}\{\max_{1\leqslant k\leqslant N}\mid S_{n+k}-S_n\mid\geqslant\varepsilon\}\leqslant$$
$$\leqslant\lim_{N\to\infty}\frac{\sum_{k=n}^{n+N}\mathbb{E}\xi_k^2}{\varepsilon^2}=\frac{\sum_{k=n}^{\infty}\mathbb{E}\xi_k^2}{\varepsilon^2}\to 0,\quad n\to\infty.$$

Поэтому ряд $\sum\xi_k$ сходится с вероятностью единица.

b) Пусть ряд $\sum\xi_k$ сходится. Тогда в силу условия фундаментальности для достаточно больших n

$$\mathbf{P}\left\{\sup_{k\geqslant 1}|S_{n+k}-S_n|\geqslant\varepsilon\right\}<\frac{1}{2}. \tag{2.2}$$

В силу неравенства Колмогорова b)

$$\mathbf{P}\left\{\sup_{k\geqslant 1}|S_{n+k}-S_n|\geqslant\varepsilon\right\}\geqslant 1-\frac{(c+\varepsilon)^2}{\sum_{k=n}^{\infty}\mathrm{E}\xi_k^2}.$$

Поэтому, если допустить, что $\sum_{k=1}^{\infty}\mathrm{E}\xi_k^2=\infty$, то получим

$$\mathbf{P}\left\{\sup_{k\geqslant 1}|S_{n+k}-S_n|\geqslant\varepsilon\right\}=1$$

что противоречит неравенству 2.2 для последовательности (S_n). □

Пример 2.13. *Если $\xi_1,\xi_2,\ldots$ — последовательность независимых бернуллиевских случайных величин с $\mathbf{P}\{\xi_n=+1\}=\mathbf{P}\{\xi_n=-1\}=1/2$, то ряд $\sum\xi_n a_n$, где $|a_n|\leqslant c$, сходится с вероятностью единица тогда и только тогда. когда $\sum a_n^2<\infty$*

2.9 Предельные теоремы теории вероятностей

Здесь мы рассмотрим ряд предельных теорем в «классической» постановке, а также и некоторые их обобщения. При доказательстве этих теорем существенную роль играет аппарат характеристических функций.

Теорема 2.20 (Закон больших чисел Хинчина)**.** *Пусть* $\{\xi_n\}_{n=1}^{\infty}$ *— последовательность независимых одинаково распределённых случайных величин, имеющих первый момент:* $\mathbb{E}\xi_n = \mu,\ n = 1, 2, 3, \ldots$. *Обозначим* $S_n = \xi_1 + \cdots + \xi_n$. *Тогда*

$$\frac{S_n}{n} \xrightarrow{\mathbf{P}} \mu$$

при $n \to \infty$.

Доказательство. Так как все ξ_k одинаково распределены, то характеристические функции ξ_k одинаковы. Обозначим $\varphi_{\xi_k}(t) = \varphi(t)$. Поскольку случайные величины имеют первый момент, то

$$\varphi(t) = 1 + it\mu + o(t).$$

Тогда

$$\varphi_{S_n}(t) = [1 + it\mu + o(t)]^n.$$

Значит,

$$\varphi_{\frac{S_n}{n}}(t) = \left[\varphi\left(\frac{t}{n}\right)\right]^n = \left[1 + i\frac{t\mu}{n} + o\left(\frac{t}{n}\right)\right]^n.$$

При $n \to \infty$ величины $\frac{t}{n} \to 0$. Тогда

$$\lim_{n\to\infty} \varphi_{\frac{S_n}{n}}(t) = \lim_{n\to\infty}\left[1 + i\frac{t}{n}\mu\right]^n = e^{it\mu}.$$

Как мы видели выше, $e^{it\mu}$ есть характеристическая функция вырожденного распределения ($\mu - const$). Значит, по теореме непрерывности

$$\frac{S_n}{n} \xrightarrow{d} \mu,$$

при ($n \to \infty$), и из теоремы о слабой сходимости к константе, имеем

$$\frac{S_n}{n} \xrightarrow{P} \mu,\ (n \to \infty).$$

□

Теорема 2.21 (Центральная предельная теорема). *Пусть* $\{\xi_n\}_{n=1}^{\infty}$ — *последовательность независимых одинаково распределённых случайных величин, имеющих второй конечный момент. Обозначим* $\mu = \mathbb{E}\xi_k$, $\sigma^2 = \mathbb{D}\xi_k$ $(k = 1, 2, \dots)$. *Тогда*

$$\mathbf{P}\left\{\frac{S_n - n\mu}{\sigma\sqrt{n}} < x\right\} \to \Phi(x)$$

равномерно по $x \in \mathbb{R}$ *при* $n \to \infty$, *где*

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{-t^2/2} dt.$$

Доказательство. Обозначим

$$\eta_n = \frac{S_n - n\mu}{\sigma\sqrt{n}} = \sum_{k=1}^{n} \frac{\xi_k - \mu}{\sigma\sqrt{n}}.$$

Рассмотрим $\varphi_{\xi_k - \mu}(t)$. Так как существует второй момент, то

$$\varphi_{\xi_k - \mu}(t) = 1 + it\mathbb{E}(\xi_k - \mu) - \frac{t^2}{2}\mathbb{E}(\xi_k - \mu)^2 + o(t^2).$$

Поскольку $\mathbb{E}(\xi_k - \mu) = 0,\ \mathbb{E}(\xi_k - \mu)^2 = \sigma^2$, то

$$\varphi_{\xi_k - \mu}(t) = 1 - \frac{t^2}{2}\sigma^2 + o(t^2).$$

Значит,

$$\varphi_{\eta_n}(t) = \left[1 - \frac{t^2}{2n} + o\left(\frac{t^2}{\sigma^2 n}\right)\right]^n.$$

Тогда

$$\lim_{n\to\infty} \varphi_{\eta_n}(t) = \lim_{n\to\infty} \left[1 - \frac{t^2}{2n}\right]^n = e^{-t^2/2}$$

для всех t.

Как мы видели выше, $e^{-t^2/2}$ есть характеристическая функция нормального распределения $\mathscr{N}(0,1)$. Отсюда по теореме непрерывности имеем

$$F_{\eta_n}(x) \xrightarrow{w} \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{-u^2/2} du.$$

Далее, так как функция $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{-u^2/2} du$ непрерывна в каждой точке $x \in \mathbb{R}$, то сходимость эта равномерная. □

2.9.1 Усиленный закон больших чисел

Усиленным законом больших чисел называется утверждение, в котором в теореме 2.20 сходимость по вероятности заменяется на сходимость почти наверное. Приведём лишь некоторые результаты.

Теорема 2.22 (УЗБЧ Колмогорова). *Пусть $\xi_1, \xi_2, \ldots$ — последовательность независимых случайных величин с конечными вторыми моментами, положительные числа b_n таковы, что $b_n \uparrow \infty$ и*

$$\sum_{n=1}^{\infty} \frac{\mathbb{D}\xi_n}{b_n^2} < \infty. \tag{2.3}$$

Тогда

$$\frac{S_n - \mathbb{E}S_n}{b_n} \to 0 \ (\mathbf{P} - ..).$$

В частности, если

$$\sum_{n=1}^{\infty} \frac{\mathbb{D}\xi_n}{n^2} < \infty,$$

то

$$\frac{S_n - \mathbb{E}S_n}{n} \to 0 \ (\mathbf{P} - ..).$$

Доказательство. Для доказательства этой теоремы, а также некоторых других утверждений нам понадобятся два вспомогательные утверждения.

Лемма 2.3 (Тёплиц). *Пусть $(a_n)_{n\geqslant 1}$ — последовательность неотрицательных чисел, $b_n = \sum_{i=1}^{n} a_i, b_1 = a_1 > 0$ и $b_n \uparrow \infty, n \to \infty$. Пусть также $(x_n)_{n\geqslant 1}$ — последовательность чисел, сходящаяся к некоторому числу x. Тогда*

$$\frac{1}{b_n} \sum_{j=1}^{n} a_j x_j \to x$$

В частности, если $a_n = 1$, то

$$\frac{x_1 + \ldots + x_n}{n} \to x.$$

Доказательство. Пусть $\varepsilon > 0$ и $n_0 = n_0(\varepsilon)$ таково, что для всех $n \geqslant n_0$ $|x_n - x| \leqslant \varepsilon/2$. Выберем $n_1 > n_0$ так, что

$$\frac{1}{b_{n_1}} \sum_{j=1}^{n_0} a_j \left| x_j - x \right| < \frac{\varepsilon}{2}.$$

Тогда для $n > n_1$

$$\left|\frac{1}{b_n}\sum_{j=1}^{n} a_j x_j - x\right| \leqslant \frac{1}{b_n}\sum_{j=1}^{n} a_j\left|x_j - x\right| = \frac{1}{b_n}\sum_{j=1}^{n_0} a_j\left|x_j - x\right| +$$
$$+\frac{1}{b_n}\sum_{j=n_0+1}^{n} a_j\left|x_j - x\right| \leqslant \frac{1}{b_{n_1}}\sum_{j=1}^{n_0} a_j\left|x_j - x\right| + \frac{1}{b_n}\sum_{j=n_0+1}^{n} a_j\left|x_j - x\right| \leqslant$$
$$\leqslant \frac{\varepsilon}{2} + \frac{b_n - b_{n_0}}{b_n}\frac{\varepsilon}{2} \leqslant \varepsilon.$$

□

Лемма 2.4 (Кронекер)**.** *Пусть* $(b_n)_{n\geqslant 1}$ *— последовательность положительных возрастающих чисел,* $b_n \uparrow \infty, n \to \infty$ *и* $(x_n)_{n\geqslant 1}$ *— последовательность чисел таких, что ряд* $\sum x_n$ *сходится. Тогда*

$$\frac{1}{b_n}\sum_{j=1}^{n} b_j x_j \to 0, \quad n \to \infty.$$

В частности, если $b_n = n, x_n = y_n/n$ *и ряд* $\sum \frac{y_n}{n}$ *сходится, то*

$$\frac{y_1 + \ldots + y_n}{n} \to 0, \quad n \to \infty.$$

Доказательство. Пусть $b_0 = 0, S_0 = 0, S_n = \sum_{j=1}^{n} x_j$. Тогда («суммирование по частям»)

$$\sum_{j=1}^{n} b_j x_j = \sum_{j=1}^{n} b_j\left(S_j - S_{j-1}\right) = b_n S_n - b_0 S_0 - \sum_{j=1}^{n} S_{j-1}\left(b_j - b_{j-1}\right)$$

и, значит, полагая $a_j = b_j - b_{j-1}$, получим

$$\frac{1}{b_n}\sum_{j=1}^{n} b_j x_j = S_n - \frac{1}{b_n}\sum_{j=1}^{n} S_{j-1} a_j \to 0$$

так как если $S_n \to x$, то по лемме Тёплица

$$\frac{1}{b_n}\sum_{j=1}^{n} S_{j-1} a_j \to x.$$

□

Доказательство теоремы 2.22. Поскольку

$$\frac{S_n - \mathbb{E}S_n}{b_n} = \frac{1}{b_n}\sum_{k=1}^{n} b_k \left(\frac{\xi_k - \mathbb{E}\xi_k}{b_k}\right)$$

то в силу леммы Кронекера для доказательства теоремы достаточно, чтобы ($\mathbf{P}$-п. н.) сходился ряд $\sum \frac{\xi_k - \mathrm{E}\xi_k}{b_k}$. Но этот ряд действительно сходится в силу условия 2.3 и теоремы 2.20. □

2.9.2 Центральная предельная теорема Линдеберга

Пусть на вероятностном пространстве $(\Omega, \mathscr{F}, \mathbf{P})$ определена последовательность независимых в совокупности случайных величин $\xi_k = \xi_k(\omega),\ k = 1, 2, \ldots$, имеющих функции распределения $F_{\xi_k}(x)$, конечные математические ожидания $\mathbb{E}\xi_k = \mu_k$ и конечные дисперсии $\sigma_k^2 = \mathbb{D}\xi_k$.

Рассмотрим последовательность частичных сумм случайных величин:

$$S_n = \sum_{k=1}^{n} \xi_k, \quad n = 1, 2, \ldots.$$

Определим соответствующее математическое ожидание $A_n = \mathbb{E}S_n = \sum_{k=1}^{n} \mu_k$ и дисперсию $B_n^2 = \mathbb{D}S_n = \sum_{k=1}^{n} \sigma_k^2$.

Рассмотрим условие Линдеберга:

$$\frac{1}{B_n^2}\sum_{k=1}^{n} \int\limits_{\{|t-\mu_k|>\tau B_n\}} (t-\mu_k)^2\, dF_{\xi_k}(t) \to 0, n \to \infty, \tag{2.4}$$

для любого $\tau > 0$.

Введём следующие обозначения:

$$\xi_{kn} = \frac{\xi_k - \mu_k}{B_n}.$$

Тогда

$$\mathbb{E}\xi_{kn} = 0,\ \mathbb{D}\xi_{kn} = \frac{\sigma_k^2}{B_n^2},\ k = 1, \ldots, n.$$

Рассмотрим далее нормированные частичные суммы:

$$S_n^0 = \sum_{k=1}^{n} \xi_{kn} = \frac{S_n - A_n}{B_n}.$$

Имеем

$$\mathbb{E}S_n^0 = 0,\ \mathbb{D}S_n^0 = 1.$$

В формуле 2.4 сделаем замену переменных: $(t-\mu_k)/B_n = x$ и получим

$$\sum_{k=1}^{n} \int\limits_{\{|x|>\tau\}} x^2 dF_{\xi_{kn}}(x) = \sum_{k=1}^{n} \mathbb{E}\xi_{kn}^2 I_{\{|\xi_{kn}|>\tau\}} \to 0,\ n \to \infty$$

для всех $\tau > 0$.

Оказывается, что при увеличении n распределение $\mathscr{L}\left\{S_n^0\right\}$ приближается к гауссовскому распределению, хотя слагаемые имеют произвольное распределение вероятностей. Это явление называют «чудом Лапласа» Основным условием выполнения этого чуда является условие Линдеберга 2.4. Выясним его вероятностный смысл.

Лемма 2.5. *Пусть на вероятностном пространстве* $(\Omega, \mathscr{F}, \mathbf{P})$ *определена последовательность независимых в совокупности случайных величин* $\xi_k = \xi_k(\omega),\ k = 1, 2, \ldots,$ *имеющих конечные математические ожидания* $\mathbb{E}\xi_k = \mu_k$ *и конечные дисперсии* $\sigma_k^2 = \mathbb{D}\xi_k$*. Тогда если выполнено условие Линдеберга 2.4, то справедливы следующие условия:*
1. Дисперсии случайных величин $\xi_{kn} = \xi_{kn}(\omega)$ *равномерно малы, то есть*

$$\max_{1\leqslant k\leqslant n} \frac{\sigma_k^2}{B_n^2} = \max_{1\leqslant k\leqslant n} \mathbb{D}\xi_{kn} \to 0,\ (n \to \infty).$$

2. для случайных величин $\xi_{kn} = \xi_{kn}(\omega)$ *выполняется соотношение*

$$\forall \tau > 0,\ \mathbf{P}\left\{\max_{1\leqslant k\leqslant n} |\xi_{kn}| > \tau\right\} \to 0,\ (n \to \infty).$$

Доказательство. 1) По построению $\{\xi_{kn}\}$ имеем

$$\max_{1\leqslant k\leqslant n} \mathbb{D}\xi_{kn} = \max_{1\leqslant k\leqslant n} \mathbb{E}\xi_{kn}^2 = \max_{1\leqslant k\leqslant n} \left(\mathbb{E}\xi_{kn}^2 I_{\{|\xi_{kn}|\leqslant\tau\}} + \mathbb{E}\xi_{kn}^2 I_{\{|\xi_{kn}|>\tau\}}\right) \leqslant$$
$$\leqslant \tau^2 + \max_{1\leqslant k\leqslant n} \mathbb{E}\xi_{kn}^2 I_{\{|\xi_{kn}|>\tau\}} \to \tau^2, (n \to \infty).$$

Отсюда в силу произвольности τ (может быть выбрано сколь угодно малым) получаем первое условие.

2) По свойствам вероятности имеем

$$\mathbf{P}\left\{\max_{1\leqslant k\leqslant n}|\xi_{kn}|>\tau\right\}=\mathbf{P}\left(\bigcup_{k=1}^{n}\{|\xi_{kn}|>\tau\}\right)\leqslant$$

$$\leqslant\sum_{k=1}^{n}\mathbf{P}\left\{|\xi_{kn}|>\tau\right\}=\sum_{k=1}^{n}\int_{|x|>\tau}dF_{\xi_{kn}}(x)\ \leqslant$$

$$\leqslant\frac{1}{\tau^2}\sum_{k=1}^{n}\mathbb{E}\xi_{kn}^2 I_{\{|\xi_{kn}|>\tau\}}\to 0,(n\to\infty)\ \forall\tau>0.$$

□

В дальнейшем нам пригодятся неравенства, приводимые в виде следующей леммы.

Лемма 2.6. *Для любых $\alpha\in\mathbb{R},\beta\in\mathbb{C},|\beta|\leqslant 1/2$ справедливы неравенства:*

$$|e^{i\alpha}-1|\leqslant|\alpha|;\ \ |e^{i\alpha}-1-i\alpha|\leqslant\tfrac{\alpha^2}{2};$$
$$|e^{i\alpha}-1-i\alpha+\tfrac{\alpha^2}{2}|\leqslant\tfrac{|\alpha|^3}{6};\ \ |\ln(1+\beta)-\beta|\leqslant|\beta|^2.$$

Доказательство. Проводится непосредственно с использованием формулы Тейлора и представлением остаточного члена в форме Лагранжа. □

Теорема 2.23 (ЦПТ Линдеберга)**.** *Пусть на вероятностном пространстве $(\Omega,\mathscr{F},\mathbf{P})$ определена последовательность независимых в совокупности случайных величин $\xi_k=\xi_k(\omega),\ k=1,2,\ldots,$ имеющих конечные математические ожидания $\mathbb{E}\xi_k=\mu_k$ и конечные дисперсии $\sigma_k^2=\mathbb{D}\xi_k$. Тогда если выполнено условие Линдеберга 2.4, то*

$$\forall x\in\mathbb{R}\quad F_{S_n^0}(x)=\mathbf{P}\left\{S_n^0<x\right\}\to\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{u^2}{2}}du,\ n\to\infty.$$

Замечание 2.6. *Последняя сходимость эквивалентна слабой сходимости $F_{S_n^0}(\cdot)\xrightarrow{w}\Phi(\cdot)$, эквивалентной, в свою очередь, сходимости по распределению $S_n^0\xrightarrow{d}\zeta$, где ζ — случайная величина, имеющая стандартное нормальное распределение.*

Доказательство. Докажем эквивалентное утверждение о сходимости последовательности соответствующих характеристических функций:

$$\varphi_{S_n^0}(t)\underset{n\to\infty}{\to}\varphi(t)=e^{-\frac{t^2}{2}},\ t\in\mathbb{R}.$$

В свою очередь, мы можем показать эквивалентную сходимости последовательности функций

$$\ln \varphi_{S_n^0}(t) \underset{n\to\infty}{\to} -\frac{t^2}{2},\ t \in \mathbb{R}.$$

Построим асимптотическое разложение для функции $\ln \varphi_{S_n^0}(t)$ в произвольной точке $t \in \mathbb{R}$, используя свойства характеристической функции и условие Линдеберга:

$$\ln \varphi_{S_n^{(0}}(t) = \ln \prod_{k=1}^{n} \varphi_{\xi kn(t)} = \sum_{k=1}^{n} \ln\left(1 + \Delta_{kn}(t)\right) \tag{2.5}$$

где

$$\Delta_{kn}(t) := \varphi_{\xi_{kn}}(t) - 1 = \mathbb{E}e^{it\xi_{kn}} - 1$$

Рассмотрим асимптотику $\Delta_{kn}(t)$ при $n \to \infty$. Используя неравенства леммы 2.6, имеем

$$|\Delta_{kn}(t)| = |\mathbb{E}(e^{it\xi_h} - 1 - it\xi_{kn})| \leqslant \mathbb{E}\left\{\frac{t^2}{2}\xi_{kn}^2\right\} = \frac{t^2}{2} \cdot \frac{\sigma_k^2}{B_n^2}.$$

Отсюда и условия равномерной малости дисперсий имеем

$$\max_{1\leqslant k\leqslant n} |\Delta_{kn}(t)| \leqslant \frac{t^2}{2} \max_{1\leqslant k\leqslant n} \frac{\sigma_k^2}{B_n^2} \underset{n\to\infty}{\to} 0.$$

Значит, существует такое $n_0 < \infty$, что для всех $n > n_0$, для всех $k \in \{1, \ldots, n\}$ выполнено неравенство $|\Delta_{kn}(t)| < 1/2$.

В формуле 2.5 воспользуемся линейной формулой Тейлора

$$\ln \varphi_{S_n^0}(t) = \sum_{k=1}^{n} \Delta_{kn}(t) + r_n, t \in \mathbb{R}$$

где для остаточного члена справедлива следующая оценка:

$$|r_n| = \left|\sum_{k=1}^{n}\{\ln\left(1 + \Delta_{kn}(t)\right) - \Delta_{kn}(t)\}\right| \leqslant \sum_{k=1}^{n} |\Delta_{kn}(t)| \cdot |\Delta_{kn}(t)| \leqslant$$

$$\leqslant \max_{1\leqslant k\leqslant n} |\Delta_{kn}(t)| \cdot \frac{t^2}{2} \cdot \sum_{k=1}^{n} \frac{\sigma_k^2}{B_n^2} = \frac{t^2}{2} \max_{1\leqslant k\leqslant n} |\Delta_{kn}(t)| \to 0$$

Таким образом, $r_n \to 0$, $(n \to \infty)$ для всех $t \in \mathbb{R}$.

Оценим теперь

$$\Delta_n = \left|\ln \varphi_{S_n^0}(t) - \left(-\frac{t^2}{2}\right)\right| = |r_n + \sum_{k=1}^{n} \mathbb{E}(e^{it\xi_{kn}} - 1 - it\xi_{kn}) +$$
$$+\frac{t^2}{2}\sum_{k=1}^{n}\mathbb{E}\xi_{kn}^2| \leqslant |r_n| + \sum_{k=1}^{n}\mathbb{E}\left\{\left|e^{it\xi_{kn}} - 1 - it\xi_{kn} + \frac{(t\xi_{kn})^2}{2}\right|\right\}$$

Выберем произвольное $\tau > 0$ и воспользуемся леммой 2.6:

$$\Delta_n \leqslant |r_n| + \sum_{k=1}^{n}\mathbb{E}\left\{\left|e^{it\xi_{kn}} - 1 - it\xi_{kn} + \frac{(t\xi_{kn})^2}{2}\right|; |\xi_{kn}| \leqslant \tau\right\} +$$
$$+\sum_{k=1}^{n}\mathbb{E}\left\{\left|e^{it\xi_{kn}} - 1 - it\xi_{kn} + \frac{(t\xi_{kn})^2}{2}\right|; |\xi_{kn}| > \tau\right\} \leqslant$$
$$\leqslant |r_n| + \sum_{k=1}^{n}\mathbb{E}\left\{\frac{|t|^3\tau\xi_{kn}^2}{6}; |\xi_{kn}| \leqslant \tau\right\} +$$
$$+\sum_{k=1}^{n}\mathbb{E}\left\{\left|e^{it\xi_{kn}} - 1 - it\xi_{kn} + \frac{(t\xi_{kn})^2}{2}\right|; |\xi_{kn}| > \tau\right\} \leqslant \frac{|t^3|\tau}{3},$$

для любого $n > n_0$ и для любых $t \in \mathbb{R}$ $\tau > 0$. Здесь мы воспользовались условием Линдеберга и предыдущей оценкой $|r_n|$. В результате получим

$$\lim_{n\to\infty} \Delta_n \leqslant \frac{|t|^3\tau}{3}$$

для любых $t \in \mathbb{R}$, $\tau > 0$. В силу произвольности τ приходим к выводу, что $\Delta_n \to 0,\ (n \to \infty)$.

□

2.9.3 Модели роста

Рассмотрим непосредственное применение центральной предельной теоремы в так называемых «моделях роста» на примере роста дерева. На самом деле, эти модели могут успешно описывать рост популяций бактерий, популяций птиц, распространение пожара и так далее.

Пусть x_0 — рост дерева в момент посадки и мы производим ежегодный замер высоты дерева $x_1, x_2, \ldots, x_n$. Нас будет интересовать рост дерева X в некоторый «далёкий» момент времени. Ясно, что рост дерева зависит от множества факторов, поэтому прогноз будет вероятностным:

с некоторой вероятностью высота дерева будет не меньше чем столько-то метров. Такое заключение нельзя, конечно, использовать, чтобы предсказать рост одного дерева, но оно может быть успешно использовано для некоторого лесного массива, посаженного в одно время.

Итак, пусть $x_1, x_2, \ldots, x_n$ — реализации изучаемой случайной величины X в предположении независимости измерений. Обозначим

$$\Delta x_k = x_k - x_{k-1}$$

— прирост дерева за год. Предположим, что «суммарный фактор роста» ξ_k влияет на величину роста линейно, то есть

$$\Delta x_k = \alpha_k \xi_k,$$

где ξ_k — независимые одинаково распределённые случайные величины. Положим $\mathbb{E}\xi_k = a$, $\mathbb{D}\xi_k = b^2$. Естественно предположить, что

$$\alpha_k = g(x_{k-1}),$$

где $g(x)$ — непрерывна и $g(x) > 0$.

Тогда

$$\xi_k = \frac{\Delta x_k}{g(x_{k-1})},\ k = 1, \ldots, n.$$

Просуммируем обе части по k:

$$\sum_{k=1}^{n} \xi_k = \sum_{k=1}^{n} \frac{\Delta x_k}{g(x_{k-1})}.$$

В правой части равенства стоит интегральная сумма, значит

$$\sum_{k=1}^{n} \xi_k \approx \int\limits_{x_0}^{x_n} \frac{dt}{g(t)}.$$

Так как в левой части стоит сумма независимых одинаково распределённых случайных величин, то по центральной предельной теореме случайная величина $\sum\limits_{k=1}^{n} \xi_k$ стремится к нормальному распределению $N(\mu, \sigma^2)$, где $\mu = na,\ \sigma^2 = nb^2$. Обозначим

$$\int\limits_{x_0}^{x_n} \frac{dt}{g(t)} = h(x).$$

Ясно, что $h(x)$ — неубывающая функция. Тогда

$$\mathbf{P}\{X < x\} = \mathbf{P}\{h(X) < h(x)\} \approx \Phi\left(\frac{h(x)-\mu}{\sigma}\right),\ x > 0.$$

Естественно здесь предположить, что $g(t) = t$. В этом случае

$$F(x) = \Phi\left(\frac{\ln(x)-\mu}{\sigma}\right),\ x > 0.$$

Распределение, соответствующее этой функции распределения, называется логнормальным распределением. Название следует из того, что

$$\ln(X) \sim N(\mu, \sigma^2).$$

Функция плотности будет равна

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}x} e^{-\frac{(\ln x-\mu)^2}{2\sigma^2}},\ x > 0.$$

2.10 Задача регрессии. Условное математическое ожидание

Пусть имеются две случайные величины ξ и η, причём значения случайной величины ξ наблюдаемы. Требуется найти функцию $g = g(x)$ так, чтобы

$$\mathbb{E}(\eta - g(\xi))^2 \longmapsto \min .$$

Эта задача называется задачей регрессии.

Как найти функцию $g = g(x)$? Начнём с самого простого случая.

1. Пусть $\xi = \xi(\omega)$ — вырожденная случайная величина, то есть, существует такая константа c, что $\mathbf{P}\{\xi = c\} = 1$. Тогда нам нужно найти $b = g(c)$, такую, что

$$\mathbb{E}(\eta - b)^2 \longmapsto \min .$$

Эту задачу мы решали когда рассматривали свойства дисперсии:

$$b = g(c) = \mathbb{E}\eta.$$

2. Пусть теперь ξ — случайная величина дискретного типа с конечным числом значений, то есть,

$$\xi(\omega) = \sum_n c_n I_{A_n}(\omega),$$

где $A_n = \{\omega : \xi(\omega) = c_n\}$.

Сведём решение этой задачи к предыдущей. Зафиксируем значения ${}_n$ и найдём $g(\xi)$ при условии, что событие $\{\xi = {}_n\}$ произошло. Положим $g({}_n) = b_n$. Тогда нам нужно минимизировать

$$\mathbb{E}(\eta - g(\xi))^2 = \sum_n \int\limits_{A_n} (\eta - b_n)^2 \mathbf{P}(d\omega).$$

Обозначим через

$$\mathbf{P}_n(\cdot) = \mathbf{P}(\cdot | A_n) = \frac{\mathbf{P}(\cdot A_n)}{\mathbf{P}(A_n)}$$

— условную вероятность. Тогда из предыдущего случая следует, что

$$b_n = \mathbb{E}_n \eta = \int\limits_{A_n} \eta \mathbf{P}_n(d\omega) = \frac{1}{\mathbf{P}(A_n)} \int\limits_{-\infty}^{\infty} \eta I_{A_n} \mathbf{P}(d\omega) = \frac{1}{\mathbf{P}(A_n)} \mathbb{E}(\eta I_{A_n}).$$

Значит,

$$g(\xi) = \sum_n g(c_n) I_{A_n} = \sum_n b_n I_{A_n} = \sum_n \frac{1}{\mathbf{P}(A_n)} \mathbb{E}(\eta I_{A_n}) I_{A_n}.$$

Последнее выражение называется условным математическим ожиданием случайной величины η относительно случайной величины ξ. Обозначается это так:

$$\mathbb{E}(\eta|\xi) =: \sum_n \frac{1}{\mathbf{P}(A_n)} \mathbb{E}(\eta I_{A_n}) I_{A_n}.$$

3. Пусть, наконец, ξ — случайная величина непрерывного типа. Нам бы хотелось применить тот же подход, что и в случае дискретного распределения, и вычислить условное математическое ожидание относительно события $\{\xi = x\}$ для всех x. Но мы знаем, что в случае непрерывного распределения $\mathbf{P}\{\xi = x\} = 0$. Тем не менее, найдём условную функцию плотности случайной величины η относительно события $\{\xi = x\}$.

Пусть $p_{\xi,\eta}(x, y)$ — совместная функция плотности случайной величины ξ и η. Вначале запишем условную функцию распределения:

$$F_{\eta|\xi}(y|x) = \mathbf{P}\{\eta < y|\xi = x\} = \lim_{\Delta x \to 0} \mathbf{P}\{\eta < y|x \leqslant \xi < x + \Delta x\} =$$

$$= \lim_{\Delta x \to 0} \frac{\mathbf{P}\{\eta < y,\ x \leqslant \xi < x + \Delta x\}}{\mathbf{P}\{x \leqslant \xi < x + \Delta x\}} = \lim_{\Delta x \to 0} \frac{\int\limits_{-\infty}^{y} dt \int\limits_{x}^{x+\Delta x} p_{\xi,\eta}(t, s) ds}{\int\limits_{x}^{x+\Delta x} p_\xi(t) dt}.$$

Применяя теорему о среднем для последних интегралов, получим

$$\mathbf{P}\{\eta < y|\xi = x\} = \lim_{\Delta x \to 0} \frac{\int\limits_{-\infty}^{y} p_{\xi,\eta}(t, c') \Delta x dt}{p_\xi(c'') \Delta x} = \frac{\int\limits_{-\infty}^{y} p_{\xi,\eta}(t, x) dt}{p_\xi(x)},$$

$(x \leqslant c' < x + \Delta x,\ x \leqslant c'' < x + \Delta x)$. Значит, условная функция плотности будет равна:

$$p_{\eta|\xi}(y|x) = \frac{p_{\xi,\eta}(x, y)}{p_\xi(x)}.$$

Тогда, следуя аналогичным рассуждениям, что и выше, получим

$$g(x) = \mathbb{E}_x \eta = \int\limits_{-\infty}^{+\infty} y p_{\eta|\xi}(y|x) dy$$

Таким образом, мы нашли вид функции $g = g(x)$. Это — условное математическое ожидание случайной величины η относительно события $\{\xi = x\}$.

Случай нормального распределения

Особо выделим случай, когда (ξ, η) — нормальный вектор.

Теорема 2.24 (о нормальной корреляции)**.** *Пусть* (ξ, η) *— нормальный случайный вектор, с вектором средних значений* (m_1, m_2) *и ковариационной матрицей*

$$\Lambda = \begin{pmatrix} \mathbb{D}\xi & Cov(\xi, \eta) \\ Cov(\xi, \eta) & \mathbb{D}\eta \end{pmatrix}$$

причём $\mathbb{D}\xi > 0$. *Тогда оптимальная оценка* $g(\xi)$ *случайной величины* η *по* ξ *есть*

$$g(\xi) = \mathbb{E}(\eta|\xi) = \mathbb{E}\eta + \frac{Cov(\xi, \eta)}{\mathbb{D}\xi}(\xi - \mathbb{E}\xi),$$

и её ошибка равна

$$\Delta \equiv \mathbb{E}(\eta - \mathbb{E}(\eta|\xi))^2 = \mathbb{D}\eta - \frac{Cov(\xi, \eta)}{\mathbb{D}(\xi)}.$$

Доказательство. Отметим, что в этом случае функция $g(\xi)$ — линейная, то есть, речь идет о линейной регрессии.

Действительно,

$$p_{\xi,\eta}(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \times$$

$$\times \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-m_1)^2}{\sigma_1^2} - 2\rho\frac{(x-m_1)(y-m_2)}{\sigma_1\sigma_2} + \frac{(y-m_2)^2}{\sigma_2^2}\right]\right\},$$

где $\sigma_1^2 = \mathbb{D}\xi$, $\sigma_2^2 = \mathbb{D}\eta$, $\rho = Cov(\xi, \eta)/(\sigma_1\sigma_2)$.

Тогда, вычислив условную функцию плотности, мы можем преобразовать её к виду

$$p_{\eta|\xi}(y|x) = \frac{1}{\sqrt{2\pi(1-\rho^2)}\,\sigma_2} \exp\left\{-\frac{(y-m(x))^2}{2\sigma_2^2(1-\rho^2)}\right\},$$

где

$$m(x) = m_2 + \frac{\sigma_2}{\sigma_1}\rho(x - m_1).$$

Тогда

$$\mathbb{E}(\eta|\xi) = \int\limits_{-\infty}^{+\infty} y p_{\eta|\xi}(y|x)dy = m(x),$$

и

$$\mathbb{D}(\eta|\xi) = \int\limits_{-\infty}^{+\infty} (y - m(x))^2 p_{\eta|\xi}(y|x) dy = \sigma_2^2(1 - \rho^2).$$

$\square$

К вопросу линейной регрессии мы ещё вернемся в курсе математической статистики.

Литература

[1] Боровков, А. А. Теория вероятностей — Москва: Наука, 1998.

[2] Володин И. Н. Лекции по теории вероятностей и математической статистике. — Казань: (Издательство), 2006.

[3] Гнеденко Б. В. Курс теории вероятностей // Изд. 6-е, перераб. и доп. — Москва: Наука. Гл. ред. физ.-мат. лит., 1988.

[4] Муштари Д.Х. Вероятность, математическая статистика, случайные процессы // Учебное пособие — Изд. Казанского университета, 2011.

[5] Харин Ю.С., Зуев Н.М., Жук Е.Е. Теория вероятностей, математическая и прикладная статистика — Минск: БГУ, 2011.

[6] Ширяев А. Н. Вероятность-1 // Элементарная теория вероятностей. Математические основания. Предельные теоремы // Издание четвертое, переработанное и дополненное. — Москва: Издательство МЦНМО, 2007.

[7] Ширяев А. Н. Вероятность-2 // Суммы и последовательности случайных величин — стационарные, мартингалы, марковские цепи // Издание четвертое, переработанное и дополненное. — Москва: Издательство МЦНМО, 2007.