

В. И. ЛОТОВ

ТЕОРИЯ ВЕРОЯТНОСТЕЙ
И
МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

*Семестровый курс лекций для студентов факультета информационных
технологий*

Содержание

I. Теория вероятностей	4
1. Вероятностные пространства. Основные формулы	4
1.1. Дискретные пространства	4
1.2. Континуальные пространства	9
1.3. Вероятностное пространство общего вида	12
1.4. Независимые события	14
1.5. Схема Бернулли	15
1.6. Условные вероятности	17
1.7. Формула полной вероятности	18
1.8. Формула Байеса	19
2. Распределения	20
2.1. Случайные величины. Функции распределения	20
2.2. Типы распределений. Примеры	23
2.3. Многомерные распределения и плотности	31
2.4. Преобразования случайных величин	35
3. Числовые характеристики распределений	39
3.1. Математическое ожидание	39
3.2. Моменты	44
3.3. Дисперсия	44
3.4. Коэффициент корреляции	46
3.5. Многомерный случай: математическое ожидание и матрица ковариаций	48
3.6. Многомерное нормальное распределение	49
4. Предельные теоремы	51
4.1. Сходимость по вероятности	51
4.2. Закон больших чисел	53
4.3. Центральная предельная теорема	54
4.4. Приближение Пуассона в схеме Бернулли	57
II. Математическая статистика	60
5. Введение	60
5.1. Основные понятия	60
5.2. Выборочные характеристики	61
6. Оценивание неизвестных параметров	64
6.1. Постановка задачи. Несмещенность и состоятельность	64
6.2. Метод моментов	65
6.3. Метод максимального правдоподобия	67
6.4. Сравнение оценок	70

7. Доверительные интервалы	71
7.1. Некоторые распределения, связанные с нормальным	71
7.2. Свойства выборок из нормального распределения	74
7.3. Доверительные интервалы для параметров нормального распределения	76
7.4. Построение доверительных интервалов с помощью нормального приближения	77
8. Проверка гипотез	78
8.1. Постановка задачи, основные понятия	78
8.2. Критерий Колмогорова	79
8.3. Критерий хи-квадрат Пирсона	80
8.4. Построение критерия с помощью доверительного интервала	82
8.5. Проверка гипотез в случае двух выборок	82
8.6. Дисперсионный анализ: однофакторная модель	85
9. Задачи линейной регрессии	87
9.1. Постановка задачи	87
9.2. Метод наименьших квадратов	90
9.3. Доверительные интервалы и проверка гипотез	91

Часть I.

Теория вероятностей

1. Вероятностные пространства. Основные формулы

1.1. Дискретные пространства

До возникновения теории вероятностей объектом исследования науки были явления или опыты, в которых условия эксперимента позволяют исследователю однозначно определить исход эксперимента. Так, например, в химии: если известны вещества, вступающие в реакцию, их свойства, условия, в которых будет протекать реакция, то однозначно можно предсказать исход реакции. В механике: если известны масса тела, все силы, которые на него действуют, координаты и начальная скорость, то нетрудно вычислить траекторию последующего движения.

Однако есть ряд явлений и экспериментов, которые называются *случайными* и которые характеризуются невозможностью предсказать их исход до начала эксперимента.

Рассмотрим некоторые примеры.

1. Однократное подбрасывание монеты. Здесь возможны два исхода, их принято обозначать «Г» (герб) и «Р» (решка).

2. Однократное бросание игральной кости (т. е. кубика, у которого на гранях нанесены числа от 1 до 6). Здесь возможны шесть исходов эксперимента: 1, 2, 3, 4, 5, 6.

3. Подсчет количества вызовов, пришедших в течение часа на АТС (автоматическую телефонную станцию) для обслуживания. Поступить может любое число вызовов: 0, 1, 2, ...

4. Определение времени безотказной работы прибора. Исходом этого эксперимента может быть любое неотрицательное число из $[0, \infty)$.

5. Движение броуновской частицы на плоскости в течение минуты. В результате этого эксперимента может осуществиться любая из бесконечного множества траекторий.

Теория вероятностей, как и всякая другая математическая дисциплина, строит и изучает математическую модель тех или иных явлений, в данном случае — случайных явлений.

Казалось бы, какие научные результаты можно получить относительно подбрасывания монеты? Если подбрасывание однократное, то, действительно, мало интересного можно сказать. Но если, к примеру, подбрасывать монету n раз и подсчитать количество S_n выпавших гербов, то окажется, что при увеличении n отношение S_n/n стремится к $1/2$. Этот факт был замечен давно, многие исследователи эмпирическим путем его перепроверяли. Так, в опытах французского исследователя Бюффона монета бросалась 4 040 раз, выпало 2 048 гербов, что привело к результату $S_n/n = 0.507$. Английский статистик Пирсон в 24 000 бросаниях получил 12 012 гербов, при этом $S_n/n = 0.5005$.

Обнаруженная закономерность — одна из простейших, она является следствием так называемого *закона больших чисел*. Эта и ряд других предельных закономерностей будут изучены нами позже.

А пока займемся построением математической модели случайных явлений. Для этого нужно выделить у изучаемых явлений общие черты и наделить ими модель. При этом надо постараться отразить наиболее существенные черты рассматриваемых

явлений и отбросить несущественные. Модель не должна быть слишком сложной, иначе изучать ее будет затруднительно.

Какие же общие черты имеются у явлений, рассмотренных в примерах 1 – 5? У каждого из них имеется некоторый набор возможных исходов эксперимента. Будем обозначать его греческой буквой Ω и называть *пространством элементарных исходов*. У каждого случайного эксперимента оно свое — подчеркнем это. Если Ω конечно или счетно, то будем называть его *дискретным*. Из уже рассмотренных примеров дискретные пространства появляются в первых трех. Элементы множества Ω обычно обозначаются буквами ω с индексами или без них и называются *элементарными исходами*. Заметим, что, несмотря на использование часто встречающегося в математике термина «пространство», в нашем случае Ω — всего лишь абстрактное множество (не обязательно числовой природы), на этом множестве не вводятся операции сложения, умножения, нет там и отношения порядка.

Далее на протяжении всего параграфа мы ограничимся рассмотрением только дискретных пространств элементарных исходов.

Введем понятие *события*. Все хорошо представляют событие как нечто могущее произойти или уже происходящее. Нам нужно ввести в рассмотрение математическую модель этого «происходящего».

Определение. Событиями называются произвольные подмножества пространства элементарных исходов Ω .

Обозначать разные события будем буквами A, B, C, \dots с индексами или без них.

Мы будем говорить, что событие $A \subset \Omega$ произошло, если в результате случайного эксперимента реализовался один из элементарных исходов $\omega \in A$.

Убедимся на примерах, что каждое подмножество Ω действительно соответствует осуществлению некоторого события в данном случайном эксперименте. Так, подмножество $\{2, 4, 6\} \subset \Omega$ в примере 2 соответствует тому, что в результате бросания игральной кости выпало четное число очков. Рассмотрим эксперимент из примера 3. Если описать здесь словами какое-нибудь событие, скажем, поступление на АТС не менее 10 вызовов за час, то ясно, что такому событию будет соответствовать множество $\{10, 11, 12, \dots\} \subset \Omega$.

Пустое множество $\emptyset \subset \Omega$ также, по определению, является событием, оно называется *невозможным* (никогда не может произойти). Все пространство $\Omega \subset \Omega$ тоже есть событие, оно называется *достоверным*. Совокупность всех возможных событий обозначим \mathcal{S} , в дискретном пространстве это совокупность всех подмножеств Ω .

Если из $\omega \in A$ следует $\omega \in B$, т. е. $A \subset B$, то мы говорим, что событие A влечет событие B (но не наоборот!).

Над событиями, как над множествами, можно производить операции объединения, пересечения, разности, перехода к дополнительному множеству, причем операции объединения и пересечения будут применяться как к конечному, так и к бесконечному набору событий. Напомним некоторые определения:

$\bigcup_{i=1}^{\infty} A_i = \{\omega : \omega \in A_i \text{ хотя бы при одном } i\}$ — *объединение* событий (означает, что происходит хотя бы одно из A_1, A_2, \dots);

$\bigcap_{i=1}^{\infty} A_i = \{\omega : \omega \in A_i \text{ при всех } i = 1, 2, \dots\}$ — *пересечение* событий (означает, что происходят одновременно все указанные события);

$A \setminus B = \{\omega : \omega \in A, \text{ но } \omega \notin B\}$ — *разность* двух событий;

$\bar{A} = \Omega \setminus A = \{\omega : \omega \notin A\}$ — *дополнительное* событие или просто *дополнение* к A .

Перечисление различных свойств этих операций не входит в программу нашего курса, мы остановимся только на одном соотношении, которое будет использоваться

в дальнейшем.

Формула двойственности. Для любой последовательности событий A_1, A_2, \dots справедливо

$$\overline{\bigcap_{i=1}^{\infty} A_i} = \bigcup_{i=1}^{\infty} \overline{A_i}.$$

Докажем это соотношение. Если $\omega \in \overline{\bigcap_{i=1}^{\infty} A_i}$, то $\omega \notin \bigcap_{i=1}^{\infty} A_i$, т. е. существует номер i такой, что $\omega \notin A_i$ или, что то же самое, $\omega \in \overline{A_i} \subset \bigcup_{i=1}^{\infty} \overline{A_i}$. Если же $\omega \in \bigcup_{i=1}^{\infty} \overline{A_i}$, то существует номер i такой, что $\omega \in \overline{A_i}$. Значит, $\omega \notin A_i$, т. е. $\omega \notin \bigcap_{i=1}^{\infty} A_i$ и, следовательно, $\omega \in \overline{\bigcap_{i=1}^{\infty} A_i}$.

Далее мы введем понятие *вероятности* события. Вообще говоря, вероятность — это числовая функция на \mathcal{S} , обладающая определенными свойствами. Для дискретных пространств мы определим ее в два этапа. Сначала только для событий, состоящих из одного единственного элементарного исхода.

Каждому элементарному исходу $\omega \in \Omega$ поставим в соответствие число $\mathbf{P}(\omega)$, называемое вероятностью этого элементарного исхода, так, чтобы были выполнены следующие два требования:

- 1) $\mathbf{P}(\omega) \geq 0$;
- 2) $\sum_{\omega \in \Omega} \mathbf{P}(\omega) = 1$.

Какие конкретно значения следует задавать — не так уж важно, это обычно определяется условиями эксперимента. Так, в примере 1 мы припишем вероятности $1/2$ каждому элементарному исходу, если монетка симметрична; в примере 2 (бросание игральной кости) можно задать равные вероятности по $1/6$ для каждого элементарного исхода. В третьем примере мы уже не можем приписать каждому элементарному исходу одну и ту же положительную вероятность — тогда сумма всех вероятностей не будет равна единице. Как показывают эксперименты, для вероятности того, что за единицу времени на АТС поступит ровно k вызовов, лучше всего подходит число $\frac{\lambda^k}{k!} e^{-\lambda}$ при некотором $\lambda > 0$.

Теперь мы можем определить вероятность произвольного события $A \subset \Omega$. Положим, по определению,

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

Будем считать, кроме того, что $\mathbf{P}(\emptyset) = 0$.

Мы тем самым завершили построение математической модели эксперимента с дискретным пространством элементарных исходов. Она состоит из тройки $\langle \Omega, \mathcal{S}, \mathbf{P} \rangle$ и называется *вероятностным пространством*.

Подчеркнем, что данное выше определение вероятности события годится только для дискретных моделей. Далее мы рассмотрим некоторые основные свойства вероятности в дискретной модели.

Свойства вероятности

1. $0 \leq \mathbf{P}(A) \leq 1$, $\mathbf{P}(\Omega) = 1$.
2. Если $A \subset B$, то $\mathbf{P}(A) \leq \mathbf{P}(B)$.

Эти два свойства очевидным образом вытекают из определения вероятности.

3. $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(AB)$, где для краткости записи обозначено $\mathbf{P}(AB) = \mathbf{P}(A \cap B)$.

Для доказательства этого соотношения обратимся сначала к его правой части. Вычисляя $\mathbf{P}(A)$, мы суммируем вероятности всех элементарных исходов из A , затем прибавляем сумму вероятностей элементарных исходов из B . Тем самым получается, что вероятности элементарных исходов из множества AB мы просуммировали дважды. Значит, один раз нужно их отнять.

События A и B называются *несовместными*, если $AB = \emptyset$. Из доказанного свойства следует, в частности, что $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B)$, если события A и B несовместны.

Последнее называется *аддитивностью* вероятности. Разумеется, с помощью индукции можно распространить это свойство на любое конечное число взаимно несовместных событий.

4. $\mathbf{P}(\bar{A}) = 1 - \mathbf{P}(A)$, поскольку $A \cup \bar{A} = \Omega$, $\mathbf{P}(A) + \mathbf{P}(\bar{A}) = \mathbf{P}(\Omega) = 1$.

5. Если события A_1, A_2, \dots попарно несовместны, т. е.

$$A_i A_j = \emptyset \quad (i \neq j),$$

то

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

Данное свойство называется *счётной аддитивностью*. Оно также легко следует из определения вероятности:

$$\mathbf{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{\omega \in \bigcup_{i=1}^{\infty} A_i} \mathbf{P}(\omega) = \sum_{\omega \in A_1} \mathbf{P}(\omega) + \sum_{\omega \in A_2} \mathbf{P}(\omega) + \dots = \sum_{i=1}^{\infty} \mathbf{P}(A_i).$$

Важный частный случай: классическое определение вероятности

Среди дискретных моделей мы более подробно рассмотрим такие, у которых:

1) $N(\Omega) = N < \infty$ (здесь $N(\Omega)$ обозначает число элементов множества Ω);

2) $\mathbf{P}(\omega_1) = \dots = \mathbf{P}(\omega_N) = 1/N$.

Вероятностные пространства, удовлетворяющие таким свойствам, используются очень часто. Первые два примера из рассмотренных приводят именно к таким моделям. Посмотрим, как будет вычисляться в такой ситуации вероятность события. Для любого события A

$$\mathbf{P}(A) = 1/N + \dots + 1/N.$$

Число слагаемых в правой части равно числу элементов множества A , т. е. $N(A)$, поэтому получаем

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

Это так называемое *классическое определение вероятности*. Как видим, в соответствии с этим определением вероятность равна отношению числа «благоприятных» исходов (т. е. тех, которые формируют интересующее нас событие) к числу всех возможных исходов эксперимента. Формула проста, но она не универсальна, ее применимость ограничивается приведенными выше двумя условиями.

Для вычисления вероятностей с помощью классического определения часто требуется применять некоторые методы и результаты из комбинаторики. Напомним кратко решения некоторых комбинаторных задач.

1. Пусть имеется совокупность из n различных объектов a_1, a_2, \dots, a_n . Сколькими способами можно расставить их в ряд?

Это задача о *перестановках*. На первое место в этом ряду можно поставить любой из n имеющихся объектов, на второе — любой из $n - 1$ оставшихся и т. д. В итоге получаем $n(n - 1)(n - 2) \cdots 2 \cdot 1 = n!$ перестановок (когда каждый из вариантов для одной позиции может объединяться с любым вариантом для другой позиции, то общее число вариантов получается перемножением, а не сложением, это легко проверить на примерах).

2. Пусть исходная совокупность a_1, a_2, \dots, a_n та же, что и в предыдущей задаче, но теперь мы будем выбирать из нее подсовкупность, состоящую из k объектов (будем говорить, что мы делаем выборку объема k), $k = 1, 2, \dots, n$. Сколько различных выборок можно получить?

Если действовать, как в предыдущем пункте, мы можем выбрать первый объект n способами, второй $n - 1$ способом, и так до тех пор, пока не наберем k объектов. Тем самым количество выборок получится равным $n(n - 1)(n - 2) \cdots (n - k + 1) = \frac{n!}{(n - k)!}$. Нетрудно убедиться на примерах, что полученное нами число выборок объема k включает выборки, различающиеся и по составу элементов, и по порядку расположения их внутри выборки.

Если мы хотим ограничить себя выборками, различающимися только составом входящих в них элементов и не принимать во внимание порядок элементов внутри выборки, то мы должны полученное выше число разделить на $k!$, так как каждая выборка фиксированного состава нами посчитана там $k!$ раз со всеми ее перестановками элементов.

Таким образом, мы получаем число $C_n^k = \frac{n!}{k!(n - k)!}$. Его обычно называют *числом сочетаний* из n по k . Числа C_n^k , $k = 0, 1, \dots, n$, также называют *биномиальными коэффициентами*, поскольку они участвуют в формуле бинома Ньютона

$$(x + y)^n = \sum_{k=0}^n C_n^k x^k y^{n-k}$$

(мы используем обозначение $C_n^0 = 1$, это удобно). Из бинома сразу следует, что $\sum_{k=0}^n C_n^k = 2^n$. Отметим также очевидное свойство $C_n^k = C_n^{n-k}$.

3. Имеется n ящиков и k различных шаров. Шары произвольным образом размещаются по ящикам без каких-либо ограничений. Скольким числом способов можно это сделать?

Здесь первый шар может быть положен в любой из n ящиков, независимо от этого у второго шара тоже n вариантов и т. д. Перемножая количества вариантов, получаем n^k различных способов размещения.

Эта задача может встретиться в другой интерпретации.

Предположим, что в алфавите n букв. Сколько слов длины k можно составить?

Ясно, что в качестве первой буквы может быть взята любая из n букв алфавита, в качестве второй — тоже любая буква алфавита и т. д. Всего получаем n^k различных слов.

Различаются выборки *с возвращением* и *без возвращения*. Примером выборок с возвращением являются разные слова в последней задаче: здесь после выбора первой буквы слова исходная совокупность (алфавит) не уменьшилась и на втором шаге мы вновь имеем n вариантов, так же и для третьей, четвертой и других букв слова. А вот при получении числа C_n^k мы делали выборки без возвращения, так как, выбирая

последовательно один объект за другим, мы уменьшали исходную совокупность.

В качестве примера применения классического определения вероятности рассмотрим одну часто встречающуюся задачу.

В ящике находится n различных шаров (скажем, шары пронумерованы), из них n_1 белых шаров и $n - n_1$ чёрных. Наугад выбираем k шаров. Какова вероятность того, что среди выбранных шаров окажется ровно k_1 белых?

Эта задача может встретиться в других терминах. Например:

1. Среди лотерейных билетов есть выигрышные (их n_1) и проигрышные ($n - n_1$). Какова вероятность того, что среди k приобретенных билетов ровно k_1 выигрышных?

2. Среди n изделий n_1 бракованных, остальные годные. Какова вероятность того, что среди k выбранных наугад изделий обнаружится ровно k_1 бракованных?

Примеров таких ситуаций много.

Для решения будем пользоваться классической моделью. Мы делаем выборки объёма k , их всего C_n^k , и все они равновозможны. Количество благоприятных исходов получается так: сначала выбираем любые k_1 белых шаров из общего количества n_1 белых шаров, это можно сделать $C_{n_1}^{k_1}$ способами. Затем набираем $k - k_1$ черных шаров из $n - n_1$ имеющихся, получаем $C_{n-n_1}^{k-k_1}$ вариантов. После чего перемножаем эти два количества, поскольку каждый из наборов белых шаров может быть объединен в выборку с каждым из наборов черных шаров, в ответе получаем

$$C_{n_1}^{k_1} C_{n-n_1}^{k-k_1} / C_n^k.$$

Мы молчаливо предполагаем, что верхние индексы не превосходят нижних в записи участвующих здесь биномиальных коэффициентов, в противном случае ответ в задаче тривиален.

Совокупность полученных вероятностей при различных допустимых значениях переменной k_1 называется *гипергеометрическим распределением*.

1.2. Континуальные пространства

Как нетрудно видеть из предыдущего раздела, свойства вероятности аналогичны свойствам массы тела. Продолжая эту аналогию, можно считать, что вероятность события — это его масса, при этом множество Ω будет иметь единичную массу. В дискретном пространстве вся эта единичная масса разбросана по конечному или счетному набору точек. Теперь мы будем рассматривать другую крайность, когда вероятность как масса «размазана» по всему пространству элементарных исходов, которое, разумеется, уже не будет дискретным.

Мы будем предполагать здесь, что $\Omega = \mathbb{R}^n$, $n \geq 1$. В качестве событий, как и раньше, будем рассматривать подмножества Ω , хотя в отличие от дискретных пространств некоторые подмножества следовало бы исключить из рассмотрения по той причине, что нам не удастся задать приемлемым образом вероятности событий на всех без исключения подмножествах Ω . Эти «неудобные» множества имеют весьма сложную структуру. Они имеют чисто теоретический интерес и в приложениях практически не встречаются. Поэтому мы сознательно не будем акцентировать на них свое внимание.

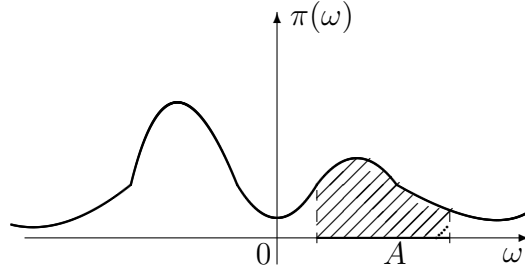
Для лучшего понимания остановимся сначала более подробно на случае $\Omega = \mathbb{R}^1$. Предположим, что у нас имеется некоторая функция $\pi : \Omega \rightarrow \mathbb{R}$ такая, что:

- 1) $\pi(\omega) \geq 0$ для любого $\omega \in \Omega$;
- 2) $\int_{-\infty}^{\infty} \pi(\omega) d\omega = 1$.

С помощью этой вспомогательной функции задается вероятность события. Положим, по определению,

$$\mathbf{P}(A) = \int_A \pi(\omega) d\omega.$$

Разумеется, мы вправе оперировать здесь только такими подмножествами $A \subset \Omega$, для которых интегрирование имеет смысл. Это определение имеет простой геометрический смысл: вероятность того или иного отрезка на прямой вычисляется как площадь криволинейной трапеции, имеющей данный отрезок своим основанием и ограниченной сверху графиком функции $\pi(\omega)$.



Вероятностное пространство, в котором таким образом задаются вероятности событий, называется *континуальным*.

Ясно, что при таком определении каждый элементарный исход имеет нулевую вероятность. Нетрудно проверить, что все свойства вероятности, перечисленные в предыдущем разделе, остаются в силе и для континуальных пространств.

Рассмотрим некоторые примеры функций π .

1. Пусть для некоторых $a < b$

$$\pi(\omega) = \begin{cases} \frac{1}{b-a}, & \omega \in [a, b]; \\ 0, & \text{иначе.} \end{cases}$$

Мы видим, что при такой функции π будет выполняться $\mathbf{P}(A) = 0$ для любого множества A , не имеющего пересечений с $[a, b]$. Поэтому можно считать, что пространство Ω сужается до размеров отрезка $[a, b]$. При этом какое бы подмножество $A = [c, d] \subset \Omega = [a, b]$ ни взять, его вероятность равна

$$\mathbf{P}(A) = \frac{d-c}{b-a} = \frac{\lambda(A)}{\lambda(\Omega)},$$

где $\lambda(A)$ обозначает длину множества A .

Вероятности событий, вычисляемые по этому простому правилу как отношение длин множеств, называются *геометрическими*. Это есть непрерывный аналог классического определения вероятностей, рассмотренного ранее для дискретных схем.

Геометрическая вероятность не зависит от сдвигов множества A внутри Ω . Можно образно сказать, что в этом случае вероятностная масса равномерно «размазана» по отрезку $[a, b]$.

2. Пусть

$$\pi(\omega) = \frac{1}{\sqrt{2\pi}} e^{-\omega^2/2}.$$

В этом случае мы уже не можем говорить о равномерности «размазывания» вероятностной массы на прямой. Вероятность любого интервала будет максимальной, если

его центр находится в нуле, и будет убывать очень быстро по мере удаления этого интервала от начала координат.

3. Еще один пример:

$$\pi(\omega) = \begin{cases} e^{-\omega}, & \omega > 0; \\ 0, & \text{иначе.} \end{cases}$$

В этом случае можно считать, что $\Omega = [0, \infty)$.

Если $\Omega = \mathbb{R}^n$ и число $n \geq 1$ произвольно, то вероятность события определяется также с помощью вспомогательной функции $\pi(\omega)$, только теперь $\omega = (\omega_1, \omega_2, \dots, \omega_n)$, и по-прежнему выполнены такие требования:

- 1) $\pi(\omega) \geq 0$ для любого $\omega \in \Omega$;
- 2) $\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \pi(\omega) d\omega_1 \dots d\omega_n = 1$.

Полагаем, по определению, для $A \subset \Omega$

$$\mathbf{P}(A) = \int_A \int \dots \int \pi(\omega) d\omega_1 \dots d\omega_n.$$

Если функция π принимает постоянное значение на некотором ограниченном множестве $D \subset \mathbb{R}^n$ и равна нулю вне него, то, как и раньше, вычисление вероятности события $A \subset D$ производится элементарным геометрическим способом:

$$\mathbf{P}(A) = \frac{\lambda(A)}{\lambda(D)},$$

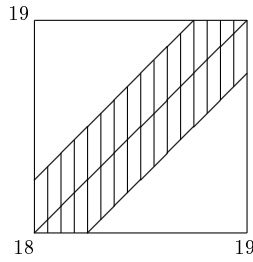
где $\lambda(A)$ здесь уже обозначает n -мерный объем множества A . Здесь, конечно, обязательно должно быть $\lambda(D) > 0$.

Рассмотрим в качестве примера задачу о встрече.

Два человека, A и B , договорились встретиться в определенном месте между 18 и 19 часами вечера. Однако момент встречи они никак не обозначили, а договорились о следующем. Тот, кто приходит первым, ждет в течение 15 минут. Если второй за это время не приходит, то первый уходит и встреча в этом случае не состоится. Разумеется, если первый придет, скажем, за 5 минут до 19 часов, то ждать все 15 минут нет никакого смысла, так как после 19 часов никто больше прийти не может.

Какова вероятность того, что встреча состоится?

Для решения задачи прежде всего нужно понять, как устроено пространство элементарных исходов. Обозначим через X момент прихода A и через Y момент прихода B . Ясно, что совокупность всевозможных пар (X, Y) , где $18 \leq X \leq 19$, $18 \leq Y \leq 19$ исчерпывает все исходы эксперимента, т. е. Ω — это квадрат на плоскости переменных X, Y . Поскольку молчаливо предполагается, что для моментов прихода каждого из них нет никаких предпочтений внутри промежутка $[18, 19]$, то мы выбираем модель с функцией π , равной единице в указанном квадрате. Иначе говоря, вычисление вероятности события будет производиться геометрическим способом, в данном случае как отношение площадей. Площадь всего Ω равна 1, нам остается выделить из квадрата подмножество точек (X, Y) , для которых встреча состоится. Это множество характеризуется неравенством $|Y - X| \leq 1/4$ или, что то же самое, $X - 1/4 \leq Y \leq X + 1/4$.



Как нетрудно видеть, площадь этого множества равна $1 - (3/4)^2 = 7/16$. Это и есть искомая вероятность.

1.3. Вероятностное пространство общего вида

Мы подробно изучили две различные вероятностные модели. В первой из них, дискретной, вероятностная масса распределялась по дискретному набору элементарных исходов; во втором — непрерывным образом «размазывалась» по пространству или по его части. Для каждой из этих моделей существует много применений. Однако рассмотренные модели являют собой всего лишь два крайних случая. Можно представить себе смешанные ситуации, когда часть вероятностной массы распределяется по дискретному множеству точек, а остальная масса «размазывается» непрерывно по другому множеству. Таким образом, разных вероятностных моделей можно строить бесконечно много. Однако, как бы они ни строились, они обязательно должны удовлетворять ряду требований, которые по сути являются аксиомами вероятностного пространства.

Итак, по-прежнему вероятностное пространство есть тройка $\langle \Omega, \mathcal{S}, \mathbf{P} \rangle$, где про Ω уже все сказано — это множество возможных исходов эксперимента, \mathcal{S} — совокупность подмножеств Ω , называемых событиями. В отличие от дискретной модели \mathcal{S} может включать в себя не все подмножества Ω . Не вдаваясь в подробности, мы будем считать тем не менее, что практически все важные с точки зрения приложений подмножества Ω входят в \mathcal{S} . Основное внимание сосредоточим на аксиомах задания вероятности. По-прежнему, вероятность — это числовая функция \mathbf{P} , областью определения которой является \mathcal{S} . Каким бы способом ни задавалась эта функция, она должна удовлетворять следующим трем аксиомам:

A1. $\mathbf{P}(A) \geq 0$ для любого $A \in \mathcal{S}$.

A2. $\mathbf{P}(\Omega) = 1$.

A3. Счётная аддитивность: если события A_1, A_2, \dots таковы, что $A_i A_j = \emptyset$ ($i \neq j$) (т. е. попарно несовместны), то

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

Из этих аксиом вытекает ряд полезных свойств вероятности. Некоторые из них мы имели возможность наблюдать в дискретных и в континуальных пространствах. Теперь установим свойства вероятности для произвольных вероятностных пространств. Все они являются следствиями введенных трех аксиом.

Свойства вероятности

1. $\mathbf{P}(\emptyset) = 0$.

Доказательство. Представим произвольное событие A в виде

$$A = A \cup \emptyset \cup \emptyset \cup \dots,$$

тогда по аксиоме АЗ

$$\mathbf{P}(A) = \mathbf{P}(A) + \mathbf{P}(\emptyset) + \mathbf{P}(\emptyset) + \dots,$$

что имеет место только при $\mathbf{P}(\emptyset) = 0$.

2. Аддитивность вероятности: для всякого конечного набора попарно несовместных событий A_1, A_2, \dots, A_n

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

Доказательство. Представим $\bigcup_{i=1}^n A_i$ в виде $A_1 \cup A_2 \cup \dots \cup A_n \cup \emptyset \cup \emptyset \dots$ и пользуемся счетной аддитивностью.

3. Для любого события A имеет место $\mathbf{P}(A) + \mathbf{P}(\bar{A}) = 1$ — это частный случай предыдущего утверждения. Выделяется в виде отдельного свойства ввиду частого использования при решении задач.

4. Для любых событий A и B

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

Доказательство. Представим событие $A \cup B$ в виде $B \cup (A \setminus B)$, тогда в силу аддитивности $\mathbf{P}(A \cup B) = \mathbf{P}(B) + \mathbf{P}(A \setminus B)$. Для нахождения последнего слагаемого воспользуемся представлением $A = AB \cup (A \setminus B)$, откуда опять по аддитивности $\mathbf{P}(A) = \mathbf{P}(AB) + \mathbf{P}(A \setminus B)$.

5. Если $A \subset B$, то $\mathbf{P}(A) \leq \mathbf{P}(B)$.

Доказательство. Поскольку $B = A \cup (B \setminus A)$, то из аддитивности и аксиомы А1 получаем $\mathbf{P}(B) = \mathbf{P}(A) + \mathbf{P}(B \setminus A) \geq \mathbf{P}(A)$.

6. Свойство непрерывности вероятности. Оно состоит из двух пунктов:

а) если события A_1, A_2, \dots таковы, что

$$A_1 \subset A_2 \subset A_3 \subset \dots,$$

то существует

$$\lim_{n \rightarrow \infty} \mathbf{P}(A_n) = \mathbf{P}\left(\bigcup_{i=1}^{\infty} A_i\right);$$

б) если $A_1 \supset A_2 \supset A_3 \supset \dots$, то существует

$$\lim_{n \rightarrow \infty} \mathbf{P}(A_n) = \mathbf{P}\left(\bigcap_{i=1}^{\infty} A_i\right).$$

Доказательство. а) Событие $\bigcup_{i=1}^{\infty} A_i$ можно представить в виде

$$\bigcup_{i=1}^{\infty} A_i = A_1 \cup (A_2 \setminus A_1) \cup (A_3 \setminus A_2) \cup \dots,$$

тогда участвующие здесь множества попарно несовместны и мы можем воспользоваться свойством счетной аддитивности:

$$\mathbf{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \mathbf{P}(A_1) + \mathbf{P}(A_2 \setminus A_1) + \mathbf{P}(A_3 \setminus A_2) + \dots$$

Поскольку сумма ряда есть предел последовательности частных сумм, то это выражение равно

$$\begin{aligned} & \lim_{n \rightarrow \infty} [\mathbf{P}(A_1) + \mathbf{P}(A_2 \setminus A_1) + \dots + \mathbf{P}(A_n \setminus A_{n-1})] = \\ & = \lim_{n \rightarrow \infty} \mathbf{P}(A_1 \cup (A_2 \setminus A_1) \cup \dots \cup (A_n \setminus A_{n-1})) = \lim_{n \rightarrow \infty} \mathbf{P}(A_n). \end{aligned}$$

Для доказательства пункта *б* перейдем к рассмотрению дополнительных событий и воспользуемся уже доказанным свойством *а*. Очевидно,

$$\bar{A}_1 \subset \bar{A}_2 \subset \bar{A}_3 \subset \dots,$$

поэтому

$$\lim_{n \rightarrow \infty} \mathbf{P}(A_n) = 1 - \lim_{n \rightarrow \infty} \mathbf{P}(\bar{A}_n) = 1 - \mathbf{P}\left(\bigcup_{i=1}^{\infty} \bar{A}_i\right) = 1 - \mathbf{P}\left(\overline{\bigcap_{i=1}^{\infty} A_i}\right) = \mathbf{P}\left(\bigcap_{i=1}^{\infty} A_i\right).$$

Здесь мы воспользовались доказанной ранее формулой двойственности.

Дальнейшее изложение материала будет относиться к вероятностным пространствам общего вида.

1.4. Независимые события

Что такое независимые события в жизни — понятно каждому. Это значит, что между событиями отсутствует причинно-следственная связь, осуществление одного никак не влияет на другое. Наша ближайшая цель — ввести для событий в нашей модели (т. е. для подмножеств пространства элементарных исходов) некоторое свойство, которое было бы отражением обиходного понимания независимости.

Определение. События A и B называются *независимыми*, если

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

Попробуем убедиться на примере, что приведенное в этом определении свойство действительно имеет место для тех событий в нашей модели, которые являются отражением независимых событий в жизни.

Пример. Из большой группы людей, где поровну мужчин и женщин, выбрали наугад человека. Пусть событие A означает, что выбрана женщина. Так как женщин — половина, то $\mathbf{P}(A) = 1/2$. Теперь выберем событие B , никак не связанное с полом, например такое: фамилия выбранного человека начинается на букву «К». Предположим, что людей с фамилией на букву «К» всего 5 %, откуда следует $\mathbf{P}(B) = 5/100 = 1/20$. Для вычисления $\mathbf{P}(AB)$ мы должны взять $1/20$ долю от половины всей группы, т. е. $\mathbf{P}(AB) = 1/20 \cdot 1/2 = \mathbf{P}(B) \cdot \mathbf{P}(A)$. С другой стороны, выберем событие C , зависящее от пола выбранного человека, например такое: выбранный человек носит женское имя. События A и C практически не различаются, вероятности $\mathbf{P}(C)$ и $\mathbf{P}(AC)$ равны примерно $1/2$. Однако для вычисления $\mathbf{P}(AC)$ вряд ли стоит брать половину от половины всей группы людей, т. е. $\mathbf{P}(AC) \neq 1/2 \cdot 1/2$.

Замечания

1. Не путать независимые и несовместные события! Несовместные события — это те, которые не имеют общих элементарных исходов. Несовместность является всего лишь свойством взаимного расположения множеств. Независимость — это свойство не только множеств, но и, главным образом, вероятности, т. е. заданной на этих

множествах функции. Более того, если события A и B несовместны, то они чаще всего зависимы, так как из $AB = \emptyset$ следует $\mathbf{P}(AB) = 0$, что может совпадать с $\mathbf{P}(A)\mathbf{P}(B)$ только если хотя бы одно из рассматриваемых событий имеет нулевую вероятность.

2. Если A и B независимы, то независимы также A и \bar{B} , \bar{A} и B , \bar{A} и \bar{B} (т. е. переход к дополнению не портит независимости).

Достаточно доказать только первое из этих утверждений. Оно следует из простых соотношений:

$$\begin{aligned}\mathbf{P}(A\bar{B}) &= \mathbf{P}(A \setminus AB) = \mathbf{P}(A) - \mathbf{P}(AB) = \mathbf{P}(A) - \mathbf{P}(A)\mathbf{P}(B) = \\ &= \mathbf{P}(A)(1 - \mathbf{P}(B)) = \mathbf{P}(A)\mathbf{P}(\bar{B}).\end{aligned}$$

3. Данное выше определение независимых событий можно распространить на случай любого количества n событий.

Определение. События A_1, A_2, \dots, A_n называются *независимыми в совокупности*, если для любого подмножества индексов

$$\{i_1, i_2, \dots, i_k\} \subset \{1, 2, \dots, n\}, \quad 2 \leq k \leq n,$$

выполняется

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

К сожалению, попарной независимости событий недостаточно для того, чтобы указанное свойство выполнялось при $k > 2$, точно так же, как выполнение этого свойства при $k = n$ не гарантирует его справедливости при меньших значениях k .

1.5. Схема Бернулли

Рассмотрим несколько задач, приводящих к одной и той же модели.

Задача 1. Известно, что вероятность рождения мальчика равна 0.515, девочки — 0.485. Некоторая супружеская пара запланировала иметь 10 детей. Какова вероятность, что мальчиков и девочек родится поровну?

Задача 2. Стрелок в тире попадает в цель с вероятностью p и промахивается с вероятностью $q = 1 - p$. Какова вероятность, что произойдет ровно k попаданий за n выстрелов? Здесь k может принимать любые значения от 0 до n .

Задача 3. Изготовлено n изделий, причем каждое из них независимо от других оказывается бракованным с вероятностью p . С какой вероятностью при проверке на пригодность будет обнаружено k бракованных изделий?

Выделим общие черты этих задач:

- 1) в каждой из них имеется некоторое количество n независимых испытаний;
- 2) каждое испытание может завершиться одним из двух возможных исходов, назовем их условно «успех» и «неуспех»;
- 3) вероятность «успеха» не меняется от испытания к испытанию и равна p .

Обозначим S_n число успехов, реализовавшихся в n испытаниях. Вопрос стоит об отыскании $\mathbf{P}(S_n = k)$ при $0 \leq k \leq n$.

Чтобы решить эту задачу, нужно сначала построить вероятностную модель.

Начнем с описания пространства элементарных исходов. Будем писать «У», если произошел успех в испытании, и «Н» в случае неуспеха. Тогда исходами эксперимента, состоящего из n испытаний, будут всевозможные последовательности длины n , у которых на каждом месте стоит один из этих двух символов. Всего таких последовательностей 2^n . Таким образом, пространство элементарных исходов является

дискретным; более того, оно конечно. Мы знаем общее правило нахождения вероятности события в дискретном пространстве; чтобы воспользоваться им, нам необходимо сначала для каждого элементарного исхода задать его вероятность.

Возьмем конкретный элементарный исход, т. е. цепочку длины n , состоящую из символов «У» и «Н», причем будем предполагать, что «У» встречается в ней k раз. Именно такие исходы формируют интересующее нас событие в задаче. Например, возьмем такую цепочку: $\omega = \langle \text{УНН...Н} \rangle$.

Для понимания того, какой должна быть вероятность такого элементарного исхода, введем n вспомогательных событий B_1, B_2, \dots, B_n , причем мы их будем строить, глядя на выбранную нами конкретную цепочку. Пусть B_1 состоит из цепочек, у которых на первом месте стоит «У», а на остальных местах может стоять что угодно, B_2 состоит из цепочек, у которых на втором месте стоит «Н», B_3 — из цепочек, у которых на третьем месте стоит «Н», и т. д. — все, как у выбранного нами элементарного исхода ω .

Введенные события должны быть независимыми в нашей модели, потому, что они независимы по условию эксперимента, так как B_1 относится только к первому испытанию, B_2 — только ко второму и т. д., а разные испытания не влияют друг на друга.

Таким образом, в нашей модели должно выполняться

$$\mathbf{P}(B_1 B_2 \dots B_n) = \mathbf{P}(B_1) \mathbf{P}(B_2) \dots \mathbf{P}(B_n).$$

Но, следуя построению, событие $B_1 B_2 \dots B_n$ состоит из одного единственного элементарного исхода $\omega = \langle \text{УНН...Н} \rangle$. С другой стороны,

$$\mathbf{P}(B_1) = p, \quad \mathbf{P}(B_2) = 1 - p, \quad \mathbf{P}(B_3) = 1 - p$$

и т. д. Поэтому для данного элементарного исхода ω должно выполняться

$$\mathbf{P}(\omega) = \mathbf{P}(B_1) \dots \mathbf{P}(B_n) = p^k q^{n-k},$$

где $q = 1 - p$, k — число успехов.

Задав вероятности элементарных исходов, мы завершили построение вероятностной модели. Она и называется *схемой Бернулли*.

Ясно, что элементарные исходы будут равновероятными только при $p = q = 1/2$. В этом случае каждый элементарный исход будет иметь вероятность $1/2^n$, и только в этом случае мы вправе использовать классическое определение вероятности.

Для нахождения вероятности того, что успехов будет ровно k в n испытаниях, мы должны просуммировать вероятности всех элементарных исходов, у которых успех встречается k раз, а таких исходов, как нетрудно видеть, C_n^k . Поэтому

$$\mathbf{P}(S_n = k) = p^k q^{n-k} + p^k q^{n-k} + \dots + p^k q^{n-k} = C_n^k p^k q^{n-k}.$$

Совокупность чисел $\{C_n^k p^k q^{n-k}, k = 0, 1, \dots, n\}$ называется *биномиальным распределением* (поскольку $(p + q)^n = 1 = \sum_{k=0}^n C_n^k p^k q^{n-k}$ — разложение по биному Ньютона).

В заключение этого параграфа выясним, при каком k вероятность $\mathbf{P}(S_n = k)$ максимальна. Для этих целей рассмотрим отношение

$$\alpha_k = \frac{\mathbf{P}(S_n = k + 1)}{\mathbf{P}(S_n = k)}$$

и выясним, при каких k имеет место $\alpha_k \geq 1$ (это будет означать неубывание $\mathbf{P}(S_n = k)$ при возрастании k) и при каких значениях k выполняется $\alpha_k \leq 1$, что соответствует невозрастанию вероятностей. На этом пути и отыщем точку максимума. Итак, запишем неравенство

$$\alpha_k = \frac{C_n^{k+1} p^{k+1} q^{n-k-1}}{C_n^k p^k q^{n-k}} = \frac{(n-k)p}{(k+1)q} \geq 1,$$

что эквивалентно $np - q \geq k(p + q) = k$. Таким образом, если $k \leq np - q$, то

$$\mathbf{P}(S_n = k + 1) \geq \mathbf{P}(S_n = k),$$

т. е. вероятности возрастают (вернее, не убывают), и, наоборот, при $k \geq np - q$ вероятности не возрастают. Поскольку число $np - q$ не обязано быть целым, то нетрудно видеть, что максимальное значение для $\mathbf{P}(S_n = k)$ будет достигаться при $k = [np - q] + 1 = [(n + 1)p]$.

1.6. Условные вероятности

Пусть A — событие, вероятностью которого мы интересуемся. Оно может произойти в результате некоторого эксперимента; мы не знаем точно, как эксперимент завершился, однако определенной информацией уже располагаем: нам сообщили, что некоторое другое событие B уже произошло.

Какова же теперь вероятность события A ?

Ясно, что знание того, что B произошло, может сильно повлиять на результат. Например, при бросании игральной кости вероятность того, что выпала шестерка (событие A), равна $1/6$. Однако если заранее известно, что выпало четное число очков (событие B), то следует ожидать, что шестерка выпадает уже с вероятностью $1/3$.

Тем самым мы приходим к необходимости введения нового понятия.

Определение. *Условной вероятностью* (или: вероятностью события A при условии, что B произошло) называется

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

Вернёмся к примеру с игральной костью. Имеем здесь

$$A = \{6\}, \quad B = \{2, 4, 6\}, \quad AB = \{6\},$$

$$\mathbf{P}(A/B) = \frac{1/6}{1/2} = \frac{1}{3},$$

что соответствует нашим интуитивным представлениям.

Сделаем несколько замечаний в связи с данным определением.

1. Поскольку $\mathbf{P}(B)$ стоит в знаменателе, то необходимо всегда требовать, чтобы $\mathbf{P}(B) > 0$. Условная вероятность не вводится, если $\mathbf{P}(B) = 0$.

2. Если $\mathbf{P}(B) > 0$, то независимость событий A и B эквивалентна условию $\mathbf{P}(A) = \mathbf{P}(A/B)$ — это очевидно. В общем, так оно и должно быть: событие B (условие) никак не должно влиять на A , если A и B независимы.

3. Если в результате эксперимента событие B уже произошло, то ни один из элементарных исходов из дополнительного множества \bar{B} уже реализоваться не может. Таким образом, пространство возможных исходов эксперимента сужается до размеров множества B . Отражением этого факта и является формула для $\mathbf{P}(A/B)$. В ней множитель $1/\mathbf{P}(B)$ выполняет нормирующую роль: суммарная вероятность всех возможных теперь исходов должна равняться единице. А использование в числителе вероятности пересечения множеств A и B соответствует тому, что из элементарных

исходов, входящих в A , произойти теперь могут только те, которые входят одновременно и в B .

Сказанное можно проследить на примере бросания наугад точки, скажем, в квадрат Ω . Пусть $\lambda(A)$ — площадь множества $A \subset \Omega$. Тогда

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

— отношение площадей. Для условной вероятности имеем

$$\mathbf{P}(A/B) = \frac{\mathbf{P}(AB)}{\mathbf{P}(B)} = \frac{\lambda(AB)/\lambda(\Omega)}{\lambda(B)/\lambda(\Omega)} = \frac{\lambda(AB)}{\lambda(B)},$$

это тоже отношение площадей, но только роль всего пространства исходов выполняет событие B .

1.7. Формула полной вероятности

Пусть нас интересует вероятность некоторого события A и предположим, что наряду с A есть некий набор вспомогательных событий H_1, H_2, \dots, H_n , которые принято называть *гипотезами* и которые удовлетворяют следующим двум требованиям.

1) $H_i H_j = \emptyset$ ($i \neq j$);

2) $A \subset \bigcup_{i=1}^n H_i$.

Тогда справедлива *формула полной вероятности*

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

Доказательство.

$$\mathbf{P}(A) = \mathbf{P}\left(A \cap \bigcup_{i=1}^n H_i\right) = \mathbf{P}\left(\bigcup_{i=1}^n AH_i\right) = \sum_{i=1}^n \mathbf{P}(AH_i).$$

Поскольку $\mathbf{P}(A/H_i) = \mathbf{P}(AH_i)/\mathbf{P}(H_i)$ (предполагаем, что $\mathbf{P}(H_i) > 0$, нет смысла использовать гипотезы с нулевой вероятностью), то остается выразить отсюда $\mathbf{P}(AH_i)$ и подставить в формулу.

Число используемых гипотез может быть и бесконечным — это ничему не противоречит.

Формула полной вероятности обычно бывает полезна при решении задач, где имеет место «двойная» (или «двухуровневая») случайность. С помощью этой формулы мы фиксируем сначала ситуацию на одном уровне (т. е. считаем, что одна из гипотез реализовалась) и перебираем все возникающие при этом возможности на другом уровне; затем ведем перебор возможностей первого уровня — это соответствует суммированию по переменной i .

Пример. На предприятии работает n рабочих, которые делают одинаковые изделия. За смену первый изготовил k_1 изделий, второй — k_2, \dots, n -й рабочий изготовил k_n изделий. Обозначим $k = k_1 + k_2 + \dots + k_n$ — общее количество деталей, изготовленных за смену.

Известно, что изделие, изготовленное первым рабочим, с вероятностью p_1 оказывается бракованным, для второго рабочего вероятность брака равна p_2 и т. д.

В конце смены все изделия ссыпали в один бункер. Какова вероятность, что наугад выбранное из бункера изделие окажется бракованным?

Обозначим через A событие, вероятностью которого мы интересуемся. Задача была бы тривиальной, если бы мы знали, кем выбранное изделие изготовлено. А так как мы не знаем, то строим облегчающие предположения (гипотезы). Пусть событие H_i означает, что выбранное нами изделие изготовлено i -м рабочим, $i = 1, 2, \dots, n$. Ясно, что любое событие из H_1, H_2, \dots, H_n исключает другие. Кроме того,

$$\bigcup_{i=1}^n H_i = \Omega \supset A.$$

Тем самым выполнены все требования, предъявляемые к гипотезам. С помощью классического определения вероятности находим

$$\mathbf{P}(H_i) = \frac{C_{k_i}^1}{C_k^1} = \frac{k_i}{k}.$$

Если же известно, что изделие изготовлено i -м рабочим, то вероятность, что оно является бракованным, равна $\mathbf{P}(A/H_i) = p_i$ по условию задачи. Тем самым получаем по формуле полной вероятности

$$\mathbf{P}(A) = \sum_{i=1}^n p_i \frac{k_i}{k}.$$

1.8. Формула Байеса

Формула Байеса используется в той же ситуации, что и формула полной вероятности, т. е. если имеется событие A и набор гипотез H_1, H_2, \dots, H_n , удовлетворяющих указанным выше требованиям.

Вероятности гипотез $\mathbf{P}(H_1), \mathbf{P}(H_2), \dots, \mathbf{P}(H_n)$ принято называть *априорными*, т. е. изначальными, доопытными. Если же событие A уже произошло, то условные вероятности гипотез $\mathbf{P}(H_1/A), \mathbf{P}(H_2/A), \dots, \mathbf{P}(H_n/A)$ могут сильно отличаться от априорных и называются *апостериорными*, т. е. послеопытными, учитывающими результаты эксперимента.

Для многих практических целей бывает полезно находить апостериорные вероятности гипотез, и делается это с помощью *формулы Байеса*. Она состоит в следующем: для любого $i = 1, \dots, n$

$$\mathbf{P}(H_i/A) = \frac{\mathbf{P}(A/H_i)\mathbf{P}(H_i)}{\sum_{j=1}^n \mathbf{P}(A/H_j)\mathbf{P}(H_j)}.$$

Для доказательства достаточно воспользоваться формулами

$$\mathbf{P}(H_i/A) = \frac{\mathbf{P}(H_i A)}{\mathbf{P}(A)}, \quad \mathbf{P}(H_i A) = \mathbf{P}(A/H_i)\mathbf{P}(H_i)$$

и уже полученной формулой полной вероятности для $\mathbf{P}(A)$.

Вернемся к предыдущему примеру. Представим себе, что взятое наугад изделие оказалось бракованным. Какова вероятность, что его изготовил i -й рабочий? По формуле Байеса получаем

$$\mathbf{P}(H_i/A) = \frac{p_i \frac{k_i}{k}}{\sum_{j=1}^n p_j \frac{k_j}{k}}.$$

2. Распределения

2.1. Случайные величины. Функции распределения

При изучении тех или иных случайных явлений или экспериментов нас интересует, как правило, не какой именно исход реализовался, а та или иная числовая характеристика этого исхода. Например, в схеме Бернулли нам не так уж важно было, какая цепочка символов реализовалась, интерес вызывало только число успехов в этой цепочке. Точно так же при стрельбе по плоской мишени мы не интересуемся точными координатами центра пробойны. Для нас важно, сколько очков мы выбили при стрельбе.

Это наводит на необходимость введения понятия случайной величины.

Определение. *Случайной величиной* X называется произвольная функция, заданная на пространстве элементарных исходов Ω и принимающая значения в \mathbb{R} (значения $\pm\infty$ исключаются), т. е. каждому элементарному исходу ω ставится в соответствие число $X(\omega) \in \mathbb{R}$.

Например, число успехов в n испытаниях Бернулли, которое мы обозначали S_n , является случайной величиной.

По-другому можно сказать, что случайная величина — это числовая переменная, значение которой меняется в зависимости от того, какой исход реализовался в результате эксперимента.

Изучение различных случайных величин — одна из основных задач теории вероятностей. В то же время следует заметить, что для изучения функций, заданных на произвольном множестве (в данном случае Ω), не существует достаточно развитого математического аппарата. В связи с этим во многих ситуациях ограничиваются изучением не самих случайных величин, а их распределений.

Определение. Мы будем говорить, что нам известно *распределение* случайной величины X , если для произвольных чисел $a \leq b$ мы можем находить вероятности вида $\mathbf{P}(\omega : a \leq X(\omega) \leq b)$ (а значит, и вероятности вида $\mathbf{P}(\omega : a \leq X(\omega) < b)$, $\mathbf{P}(\omega : a < X(\omega) \leq b)$, $\mathbf{P}(\omega : a < X(\omega) < b)$).

В дальнейшем будем использовать краткую запись:

$$\mathbf{P}(\omega : a \leq X(\omega) \leq b) = \mathbf{P}(a \leq X \leq b) = \mathbf{P}(X \in [a, b]).$$

Далее мы увидим следующее: что для того, чтобы знать распределение случайной величины X , достаточно знать всего одну функцию — функцию распределения этой случайной величины.

Определение. *Функцией распределения* случайной величины X называется

$$F_X(y) = \mathbf{P}(\omega : X(\omega) < y) = \mathbf{P}(X < y), \quad -\infty < y < \infty.$$

Основные свойства функций распределения

1. $0 \leq F_X(y) \leq 1$ для всех значений y . Свойство очевидно.
2. Если $y_1 < y_2$, то $F_X(y_1) \leq F_X(y_2)$, т. е. функция распределения монотонно не убывает.

Доказательство. Введем события $A_1 = \{X < y_1\}$, $A_2 = \{X < y_2\}$, тогда $A_1 \subset A_2$, поэтому $F_X(y_1) = \mathbf{P}(A_1) \leq \mathbf{P}(A_2) = F_X(y_2)$.

3. Существуют пределы $\lim_{y \rightarrow -\infty} F_X(y) = 0$ и $\lim_{y \rightarrow \infty} F_X(y) = 1$.

Доказательство. Существование пределов следует из монотонности и ограниченности функции распределения. Чтобы найти значения пределов, достаточно вместо

непрерывно меняющейся переменной y рассмотреть какую-нибудь последовательность $y_k \rightarrow -\infty$ в первом случае и $y_k \rightarrow \infty$ во втором.

Пусть последовательность $\{y_k\}$, монотонно убывая, стремится к $-\infty$ (например, можно взять $y_k = -k$). Введем события

$$A_k = \{X < y_k\}, \quad k = 1, 2, \dots$$

Нетрудно видеть, что

$$A_1 \supset A_2 \supset \dots$$

Используя свойство непрерывности вероятности, получаем

$$\lim_{k \rightarrow \infty} F_X(y_k) = \lim_{k \rightarrow \infty} \mathbf{P}(A_k) = \mathbf{P}\left(\bigcap_{k=1}^{\infty} A_k\right) \stackrel{?}{=} \mathbf{P}(\emptyset) = 0.$$

Равенство, помеченное вопросом, требует комментариев. Докажем, что указанное пересечение множеств пусто. От противного: предположим, что существует хотя бы один элементарный исход $\omega \in \bigcap A_k$. Поскольку $X(\omega)$ — конечное число, то существует индекс k_0 такой, что $y_{k_0} < X(\omega)$, т. е. $\omega \notin A_{k_0}$, а значит, $\omega \notin \bigcap A_k$, что противоречит исходному предположению.

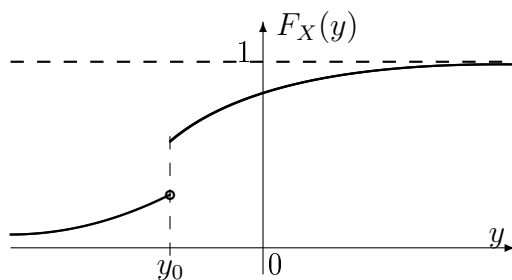
Для доказательства второго предельного соотношения рассмотрим последовательность чисел y_k , монотонно стремящуюся к бесконечности (например, $y_k = k$), и введем события $A_k = \{X < y_k\}$, $k = 1, 2, \dots$. Очевидно, $A_1 \subset A_2 \subset \dots$; и опять в силу свойства непрерывности вероятности

$$\lim_{k \rightarrow \infty} F_X(y_k) = \lim_{k \rightarrow \infty} \mathbf{P}(A_k) = \mathbf{P}\left(\bigcup_{k=1}^{\infty} A_k\right) = \mathbf{P}(\Omega) = 1.$$

Поясним, почему здесь $\bigcup_{k=1}^{\infty} A_k = \Omega$. Включение $\bigcup A_k \subset \Omega$ очевидно. Обратно: пусть $\omega \in \Omega$, вычислим $X(\omega)$ — это некоторое конечное число. Поэтому найдется индекс k_0 такой, что $y_{k_0} > X(\omega)$, т. е. $\omega \in A_{k_0} \subset \bigcup_{k=1}^{\infty} A_k$.

Установленные свойства уже позволяют в общих чертах представить себе, как выглядят графики функций распределения. Располагаясь полностью в полосе $0 \leq y \leq 1$ на координатной плоскости точек (x, y) , кривые являют собой неубывающие функции, которые проходят путь по вертикали от 0 до 1 при возрастании аргумента от $-\infty$ до $+\infty$.

Однако путь этот не обязан быть непрерывным: возможны скачки. Например, график может быть таким.



Возникает вопрос: чему равно значение функции распределения в точке разрыва, коль скоро он имеет место? Ответ содержится в следующем свойстве.

4. Для любого y имеет место $F_X(y - 0) = F_X(y)$, т. е. функция распределения всегда непрерывна слева.

Доказательство. Выбираем возрастающую последовательность точек $\{y_k\}$, сходящуюся слева к y (например, $y_k = y - \frac{1}{k}$). Введем события $A_k = \{X < y_k\}$, $k = 1, 2, \dots$. Здесь, очевидно, $A_1 \subset A_2 \subset \dots$, значит,

$$F_X(y - 0) = \lim_{k \rightarrow \infty} F_X(y_k) = \lim_{k \rightarrow \infty} \mathbf{P}(A_k) = \mathbf{P}\left(\bigcup_{k=1}^{\infty} A_k\right) \stackrel{?}{=} \mathbf{P}(X < y) = F_X(y).$$

Поясним равенство, отмеченное вопросом.

Пусть $\omega \in \bigcup A_k$, тогда существует индекс k_0 такой, что $\omega \in A_{k_0}$, т. е. $X(\omega) < y_{k_0} < y$.

В другую сторону: пусть ω таково, что $X(\omega) < y$, тогда существует индекс k_0 такой, что $X(\omega) < y_{k_0}$, т. е. $\omega \in A_{k_0} \subset \bigcup A_k$.

Свойства 1–4, доказанные нами, являются характеристическими для функций распределения в том смысле, что любая функция, ими обладающая, является функцией распределения какой-то случайной величины в подходящем вероятностном пространстве.

Доказательство этого факта выходит за рамки нашего курса.

Отметим еще одно (дополнительное) свойство: мы доказали, что

$$F_X(y - 0) = \mathbf{P}(X < y);$$

оказывается, что

$$F_X(y + 0) = \mathbf{P}(X \leq y).$$

Мы не будем доказывать это соотношение, для этого потребовалось бы вновь (в четвертый раз!) построить нужную последовательность точек и воспользоваться свойством непрерывности вероятности. Отметим только одно полезное следствие этих фактов. Из аддитивности следует, что

$$\mathbf{P}(X \leq y) = \mathbf{P}(X < y) + \mathbf{P}(X = y),$$

откуда

$$\mathbf{P}(X = y) = \mathbf{P}(X \leq y) - \mathbf{P}(X < y) = F_X(y + 0) - F_X(y - 0),$$

что равно величине скачка функции распределения в точке y .

Таким образом, $\mathbf{P}(X = y) = 0$ для всех точек y , в которых функция распределения случайной величины X непрерывна. Далее, для любых чисел $a < b$ можно записать $\{X < b\} = \{X < a\} \cup \{a \leq X < b\}$, то

$$\mathbf{P}(X < b) = \mathbf{P}(X < a) + \mathbf{P}(a \leq X < b),$$

поэтому

$$\mathbf{P}(a \leq X < b) = \mathbf{P}(X < b) - \mathbf{P}(X < a) = F_X(b) - F_X(a).$$

Точно так же

$$\begin{aligned} \mathbf{P}(a \leq X \leq b) &= \mathbf{P}(X \leq b) - \mathbf{P}(X < a) = F_X(b + 0) - F_X(a); \\ \mathbf{P}(a < X \leq b) &= \mathbf{P}(X \leq b) - \mathbf{P}(X \leq a) = F_X(b + 0) - F_X(a + 0); \\ \mathbf{P}(a < X < b) &= \mathbf{P}(X < b) - \mathbf{P}(X \leq a) = F_X(b) - F_X(a + 0). \end{aligned}$$

Тем самым мы подтвердили высказанное ранее утверждение о том, что функция распределения полностью определяет распределение случайной величины. В связи с этим термины «распределение» и «функция распределения» (а также «закон распределения») часто используются как синонимы.

2.2. Типы распределений. Примеры

Определение. Случайная величина X называется *дискретной*, если существует конечная или счетная последовательность чисел y_1, y_2, y_3, \dots такая, что

$$\sum_{k=1}^{\infty} \mathbf{P}(X = y_k) = 1.$$

Функция распределения дискретной случайной величины называется *дискретной*.

Дискретное распределение удобно задавать с помощью таблицы. Обозначим

$$p_k = \mathbf{P}(X = y_k), \quad k = 1, 2, \dots,$$

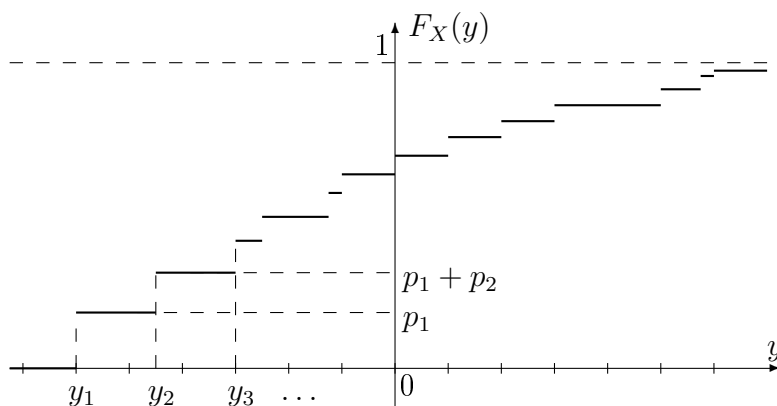
тогда приведенная ниже таблица полностью характеризует распределение.

Значения	y_1	y_2	y_3	\dots
Вероятности	p_1	p_2	p_3	\dots

Например, вероятности попадания значений случайной величины в интервал легко находятся суммированием элементов таблицы:

$$\mathbf{P}(a < X < b) = \sum_{k: a < y_k < b} p_k.$$

Пусть значения случайной величины y_1, y_2, y_3, \dots пронумерованы в порядке их возрастания. Тогда график функции распределения будет выглядеть примерно так:

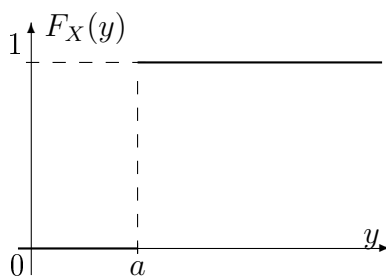


Действительно, если $y < y_1$, то $F_X(y) = \mathbf{P}(X < y) = 0$; если $y_1 < y < y_2$, то $F_X(y) = \mathbf{P}(X < y) = \mathbf{P}(X = y_1) = p_1$, и т. д.

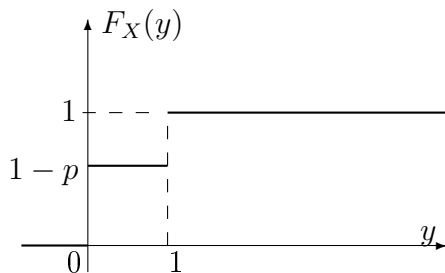
В дальнейшем мы будем писать $X \in F$, если X имеет функцию распределения F .

Примеры дискретных распределений

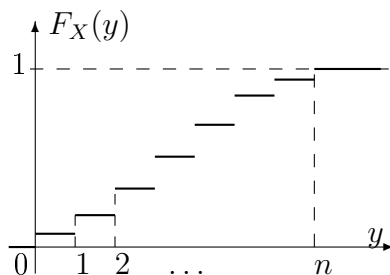
1. *Вырожденное распределение* $I_a : X \in I_a$, если $\mathbf{P}(X = a) = 1$.



2. *Распределение Бернулли* B_p : $X \in B_p$, если $\mathbf{P}(X = 1) = p$, $\mathbf{P}(X = 0) = 1 - p$, $0 < p < 1$.

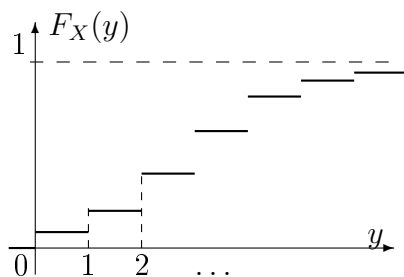


3. *Биномиальное распределение* $B_{n,p}$: $X \in B_{n,p}$, если $\mathbf{P}(X = k) = C_n^k p^k (1 - p)^{n-k}$, $k = 0, 1, \dots, n$ (в частности, $B_{1,p} = B_p$).



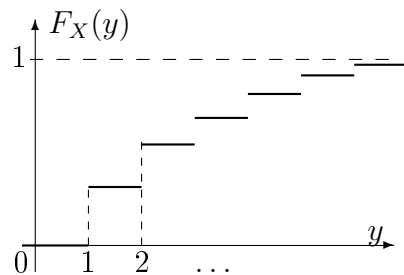
Биномиальное распределение, как мы уже видели, возникает при рассмотрении схемы Бернулли — это распределение числа успехов в n испытаниях.

4. *Распределение Пуассона* Π_λ : $X \in \Pi_\lambda$, если $\mathbf{P}(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$, $k = 0, 1, 2, \dots$; $\lambda > 0$.



Распределение Пуассона может использоваться при описании числа клиентов, поступивших в течение определенного времени в систему обслуживания, числа частиц, зарегистрированных прибором, числа особей биологической популяции и т. д.

5. *Геометрическое распределение* G_p : $X \in G_p$, если $\mathbf{P}(X = k) = (1 - p) p^{k-1}$, $k = 1, 2, 3, \dots$, $0 < p < 1$.



Если в схеме Бернулли производить испытания до первого получения неуспеха включительно, то количество требуемых для этого испытаний будет случайной величиной, имеющей геометрическое распределение.

Данное распределение может встретиться и в другом варианте:

$$\mathbf{P}(X = k) = (1 - p) p^k, \quad k = 0, 1, 2, 3, \dots$$

Далее рассмотрим другой тип распределений.

Определение. Функция распределения $F_X(y)$ называется *абсолютно непрерывной*, если для любого значения y

$$F_X(y) = \int_{-\infty}^y f(t) dt;$$

стоящая под знаком интеграла функция $f(t)$ называется *плотностью* распределения.

Чтобы подчеркнуть, что плотность относится к случайной величине X , ее также снабжают индексом $f(t) = f_X(t)$.

Требование абсолютной непрерывности является более сильным, нежели просто непрерывность. Из определения абсолютно непрерывной функции распределения вытекает, что $F_X(y)$ почти всюду имеет производную (в некоторых точках производная может не существовать, хотя непрерывность сохраняется). Поскольку функция распределения есть интеграл от плотности, то плотность, в свою очередь, равна производной функции распределения

$$f_X(t) = \frac{dF_X(t)}{dt}$$

и это соотношение выполняется для всех точек, где производная существует.

Поскольку абсолютно непрерывная функция распределения не имеет скачков, то $\mathbf{P}(X = y) = 0$ для любого y и совпадают, к примеру, вероятности $\mathbf{P}(X \in [a, b])$ и $\mathbf{P}(X \in (a, b))$, $a < b$.

Свойства плотности

1) $f_X(t) \geq 0$ — как производная неубывающей функции;

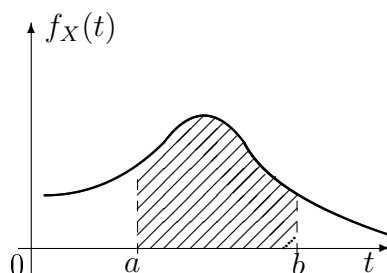
2) $\int_{-\infty}^{\infty} f_X(t) dt = 1$.

Для доказательства последнего достаточно устремить $y \rightarrow \infty$ в определении абсолютно непрерывной функции распределения.

Любая функция $f(t)$, обладающая этими двумя свойствами, может быть плотностью распределения.

Отметим еще одно важное свойство плотностей. Для любых чисел $a < b$

$$\mathbf{P}(a \leq X < b) = F_X(b) - F_X(a) = \int_{-\infty}^b f_X(t) dt - \int_{-\infty}^a f_X(t) dt = \int_a^b f_X(t) dt.$$



Таким образом, плотность есть неотрицательная интегрируемая функция, площадь под графиком которой равна единице. Если вообразить опять, что вероятность — это масса, то суммарная масса значений случайной величины равна единице. Эти значения разбросаны (или, лучше сказать, размазаны) по вещественной прямой и график плотности показывает нам толщину получившегося «бутерброда». Вероятность того, что значения случайной величины попадают в промежуток $[a, b]$, равна площади под графиком плотности, приходящейся на отрезок $[a, b]$. В нашей интерпретации данная вероятность — это масса «бутерброда» с основанием $[a, b]$.

Вообще, если множество $B \subset \mathbb{R}$ допускает возможность интегрирования по нему, то

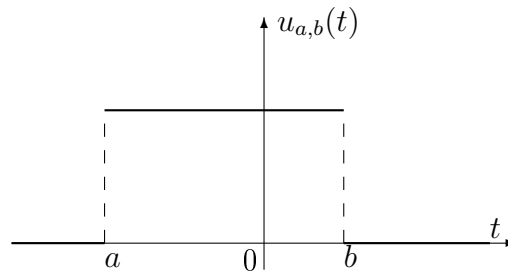
$$\mathbf{P}(X \in B) = \int_B f_X(t) dt.$$

Примеры абсолютно непрерывных распределений

Здесь мы используем заглавные буквы для обозначения функций распределения, а соответствующие малые буквы — для обозначения плотностей.

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

$$u_{a,b}(t) = \begin{cases} \frac{1}{b-a}, & t \in [a, b], \\ 0, & \text{иначе.} \end{cases},$$



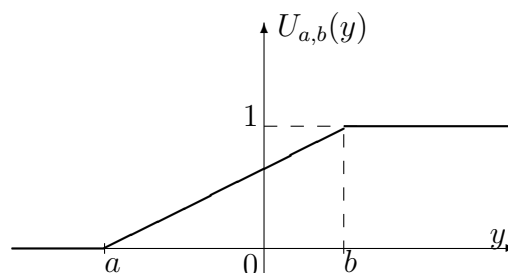
Ясно, что в данном случае все значения случайной величины располагаются на отрезке $[a, b]$ и равномерно там разбросаны; вероятность попадания в любой промежуток $[c, d] \subset [a, b]$ равна отношению длин

$$\mathbf{P}(X \in [c, d]) = \int_c^d \frac{1}{b-a} dt = \frac{d-c}{b-a},$$

что уже встречалось нам в задачах на геометрические вероятности.

Для функции распределения имеем формулу

$$U_{a,b}(y) = \begin{cases} 0, & y \leq a, \\ \frac{y-a}{b-a}, & y \in [a, b], \\ 1, & y > b. \end{cases}$$

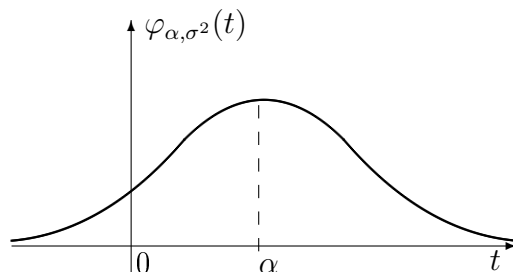


Как видим, в двух точках эта функция производной не имеет.

2. *Нормальное (гауссовское) распределение* Φ_{α, σ^2} . Плотность задается формулой

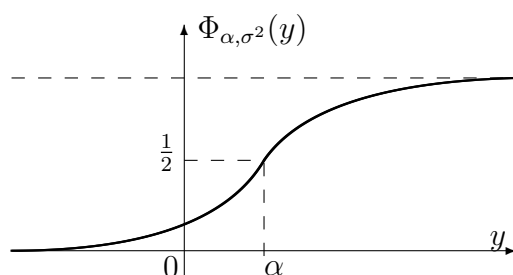
$$\varphi_{\alpha, \sigma^2}(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(t-\alpha)^2/2\sigma^2}, \quad -\infty < t < \infty.$$

Здесь α — параметр сдвига, $-\infty < \alpha < \infty$, другой параметр $\sigma^2 > 0$ отвечает за угол развала ветвей графика плотности и за максимальное значение этой функции.



Функция распределения задается формулой (к сожалению, интеграл не берется в элементарных функциях)

$$\Phi_{\alpha, \sigma^2}(y) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^y e^{-\frac{(t-\alpha)^2}{2\sigma^2}} dt.$$



Если $\alpha = 0$, $\sigma^2 = 1$, то мы получаем *стандартное нормальное распределение* $\Phi_{0,1}$ с плотностью

$$\varphi_{0,1}(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}$$

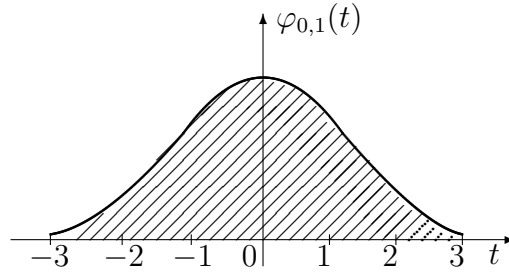
и с функцией распределения

$$\Phi_{0,1}(y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^y e^{-t^2/2} dt.$$

График этой функции имеет центр симметрии — точку с координатами $(0, 1/2)$, $\Phi_{0,1}(y) = 1 - \Phi_{0,1}(-y)$. Функция $\Phi_{0,1}(y)$ очень быстро стремится к нулю при $y \rightarrow -\infty$ (и соответственно так же быстро к единице при $y \rightarrow \infty$):

$$\Phi_{0,1}(-3) = 0.00135; \quad \Phi_{0,1}(-1.96) = 0.025; \quad \Phi_{0,1}(-1.64) = 0.05.$$

Эти данные взяты из таблиц стандартного нормального распределения, которыми снабжены почти все пособия по теории вероятностей и математической статистике ввиду важности этого распределения для приложений. Несмотря на то что значения случайной величины $Y \in \Phi_{0,1}$ разбросаны по всей прямой, видно, что с вероятностью 0.9973 они попадают в интервал $(-3, 3)$.



Позже мы покажем, что если $X \in \Phi_{\alpha, \sigma^2}$, то $Y = (X - \alpha)/\sigma \in \Phi_{0,1}$. Значит, $\mathbf{P}(|Y| < 3) = 0.9973$, или, что то же самое,

$$\mathbf{P}(|X - \alpha| < 3\sigma) = 0.9973.$$

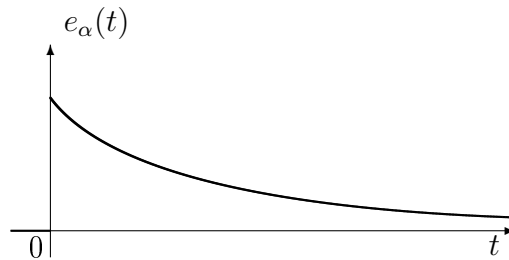
Последнее известно как *правило трех сигм*.

Насколько важно нормальное распределение для приложений, станет ясно позже, когда будет изучена центральная предельная теорема. Забегая вперед, скажем, что очень часто распределение случайной величины будет близко к нормальному, если она сформировалась в результате накопления большого числа более «мелких» случайных факторов.

3. *Показательное (экспоненциальное) распределение* E_α . Плотность показательного распределения задается формулой

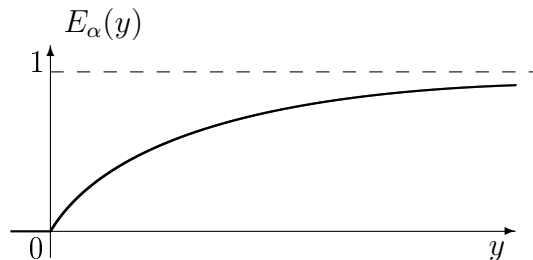
$$e_\alpha(t) = \begin{cases} \alpha e^{-\alpha t}, & t > 0, \\ 0, & t \leq 0. \end{cases}$$

Здесь $\alpha > 0$ — параметр распределения.



Функция распределения легко получается интегрированием:

$$E_\alpha(y) = \begin{cases} 0, & y \leq 0, \\ 1 - e^{-\alpha y}, & y > 0. \end{cases}$$



Показательно распределенными оказываются длительности телефонных разговоров, промежутки времени между последовательными приходами клиентов на обслуживание (например, кораблей в порт или покупателей в магазин), длительности обслуживания клиентов, время безотказной работы прибора и многое другое.

Остановимся более подробно на одном замечательном свойстве показательного распределения. Пусть X — продолжительность телефонного разговора и пусть $X \in E_\alpha$, т. е.

$$\mathbf{P}(X \geq y) = e^{-\alpha y}, \quad y > 0.$$

Телефонный разговор начался в момент времени 0 и, когда в момент времени y мы решили подключиться к нему (с неблагоприятной целью подслушивания), он все еще продолжался, то есть $X \geq y$. Каково будет распределение у оставшейся продолжительности разговора $X - y$? Оказывается, в точности такое же, как и у всей продолжительности X . Действительно, вероятность того, что оставшаяся длительность разговора будет не меньше t , равна

$$\begin{aligned} \mathbf{P}(X - y \geq t / X \geq y) &= \frac{\mathbf{P}(X \geq y + t, X \geq y)}{\mathbf{P}(X \geq y)} = \frac{\mathbf{P}(X \geq y + t)}{\mathbf{P}(X \geq y)} = \\ &= \frac{e^{-\alpha(y+t)}}{e^{-\alpha y}} = e^{-\alpha t} = \mathbf{P}(X \geq t), \quad t > 0. \end{aligned}$$

4. *Гамма-распределение* $\Gamma_{\alpha, \lambda}$. Плотность гамма-распределения равна

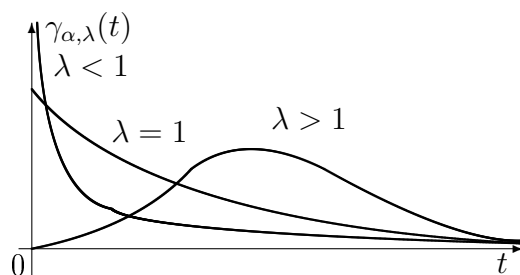
$$\gamma_{\alpha, \lambda}(t) = \begin{cases} \frac{\alpha^\lambda}{\Gamma(\lambda)} t^{\lambda-1} e^{-\alpha t}, & t > 0, \\ 0, & t \leq 0. \end{cases}$$

Здесь участвуют два параметра $\alpha > 0$, $\lambda > 0$. Напомним, что

$$\Gamma(\lambda) = \int_0^\infty t^{\lambda-1} e^{-t} dt$$

— известная гамма-функция Эйлера; она обладает свойством $\Gamma(\lambda + 1) = \lambda \Gamma(\lambda)$. Для целых значений $\lambda = n$ имеет место по этой причине $\Gamma(n + 1) = n!$.

Графики плотности гамма-распределения существенно различаются в зависимости от значений параметра λ (см. рисунок). При $\lambda < 1$ плотность неограниченна в окрестности нуля, при $\lambda = 1$ получается плотность показательного распределения ($\Gamma_{\alpha, 1} = E_\alpha$). При $\lambda > 1$ график плотности имеет одну вершину, которая удаляется вправо с увеличением λ .



Функция гамма-распределения задается формулой

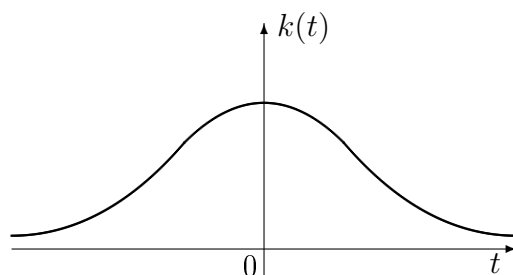
$$\Gamma_{\alpha, \lambda}(y) = \int_0^y \gamma_{\alpha, \lambda}(t) dt$$

при $y > 0$ и $\Gamma_{\alpha, \lambda}(y) = 0$ при $y \leq 0$. Этот интеграл можно вычислить с помощью неоднократного интегрирования по частям, если λ целое, и не берется в элементарных функциях при прочих λ .

Гамма-распределение широко используется в теории систем обслуживания, математической статистике, теории надежности.

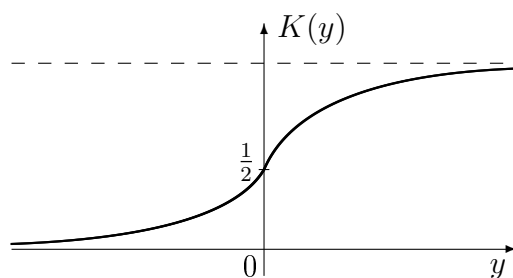
5. *Распределение Коши K*. Плотность задается формулой

$$k(t) = \frac{1}{\pi} \frac{1}{1+t^2}, \quad -\infty < t < \infty.$$



По своему виду график плотности напоминает плотность стандартного нормального распределения, только в отличие от последнего стремление $k(t) \rightarrow 0$ при $|t| \rightarrow \infty$ происходит значительно медленнее. Интегрируя плотность, находим функцию распределения:

$$K(y) = \frac{1}{\pi} \int_{-\infty}^y \frac{1}{1+t^2} dt = \frac{1}{2} + \frac{1}{\pi} \operatorname{arctg} y.$$



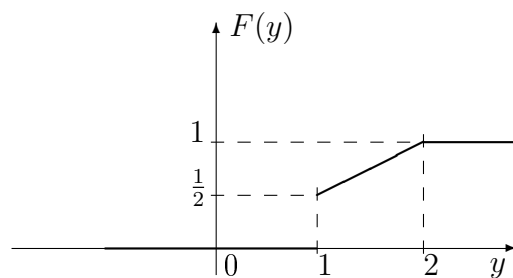
Определение. Функция распределения F относится к *смешанному типу*, если при всех значениях y

$$F(y) = \alpha F_1(y) + \beta F_2(y),$$

где $F_1(y)$ — абсолютно непрерывная, а $F_2(y)$ — дискретная функции распределения, $\alpha \geq 0$, $\beta \geq 0$, $\alpha + \beta = 1$.

Ясно, что частными случаями смешанных распределений являются абсолютно непрерывные (им соответствуют $\alpha = 1$, $\beta = 0$) и дискретные распределения (при $\alpha = 0$, $\beta = 1$).

Пример функции распределения смешанного типа. На рисунке изображен график некоторой функции распределения. Ясно, что эта функция не является дискретной (имеется участок непрерывного роста) и не является абсолютно непрерывной в силу наличия скачка. Это распределение смешанного типа.



Для разложения этой функции распределения на компоненты проще всего выделить сначала дискретную часть: она должна иметь единственный скачок в точке $y = 1$. Берем $F_2(y) = I_1(y)$ (вырожденное распределение в единице), $\beta = 1/2$. Тогда ясно, что $F_1(y) = U_{1,2}(y)$, $\alpha = 1/2$.

Таким образом, поставив первоначальную задачу изучения случайных величин, мы на самом деле стали подробно изучать их распределения. Тем самым произошла некоторая подмена.

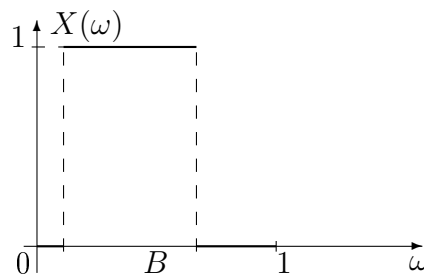
Можно ли утверждать, что распределение однозначно характеризует случайную величину? Оказывается, нет. По случайной величине мы известным образом строим распределение, а вот по распределению восстановить случайную величину невозможно.

Следующий пример показывает, что на одном и том же вероятностном пространстве можно построить бесконечно много различных случайных величин, имеющих одно и то же распределение.

Пример. Пусть $\Omega = [0, 1]$. Для всякого интервала $A \subset \Omega$ положим $\mathbf{P}(A) = \lambda(A)$, где $\lambda(A)$ — длина интервала. Возьмем далее произвольный интервал $B \subset \Omega$, имеющий длину $1/2$, и зададим случайную величину

$$X(\omega) = \begin{cases} 1, & \omega \in B, \\ 0, & \omega \notin B. \end{cases}$$

Она представлена на рисунке.



Ясно, что $X \in B_{1/2}$:

$$\mathbf{P}(X = 1) = \lambda(B) = \frac{1}{2}, \quad \mathbf{P}(X = 0) = \frac{1}{2}.$$

Перемещая множество B внутри Ω , мы будем получать все новые случайные величины, однако все они будут иметь одно и то же распределение $B_{1/2}$.

2.3. Многомерные распределения и плотности

В ряде прикладных задач возникает необходимость рассматривать *случайные векторы*. Мы будем называть случайным всякий вектор $X = (X_1, X_2, \dots, X_n)$, компонентами которого являются случайные величины. Изображаться случайные векторы будут в виде строк или в виде столбцов (как это удобно).

На многомерный случай можно распространить понятие функции распределения.

Определение. Функцией распределения случайного вектора X (многомерной функцией распределения, совместной функцией распределения) называется

$$F_{X_1, X_2, \dots, X_n}(y_1, y_2, \dots, y_n) = \mathbf{P}(X_1 < y_1, X_2 < y_2, \dots, X_n < y_n),$$

где перечисление событий через запятую означает одновременное их осуществление, то есть пересечение.

Свойства многомерных функций распределения

1. $0 \leq F_{X_1, \dots, X_n}(y_1, \dots, y_n) \leq 1$.
2. Если $y_1 \leq z_1, y_2 \leq z_2, \dots, y_n \leq z_n$, то

$$F_{X_1, \dots, X_n}(y_1, \dots, y_n) \leq F_{X_1, \dots, X_n}(z_1, \dots, z_n).$$

Эти два свойства очевидны.

По аналогии со свойствами одномерных функций распределений рассмотрим далее предельное поведение многомерных функций распределения на бесконечности. Но здесь, впрочем, присутствует n аргументов. Мы будем устремлять к $-\infty$ и к $+\infty$ один из них (допустим, последний).

3. а) $\lim_{y_n \rightarrow -\infty} F_{X_1, \dots, X_n}(y_1, \dots, y_n) = 0$,
- б) $\lim_{y_n \rightarrow \infty} F_{X_1, \dots, X_n}(y_1, \dots, y_n) = F_{X_1, \dots, X_{n-1}}(y_1, \dots, y_{n-1})$.

В частности, $F_{X_1}(y_1) = F_{X_1, \dots, X_n}(y_1, \infty, \dots, \infty)$.

Идея доказательства. Если устремить $y_n \rightarrow -\infty$, то событие $\{X_n < y_n\}$ будет уменьшаться до размеров пустого множества и потянет за собой все пересечение $\{X_1 < y_1, X_2 < y_2, \dots, X_n < y_n\}$. Поэтому вероятность этого пересечения будет сходиться к нулю.

Если же $y_n \rightarrow \infty$, то событие $\{X_n < y_n\}$ будет разрастаться до размеров всего пространства Ω , поэтому пересечение событий $\{X_1 < y_1, X_2 < y_2, \dots, X_n < y_n\}$ в пределе превратится в $\{X_1 < y_1, X_2 < y_2, \dots, X_{n-1} < y_{n-1}\}$.

Определение. Случайные величины X_1, X_2, \dots, X_n называются *независимыми*, если для любых $B_1 \subset \mathbb{R}, \dots, B_n \subset \mathbb{R}$ выполняется соотношение

$$\mathbf{P}(X_1 \in B_1, X_2 \in B_2, \dots, X_n \in B_n) = \mathbf{P}(X_1 \in B_1)\mathbf{P}(X_2 \in B_2) \dots \mathbf{P}(X_n \in B_n).$$

Из этого определения вытекает, к примеру, попарная независимость случайных величин: если положить $B_3 = B_4 = \dots = B_n = \mathbb{R}$, то будем иметь

$$\mathbf{P}(X_1 \in B_1, X_2 \in B_2) = \mathbf{P}(X_1 \in B_1)\mathbf{P}(X_2 \in B_2).$$

Если X_1, X_2, \dots, X_n независимы, то

$$F_{X_1, \dots, X_n}(y_1, \dots, y_n) = F_{X_1}(y_1) \dots F_{X_n}(y_n). \quad (1)$$

Это соотношение получается, если в определении независимости положить $B_i = (-\infty, y_i), i = 1, \dots, n$.

Таким образом, для независимых случайных величин введение многомерной функции распределения по существу не дает ничего нового: она выражается через одномерные функции распределения. Для случайного вектора с зависимыми компонентами его функция распределения содержит информацию как о распределении отдельных компонент, так и о зависимости между ними.

Можно показать, что если верно соотношение (1) для всех значений y_1, y_2, \dots, y_n , то X_1, X_2, \dots, X_n независимы.

Доказательство этого факта выходит за рамки нашего курса, однако мы впоследствии будем пользоваться этим утверждением.

Если каждая компонента вектора (X_1, X_2, \dots, X_n) дискретна, то его многомерное распределение также будет называться дискретным. Для дискретного случая

определение независимости случайных величин удобно использовать в следующей эквивалентной форме.

Определение. Дискретные случайные величины X_1, X_2, \dots, X_n независимы, если для всех возможных значений этих случайных величин

$$\mathbf{P}(X_1 = y_1, X_2 = y_2, \dots, X_n = y_n) = \mathbf{P}(X_1 = y_1)\mathbf{P}(X_2 = y_2) \dots \mathbf{P}(X_n = y_n).$$

Дискретное распределение двумерного случайного вектора (X, Y) удобно задавать таблицей. Пусть X принимает возможные значения x_1, x_2, \dots , а Y — значения y_1, y_2, \dots . Обозначим

$$p_{ij} = \mathbf{P}(X = x_i, Y = y_j), \quad i = 1, 2, \dots, \quad j = 1, 2, \dots$$

Приведенная ниже таблица полностью задает распределение вектора (X, Y) .

$X \setminus Y$	y_1	y_2	y_3	\dots
x_1	p_{11}	p_{12}	p_{13}	\dots
x_2	p_{21}	p_{22}	p_{23}	\dots
x_3	p_{31}	p_{32}	p_{33}	\dots
\dots	\dots	\dots	\dots	\dots

Ясно, что

$$\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} p_{ij} = 1.$$

Если суммировать только элементы i -й строки, то получим

$$\sum_{j=1}^{\infty} p_{ij} = \sum_{j=1}^{\infty} \mathbf{P}(X = x_i, Y = y_j) = \mathbf{P}(X = x_i).$$

Точно так же сумма элементов j -го столбца равна

$$\sum_{i=1}^{\infty} p_{ij} = \sum_{i=1}^{\infty} \mathbf{P}(X = x_i, Y = y_j) = \mathbf{P}(Y = y_j).$$

Эти формулы демонстрируют способ получения одномерных распределений из двумерных.

Определение. Функция распределения $F_{X_1, \dots, X_n}(y_1, \dots, y_n)$ называется *абсолютно непрерывной*, если для всех значений аргументов

$$F_{X_1, \dots, X_n}(y_1, \dots, y_n) = \int_{-\infty}^{y_1} \int_{-\infty}^{y_2} \dots \int_{-\infty}^{y_n} f(t_1, t_2, \dots, t_n) dt_n \dots dt_2 dt_1.$$

Подынтегральная функция называется плотностью многомерного распределения, как и в одномерном случае ее принято снабжать индексами, указывающими на связь со случайным вектором: $f(t_1, t_2, \dots, t_n) = f_{X_1, \dots, X_n}(t_1, t_2, \dots, t_n)$. Как и в одномерном случае, плотность получается из функции распределения дифференцированием, только здесь требуется брать частные производные по каждой переменной:

$$f_{X_1, \dots, X_n}(t_1, \dots, t_n) = \frac{\partial^n F_{X_1, \dots, X_n}(t_1, \dots, t_n)}{\partial t_1 \dots \partial t_n}.$$

Свойства многомерных плотностей

1. $f_{X_1, \dots, X_n}(t_1, t_2, \dots, t_n) \geq 0$.

2.
$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f_{X_1, \dots, X_n}(t_1, t_2, \dots, t_n) dt_n \dots dt_1 = 1.$$

3.
$$\mathbf{P}((X_1, X_2, \dots, X_n) \in B) = \int \int \dots \int_B f_{X_1, \dots, X_n}(t_1, t_2, \dots, t_n) dt_n \dots dt_2 dt_1$$

для любого прямоугольника $B = [a_1, b_1] \times [a_2, b_2] \times \dots \times [a_n, b_n] \subset \mathbb{R}^n$.

4. Если случайные величины X_1, X_2, \dots, X_n имеют абсолютно непрерывное совместное распределение, то они независимы тогда и только тогда, когда

$$f_{X_1, \dots, X_n}(t_1, \dots, t_n) = f_{X_1}(t_1) f_{X_2}(t_2) \dots f_{X_n}(t_n).$$

Это свойство получается из формулы (1) поочередным дифференцированием по каждой переменной, а само оно превращается в соотношение (1) после поочередного интегрирования по каждой из переменных.

5. Если известна n -мерная плотность $f_{X_1, \dots, X_n}(t_1, \dots, t_n)$, то получить плотность меньшей размерности можно с помощью интегрирования:

$$f_{X_1, \dots, X_{n-1}}(t_1, t_2, \dots, t_{n-1}) = \int_{-\infty}^{\infty} f_{X_1, \dots, X_n}(t_1, t_2, \dots, t_n) dt_n.$$

Для доказательства этого свойства мы должны представить в виде соответствующего интеграла $(n-1)$ -мерную функцию распределения:

$$\begin{aligned} F_{X_1, \dots, X_{n-1}}(y_1, \dots, y_{n-1}) &= F_{X_1, \dots, X_n}(y_1, \dots, y_{n-1}, \infty) = \\ &= \int_{-\infty}^{y_1} \dots \int_{-\infty}^{y_{n-1}} \left\{ \int_{-\infty}^{\infty} f_{X_1, \dots, X_n}(t_1, \dots, t_n) dt_n \right\} dt_{n-1} \dots dt_1. \end{aligned}$$

Выражение, стоящее в фигурных скобках, и будет искомой плотностью.

В дискретном случае уменьшение размерности производилось аналогично, но только с помощью суммирования (см. рассмотренный выше табличный способ задания двумерных дискретных распределений).

Примеры многомерных плотностей

1. *Многомерное равномерное распределение.* Плотность задается формулой

$$f(t) = \begin{cases} \frac{1}{\lambda(D)}, & t \in D, \\ 0, & \text{иначе,} \end{cases}$$

где $D \subset \mathbb{R}^n$ — ограниченное множество, у которого n -мерный объем $\lambda(D) > 0$. Легко видеть, что для случайного вектора X с такой плотностью

$$\mathbf{P}(X \in B) = \frac{\lambda(B)}{\lambda(D)},$$

если $B \subset D$, т. е. вероятность вычисляется геометрическим способом.

2. Многомерное стандартное нормальное распределение с плотностью

$$f(t) = \frac{1}{(2\pi)^{n/2}} \exp \left\{ -\frac{1}{2} \sum_{i=1}^n t_i^2 \right\} = \prod_{i=1}^n \varphi_{0,1}(t_i), \quad t = (t_1, \dots, t_n).$$

Компоненты случайного вектора, имеющего такую плотность, независимы и имеют стандартное нормальное распределение.

2.4. Преобразования случайных величин

В этом параграфе мы изучим, как изменяются распределения при преобразованиях случайных величин.

Теорема 1. Пусть случайные величины X и Y независимы, g и h — функции из \mathbb{R} в \mathbb{R} . Тогда случайные величины $g(X)$ и $h(Y)$ также независимы.

Доказательство. Для любых $B_1 \subset \mathbb{R}$, $B_2 \subset \mathbb{R}$

$$\begin{aligned} \mathbf{P}(g(X) \in B_1, h(Y) \in B_2) &= \mathbf{P}(X \in g^{-1}(B_1), Y \in h^{-1}(B_2)) = \\ &= \mathbf{P}(X \in g^{-1}(B_1)) \mathbf{P}(Y \in h^{-1}(B_2)) = \mathbf{P}(g(X) \in B_1) \mathbf{P}(h(Y) \in B_2), \end{aligned}$$

где $g^{-1}(B_1) = \{y : g(y) \in B_1\}$, $h^{-1}(B_2) = \{y : h(y) \in B_2\}$.

Пусть теперь случайная величина X обладает плотностью распределения $f_X(t)$. Образует новую случайную величину $Y = g(X)$, где g — некоторая неслучайная функция. Разумеется, Y не обязательно обладает плотностью, достаточно взять $g(t) \equiv C$, чтобы убедиться в этом. Однако если g такова, что $f_Y(t)$ все-таки существует, то как ее найти?

Начнем с рассмотрения функции распределения $F_Y(y)$.

$$F_Y(y) = \mathbf{P}(g(X) < y) = \mathbf{P}(X \in g^{-1}((-\infty, y))) = \int_{g^{-1}((-\infty, y))} f_X(u) du.$$

Теперь задача состоит в том, чтобы преобразовать полученный интеграл к виду

$$\int_{-\infty}^y h(t) dt$$

с некоторой подынтегральной функцией $h(t)$, которая и будет являться плотностью для Y в соответствии с определением. Единого подхода здесь не существует, чаще всего помогает преобразовать интеграл к нужному виду подходящая замена переменных.

Проиллюстрируем все это более подробно на примере преобразования $Y = aX + b$, где $a \neq 0$.

Теорема 2. Пусть распределение случайной величины X обладает плотностью $f_X(t)$. Тогда для любых чисел $a \neq 0$ и b

$$f_{aX+b}(t) = \frac{1}{|a|} f_X\left(\frac{t-b}{a}\right). \quad (2)$$

Доказательство. Пусть $a > 0$. Тогда

$$F_Y(y) = \mathbf{P}(aX + b < y) = \mathbf{P}\left(X < \frac{y-b}{a}\right) = \int_{-\infty}^{(y-b)/a} f_X(u) du.$$

Сделаем замену $t = au + b$. Тогда

$$F_Y(y) = \int_{-\infty}^y \frac{1}{a} f_X\left(\frac{t-b}{a}\right) dt.$$

Если $a < 0$, то, используя ту же замену переменной, получаем

$$\begin{aligned} F_Y(y) &= \mathbf{P}(aX + b < y) = \mathbf{P}\left(X > \frac{y-b}{a}\right) = \int_{(y-b)/a}^{\infty} f_X(u) du = \\ &= \int_y^{-\infty} \frac{1}{a} f_X\left(\frac{t-b}{a}\right) dt = \int_{-\infty}^y \frac{1}{|a|} f_X\left(\frac{t-b}{a}\right) dt. \end{aligned}$$

Таким образом, при всех $a \neq 0$ верна формула (2).

Выведем отсюда несколько полезных следствий для гауссовских распределений.

Следствие 1. Если $X \in \Phi_{\alpha, \sigma^2}$, то $Y = (X - \alpha)/\sigma \in \Phi_{0,1}$.

Следствие 2. Если $Y \in \Phi_{0,1}$, то $X = \sigma Y + \alpha \in \Phi_{\alpha, \sigma^2}$.

Следствие 3. Если $X \in \Phi_{\alpha, \sigma^2}$, то $Y = AX + B \in \Phi_{A\alpha+B, \sigma^2 A^2}$.

Доказательство. Утверждения первых двух следствий прямо вытекают из формулы (2). Для доказательства третьего утверждения удобно сначала представить

$$AX + B = \sigma A \left(\frac{X - \alpha}{\sigma} \right) + A\alpha + B,$$

и затем воспользоваться предыдущими двумя утверждениями.

Пусть теперь X и Y — две случайные величины с известными функциями распределения. Можно ли выразить F_{X+Y} через F_X и F_Y ?

Ответ отрицательный: если больше ничего не предполагать про X и Y , то информации, заложенной в F_X и F_Y , недостаточно, чтобы найти F_{X+Y} . При одних и тех же F_X и F_Y можно получать разные результаты.

Пример. Пусть $X \in \Phi_{0,1}$, $Y = X$, тогда $X + Y = 2X \in \Phi_{0,4}$.

Если же взять $Y = -X$, то по-прежнему $Y \in \Phi_{0,1}$, и получаем, что $X + Y = 0 \in I_0$ при тех же F_X и F_Y .

Если дополнительно предположить, что X и Y независимы, то F_{X+Y} полностью определяется через F_X и F_Y . Мы покажем, как это делается, отдельно для целочисленных случайных величин и для распределений с плотностью.

Теорема 3. Пусть случайные величины X и Y независимы и каждая из них принимает целые неотрицательные значения, при этом

$$\mathbf{P}(X = k) = p_k, \quad \mathbf{P}(Y = k) = q_k, \quad k = 0, 1, 2, \dots$$

Тогда для $k = 0, 1, 2, \dots$

$$r_k = \mathbf{P}(X + Y = k) = \sum_{i=0}^k p_i q_{k-i}.$$

Доказательство.

$$\begin{aligned}\mathbf{P}(X+Y=k) &= \mathbf{P}(\{X=0, Y=k\} \cup \{X=1, Y=k-1\} \cup \dots \cup \{X=k, Y=0\}) = \\ &= \sum_{i=0}^k \mathbf{P}(X=i, Y=k-i) = \sum_{i=0}^k \mathbf{P}(X=i) \mathbf{P}(Y=k-i) = \sum_{i=0}^k p_i q_{k-i}.\end{aligned}$$

Последовательность чисел $\{r_k = \sum_{i=0}^k p_i q_{k-i}, \quad k=0, 1, 2, \dots\}$ называется *сверткой* последовательностей $\{p_k, \quad k=0, 1, 2, \dots\}$ и $\{q_k, \quad k=0, 1, 2, \dots\}$.

В качестве следствия получим следующий интересный результат.

Теорема 4. Пусть X_1 и X_2 независимы, $X_1 \in \Pi_{\lambda_1}$, $X_2 \in \Pi_{\lambda_2}$. Тогда $X_1 + X_2 \in \Pi_{\lambda_1 + \lambda_2}$.

Доказательство. Воспользуемся полученной формулой:

$$\mathbf{P}(X+Y=k) = \sum_{i=0}^k \frac{\lambda_1^i}{i!} e^{-\lambda_1} \frac{\lambda_2^{k-i}}{(k-i)!} e^{-\lambda_2} = \frac{e^{-(\lambda_1+\lambda_2)}}{k!} \sum_{i=0}^k C_k^i \lambda_1^i \lambda_2^{k-i} = \frac{(\lambda_1 + \lambda_2)^k}{k!} e^{-(\lambda_1+\lambda_2)}.$$

Перейдем к рассмотрению плотностей сумм независимых случайных величин.

Теорема 5. Пусть X и Y независимы и имеют плотности распределения f_X и f_Y соответственно. Тогда

$$f_{X+Y}(t) = \int_{-\infty}^{\infty} f_X(u) f_Y(t-u) du = \int_{-\infty}^{\infty} f_Y(v) f_X(t-v) dv.$$

Доказательство. Достаточно доказать первое соотношение, второе получается из него заменой $v = t - u$. Имеем для функции распределения

$$\begin{aligned}F_{X+Y}(y) &= \mathbf{P}(X+Y < y) = \mathbf{P}((X, Y) \in \{(u, v) : u+v < y\}) = \int \int_{u+v < y} f_{X,Y}(u, v) dudv \\ &= \int_{-\infty}^{\infty} f_X(u) \int_{-\infty}^{y-u} f_Y(v) dv du = \int_{-\infty}^{\infty} f_X(u) \int_{-\infty}^y f_Y(t-u) dt du = \\ &= \int_{-\infty}^y \left\{ \int_{-\infty}^{\infty} f_X(u) f_Y(t-u) du \right\} dt.\end{aligned}$$

Мы здесь воспользовались свойством $f_{X,Y}(u, v) = f_X(u) f_Y(v)$ для независимых X и Y и заменой переменной $t = u + v$. Выражение, стоящее в фигурных скобках, и будет искомой плотностью распределения суммы. Теорема доказана.

Оба интеграла, присутствующие в формулировке теоремы, называются *свертками* плотностей f_X и f_Y .

Продemonстрируем на примерах, как работает операция свертки.

Теорема 6. Пусть X_1 и X_2 независимы, $X_1 \in \Gamma_{\alpha, \lambda_1}$, $X_2 \in \Gamma_{\alpha, \lambda_2}$. Тогда $X_1 + X_2 \in \Gamma_{\alpha, \lambda_1 + \lambda_2}$.

Доказательство.

$$f_{X_1+X_2}(t) = \int_{-\infty}^{\infty} \gamma_{\alpha, \lambda_1}(u) \gamma_{\alpha, \lambda_2}(t-u) du.$$

Поскольку $\gamma_{\alpha,\lambda}(u) = 0$ при $u \leq 0$, то стоящие под интегралом функции обе отличны от нуля только если одновременно $u > 0$ и $t - u > 0$. При $t \leq 0$ эти неравенства несовместны, т. е. $f_{X_1+X_2}(t) = 0$. Если $t > 0$, то подынтегральные функции отличны от нуля при $0 < u < t$, поэтому

$$\begin{aligned} f_{X_1+X_2}(t) &= \int_0^t \frac{\alpha^{\lambda_1}}{\Gamma(\lambda_1)} u^{\lambda_1-1} e^{-\alpha u} \frac{\alpha^{\lambda_2}}{\Gamma(\lambda_2)} (t-u)^{\lambda_2-1} e^{-\alpha(t-u)} du \\ &= \frac{\alpha^{\lambda_1+\lambda_2}}{\Gamma(\lambda_1)\Gamma(\lambda_2)} e^{-\alpha t} \int_0^t u^{\lambda_1-1} (t-u)^{\lambda_2-1} du. \end{aligned}$$

Сделаем замену $u = vt$. Тогда

$$f_{X_1+X_2}(t) = \frac{\alpha^{\lambda_1+\lambda_2}}{\Gamma(\lambda_1)\Gamma(\lambda_2)} t^{\lambda_1+\lambda_2-1} e^{-\alpha t} \int_0^1 v^{\lambda_1-1} (1-v)^{\lambda_2-1} dv.$$

Последний интеграл от t уже не зависит. Это константа, которую можно объединить с константами, стоящими в начале формулы. На этом доказательство можно завершить, потому что мы получили выражение вида $C t^{\lambda_1+\lambda_2-1} e^{-\alpha t}$, т.е. плотность гамма-распределения с параметрами $(\alpha, \lambda_1 + \lambda_2)$. Можно, впрочем, и уточнить значение константы. Указанный интеграл известен в теории как бета-функция

$$B(\lambda_1, \lambda_2) = \int_0^1 v^{\lambda_1-1} (1-v)^{\lambda_2-1} dv = \frac{\Gamma(\lambda_1)\Gamma(\lambda_2)}{\Gamma(\lambda_1 + \lambda_2)}.$$

Последнее можно найти в таблицах интегралов. Теорема доказана.

Следствие. Если случайные величины X_1, \dots, X_n независимы и все $X_i \in E_\alpha$, то $X_1 + \dots + X_n \in \Gamma_{\alpha,n}$.

Доказательство следует из того, что $E_\alpha = \Gamma_{\alpha,1}$.

Теорема 7. Пусть X_1 и X_2 независимы, $X_1 \in \Phi_{\alpha_1, \sigma_1^2}$, $X_2 \in \Phi_{\alpha_2, \sigma_2^2}$. Тогда $X_1 + X_2 \in \Phi_{\alpha_1+\alpha_2, \sigma_1^2+\sigma_2^2}$.

Доказательство. Введем новые случайные величины

$$Y_1 = \frac{X_1 - \alpha_1}{\sigma_1}, \quad Y_2 = \frac{X_2 - \alpha_2}{\sigma_1}.$$

Тогда

$$Y_1 \in \Phi_{0,1}, \quad Y_2 = \frac{\sigma_2}{\sigma_1} \frac{X_2 - \alpha_2}{\sigma_2} \in \Phi_{0, \sigma_2^2/\sigma_1^2}.$$

Если мы докажем, что $Y_1 + Y_2 \in \Phi_{0, 1+\sigma_2^2/\sigma_1^2}$, то по свойству линейных преобразований

$$X_1 + X_2 = \sigma_1(Y_1 + Y_2) + \alpha_1 + \alpha_2 \in \Phi_{\alpha_1+\alpha_2, \sigma_1^2+\sigma_2^2}.$$

Обозначим для краткости $\theta^2 = \sigma_2^2/\sigma_1^2$. Тогда

$$\begin{aligned} f_{Y_1+Y_2}(t) &= \int_{-\infty}^{\infty} \frac{1}{\theta\sqrt{2\pi}} \exp\left\{-\frac{u^2}{2\theta^2}\right\} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{(t-u)^2}{2}\right\} du = \\ &= \frac{1}{2\pi\theta} \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2}\left(\frac{u^2}{\theta^2} + u^2 - 2tu + t^2\right)\right\} du = \\ &= \frac{1}{2\pi\theta} \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2}\left(u^2 \frac{1+\theta^2}{\theta^2} - 2u\sqrt{\frac{1+\theta^2}{\theta^2}}t\sqrt{\frac{\theta^2}{1+\theta^2}} + t^2 \frac{\theta^2}{1+\theta^2} + \frac{t^2}{1+\theta^2}\right)\right\} du = \end{aligned}$$

$$= \frac{1}{2\pi\theta} \exp \left\{ -\frac{t^2}{2(1+\theta^2)} \right\} \int_{-\infty}^{\infty} \exp \left\{ -\frac{1}{2} \left(u\sqrt{\frac{1+\theta^2}{\theta^2}} - t\sqrt{\frac{\theta^2}{1+\theta^2}} \right)^2 \right\} du.$$

Сделаем замену переменной

$$v = u\sqrt{\frac{1+\theta^2}{\theta^2}} - t\sqrt{\frac{\theta^2}{1+\theta^2}}.$$

Тогда

$$\begin{aligned} f_{Y_1+Y_2}(t) &= \frac{1}{2\pi\sqrt{1+\theta^2}} \exp \left\{ -\frac{t^2}{2(1+\theta^2)} \right\} \int_{-\infty}^{\infty} \exp \left\{ -\frac{v^2}{2} \right\} dv \\ &= \frac{1}{\sqrt{2\pi(1+\theta^2)}} \exp \left\{ -\frac{t^2}{2(1+\theta^2)} \right\}. \end{aligned}$$

Теорема доказана.

Следствие. Если случайные величины X_1, \dots, X_n независимы и все $X_i \in \Phi_{\alpha, \sigma^2}$, то $X_1 + \dots + X_n \in \Phi_{n\alpha, n\sigma^2}$, а также

$$\bar{X} = \frac{X_1 + \dots + X_n}{n} \in \Phi_{\alpha, \sigma^2/n}.$$

Последнее следует из того, что $X_i/n \in \Phi_{\alpha/n, \sigma^2/n^2}$.

3. Числовые характеристики распределений

3.1. Математическое ожидание

При рассмотрении случайной величины часто возникает вопрос: чему равно ее среднее значение и как его найти?

В математике известны разные виды средних: среднее арифметическое, среднее геометрическое, среднее квадратическое и т. д. Попробуем понять, какое из них более всего подойдет для наших целей.

Рассмотрим для примера случайную величину X , принимающую всего два значения: 1 и 2. Среднее арифметическое этих значений равно 1.5 — оно одинаково удалено от 1 и 2. Однако если значение 1 принимается с гораздо большей вероятностью, чем значение 2 (например, если $\mathbf{P}(X = 1) = 0.99$, а $\mathbf{P}(X = 2) = 0.01$), то по логике вещей среднее значение должно быть смещено ближе к единице, ведь значение $X = 1$ принимается существенно чаще, чем 2. В связи с этим вместо среднего арифметического $1 \cdot 0.5 + 2 \cdot 0.5$ более естественно использовать среднее взвешенное $1 \cdot 0.99 + 2 \cdot 0.01$, в котором весовыми коэффициентами значений случайной величины являются вероятности этих значений. В итоге получаем, что чем более вероятно значение, тем с большим вкладом оно входит в эту сумму.

Эти соображения и легли в основу определения *математического ожидания* случайной величины. Математическое ожидание есть среднее значение случайной величины. Мы дадим его определение отдельно для дискретных, абсолютно непрерывных и смешанных распределений.

Пусть случайная величина X дискретна, т. е. для некоторого набора чисел y_1, y_2, \dots

$$\sum_{k=1}^{\infty} \mathbf{P}(X = y_k) = 1.$$

Определение. Математическим ожиданием введенной дискретной случайной величины называется

$$\mathbf{E} X = \sum_{k=1}^{\infty} y_k \mathbf{P}(X = y_k),$$

если этот ряд абсолютно сходится, т. е. если

$$\sum_{k=1}^{\infty} |y_k| \mathbf{P}(X = y_k) < \infty.$$

В противном случае мы говорим, что математическое ожидание случайной величины X не существует.

Примеры. 1. Пусть $X \in B_p$. Тогда

$$\mathbf{E} X = 1 \cdot p + 0 \cdot (1 - p) = p.$$

2. Если $X \in B_{n,p}$, то

$$\begin{aligned} \mathbf{E} X &= \sum_{k=0}^n k C_n^k p^k (1-p)^{n-k} = \sum_{k=1}^n \frac{n!}{(k-1)!(n-k)!} p^k (1-p)^{n-k} \\ &= np \sum_{m=0}^{n-1} C_{n-1}^m p^m (1-p)^{n-1-m} = np. \end{aligned}$$

Определение. Математическим ожиданием случайной величины X , имеющей абсолютно непрерывное распределение с плотностью $f_X(t)$, называется

$$\mathbf{E} X = \int_{-\infty}^{\infty} t f_X(t) dt,$$

если только

$$\int_{-\infty}^{\infty} |t| f_X(t) dt < \infty.$$

В противном случае считаем, что $\mathbf{E} X$ не существует.

Ясно, что в этом случае $\mathbf{E} X$ также может восприниматься как среднее взвешенное значений случайной величины, только здесь мы используем интегральный аналог формулы. Роль весовой функции играет плотность распределения.

Заметим, что для вычисления математического ожидания достаточно знать распределение случайной величины, т. е. матожидание — это на самом деле числовая характеристика распределения.

Нетрудно видеть, что матожидание не существует для распределения Коши или, например, для случайной величины X такой, что $\mathbf{P}(X = 2^k) = 2^{-k}$, $k = 1, 2, \dots$

Пример. Пусть $X \in \Phi_{\alpha, \sigma^2}$. Тогда

$$\begin{aligned} \mathbf{E} X &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} t \exp \left\{ -\frac{(t-\alpha)^2}{2\sigma^2} \right\} dt = \\ &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (t-\alpha) \exp \left\{ -\frac{(t-\alpha)^2}{2\sigma^2} \right\} dt + \frac{\alpha}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp \left\{ -\frac{(t-\alpha)^2}{2\sigma^2} \right\} dt = \\ &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} y \exp \left\{ -\frac{y^2}{2\sigma^2} \right\} dy + \alpha \int_{-\infty}^{\infty} \varphi_{\alpha, \sigma^2}(t) dt = \alpha. \end{aligned}$$

Здесь интеграл от плотности нормального распределения равен единице, а предпоследний интеграл равен нулю, так как в нем интегрируется нечетная функция.

Предположим теперь, что случайная величина X имеет функцию распределения смешанного типа

$$F_X(y) = \alpha F_1(y) + \beta F_2(y),$$

где $\alpha + \beta = 1$, $\alpha \geq 0$, $\beta \geq 0$, $F_1(y)$ — абсолютно непрерывная функция распределения, имеющая плотность $f(t)$, а $F_2(y)$ — дискретная функция распределения, имеющая скачки величиной p_1, p_2, \dots в точках y_1, y_2, \dots .

Тогда, по определению,

$$\mathbf{E}X = \alpha \int_{-\infty}^{\infty} t f(t) dt + \beta \sum_{k=1}^{\infty} y_k p_k,$$

если только абсолютно сходятся участвующие здесь интеграл и сумма ряда.

В ряде случаев возникает задача нахождения математического ожидания некоторой функции $g(X)$ от случайной величины (или функции $g(X_1, X_2, \dots, X_n)$ от нескольких случайных величин), при этом изначально известным является только распределение X . Для того чтобы применить данное выше определение математического ожидания, нам сначала следовало бы найти распределение случайной величины $g(X)$, а потом воспользоваться им для вычисления $\mathbf{E}g(X)$.

Оказывается, все можно сделать проще.

Если случайная величина X дискретна и принимает возможные значения y_1, y_2, \dots , то $g(X)$ также будет дискретной со значениями $g(y_1), g(y_2), \dots$ (среди них могут быть повторяющиеся) и

$$\mathbf{P}(g(X) = g(y_k)) = \sum_{i: g(y_i) = g(y_k)} \mathbf{P}(X = y_i).$$

Поэтому

$$\mathbf{E}g(X) = \sum_k g(y_k) \mathbf{P}(g(X) = g(y_k)) = \sum_i g(y_i) \mathbf{P}(X = y_i),$$

т. е. в итоге мы воспользовались распределением исходной случайной величины X . Так же дело обстоит и в случае абсолютно непрерывного распределения X . Имеет место формула

$$\mathbf{E}g(X) = \int_{-\infty}^{\infty} t f_{g(X)}(t) dt = \int_{-\infty}^{\infty} g(t) f_X(t) dt.$$

Разумеется, в ней первый интеграл может быть записан только если распределение $g(X)$ обладает плотностью, для использования второго интеграла наличие плотности у $g(X)$ не обязательно. Интуитивно эта формула понятна: мы усредняем значения случайной величины $g(X)$, которые имеют вид $g(t)$, где t — значение для X . Более строгого обоснования этой формулы мы здесь приводить не будем.

Аналог этой формулы в случае функции от нескольких случайных величин выглядит так:

$$\mathbf{E}g(X_1, X_2, \dots, X_n) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} g(t_1, t_2, \dots, t_n) f_{X_1, X_2, \dots, X_n}(t_1, t_2, \dots, t_n) dt_1 \dots dt_n,$$

где $f_{X_1, X_2, \dots, X_n}(t_1, t_2, \dots, t_n)$ — плотность совместного распределения случайного вектора (X_1, X_2, \dots, X_n) .

Заметим, что распределение случайной величины $g(X_1, X_2, \dots, X_n)$ здесь также может и не быть абсолютно непрерывным — формула остается в силе.

Свойства математического ожидания

1. Если $\mathbf{P}(X = C) = 1$, то $\mathbf{E}X = C$, т. е. математическое ожидание константы равно этой константе. Свойство очевидно.

2. Постоянный множитель можно выносить за знак математического ожидания:

$$\mathbf{E}(\alpha X) = \alpha \mathbf{E}X.$$

Это свойство вытекает из формул для вычисления матожидания функции от случайной величины, в данном случае $g(X) = \alpha X$.

3. $\mathbf{E}(X + Y) = \mathbf{E}X + \mathbf{E}Y$, если все участвующие здесь математические ожидания существуют.

Мы проведем доказательство этого свойства отдельно для случаев, когда X и Y дискретны и когда вектор (X, Y) обладает плотностью совместного распределения.

Если X принимает возможные значения x_1, x_2, \dots , а Y — возможные значения y_1, y_2, \dots , то $X + Y$ будет принимать возможные значения вида $x_i + y_j$, $i = 1, 2, \dots$, $j = 1, 2, \dots$, и

$$\begin{aligned} \mathbf{E}(X + Y) &= \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} (x_i + y_j) \mathbf{P}(X = x_i, Y = y_j) = \sum_{i=1}^{\infty} x_i \sum_{j=1}^{\infty} \mathbf{P}(X = x_i, Y = y_j) + \\ &+ \sum_{j=1}^{\infty} y_j \sum_{i=1}^{\infty} \mathbf{P}(X = x_i, Y = y_j) = \sum_{i=1}^{\infty} x_i \mathbf{P}(X = x_i) + \sum_{j=1}^{\infty} y_j \mathbf{P}(Y = y_j) = \\ &= \mathbf{E}X + \mathbf{E}Y. \end{aligned}$$

Для абсолютно непрерывных распределений имеем по той же схеме

$$\begin{aligned} \mathbf{E}(X + Y) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (u + v) f_{X,Y}(u, v) du dv = \int_{-\infty}^{\infty} u \int_{-\infty}^{\infty} f_{X,Y}(u, v) dv du + \\ &+ \int_{-\infty}^{\infty} v \int_{-\infty}^{\infty} f_{X,Y}(u, v) du dv = \int_{-\infty}^{\infty} u f_X(u) du + \int_{-\infty}^{\infty} v f_Y(v) dv = \\ &= \mathbf{E}X + \mathbf{E}Y. \end{aligned}$$

Доказательство этого свойства для остальных случаев (смешанные распределения) мы опускаем.

Вернемся к рассмотренному выше одному из примеров. Пусть X — число успехов в n испытаниях схемы Бернулли, т. е. $X \in B_{n,p}$. Мы уже нашли, что $\mathbf{E}X = np$. С помощью доказанного свойства мы найдем $\mathbf{E}X$ другим способом. Введем вспомогательные случайные величины X_i , $i = 1, 2, \dots, n$, где X_i — число успехов в i -м испытании, т. е. $\mathbf{P}(X_i = 1) = p$, $\mathbf{P}(X_i = 0) = 1 - p$, и поэтому $\mathbf{E}X_i = p$. Тогда $X = X_1 + \dots + X_n$ и $\mathbf{E}X = \mathbf{E}X_1 + \dots + \mathbf{E}X_n = np$.

4. Если X и Y независимы, то $\mathbf{E}(XY) = \mathbf{E}X \cdot \mathbf{E}Y$. Мы вновь предполагаем, что все участвующие здесь матожидания существуют.

Заметим, что обратное утверждение неверно: можно привести пример зависимых случайных величин, для которых это свойство также выполняется. Достаточно взять несомненно зависимые случайные величины X и X^2 , где $X \in U_{-1,1}$. Тогда

$$\mathbf{E}(X \cdot X^2) = \int_{-1}^1 t^3 \frac{1}{2} dt = 0, \quad \mathbf{E}X = \int_{-1}^1 t \frac{1}{2} dt = 0, \quad \mathbf{E}X^2 = \frac{1}{3}.$$

Доказательство свойства 4, как и в предыдущем случае, проведем отдельно для дискретных и абсолютно непрерывных распределений вектора (X, Y) , сохраняя обозначения предыдущего пункта. Имеем в дискретном случае

$$\mathbf{P}(X = x_i, Y = y_j) = \mathbf{P}(X = x_i)\mathbf{P}(Y = y_j)$$

(свойство независимости), поэтому

$$\begin{aligned}\mathbf{E}(XY) &= \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} x_i y_j \mathbf{P}(X = x_i, Y = y_j) = \sum_{i=1}^{\infty} x_i \mathbf{P}(X = x_i) \sum_{j=1}^{\infty} y_j \mathbf{P}(Y = y_j) = \\ &= \mathbf{E}X \mathbf{E}Y.\end{aligned}$$

Далее, в силу независимости, для совместной плотности распределения имеем $f_{X,Y}(u, v) = f_X(u)f_Y(v)$, поэтому

$$\begin{aligned}\mathbf{E}(XY) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} uv f_{X,Y}(u, v) du dv = \int_{-\infty}^{\infty} u f_X(u) du \int_{-\infty}^{\infty} v f_Y(v) dv = \\ &= \mathbf{E}X \mathbf{E}Y.\end{aligned}$$

5. Если $X \geq Y$, то $\mathbf{E}X \geq \mathbf{E}Y$.

Обозначим $Z = X - Y$, тогда свойство 5 эквивалентно утверждению: если $Z \geq 0$, то $\mathbf{E}Z \geq 0$.

Доказательство проведем для общего случая, когда $F(y) = \mathbf{P}(Z < y)$ — функция распределения смешанного типа. В силу того что $F(0) = 0$, делаем следующий вывод: в разложении функции F на абсолютно непрерывную и дискретную компоненты

$$F(y) = \alpha F_1(y) + \beta F_2(y)$$

имеет место $F_1(0) = F_2(0) = 0$. Поэтому в формуле для математического ожидания

$$\mathbf{E}Z = \alpha \int_{-\infty}^{\infty} t f(t) dt + \beta \sum_{k=1}^{\infty} y_k p_k$$

$f(t) = 0$ при $t < 0$ и все y_k неотрицательны. Отсюда следует, что $\mathbf{E}Z \geq 0$.

6. Если $\mathbf{E}X = 0$ и $X \geq 0$, то $\mathbf{P}(X = 0) = 1$.

Для доказательства нам потребуется *неравенство Чебышева*. Назовем его первым неравенством Чебышева, поскольку далее в курсе будет предложено второе.

Первое неравенство Чебышева. Если $X \geq 0$, то для любого $\delta > 0$

$$\mathbf{P}(X \geq \delta) \leq \frac{\mathbf{E}X}{\delta}.$$

Доказательство. Если $\mathbf{E}X = \infty$, то неравенство очевидно. Пусть теперь $\mathbf{E}X < \infty$. Введем случайную величину

$$X' = \begin{cases} 0, & \text{если } 0 \leq X < \delta, \\ \delta, & \text{если } X \geq \delta. \end{cases}$$

Ясно по построению, что $X \geq X'$, поэтому

$$\mathbf{E}X \geq \mathbf{E}X' = 0 \cdot \mathbf{P}(0 \leq X < \delta) + \delta \mathbf{P}(X \geq \delta) = \delta \mathbf{P}(X \geq \delta).$$

Неравенство доказано.

Применим его к доказательству свойства 6. Для любого $\delta > 0$ имеем

$$0 \leq \mathbf{P}(X \geq \delta) \leq \frac{\mathbf{E}X}{\delta} = 0,$$

то есть $\mathbf{P}(X \geq \delta) = 0$, что возможно только при $\mathbf{P}(X = 0) = 1$.

3.2. Моменты

Определение. Моментом k -го порядка случайной величины X называется $\mathbf{E}X^k$, $k > 0$.

Как и всякое математическое ожидание, момент k -го порядка существует тогда и только тогда, когда $\mathbf{E}|X|^k < \infty$. Последнее называется *абсолютным* моментом k -го порядка. Пользуясь формулами вычисления математических ожиданий функций от случайных величин, можем записать

$$\mathbf{E}X^k = \sum_i y_i^k \mathbf{P}(X = y_i)$$

для дискретных распределений и

$$\mathbf{E}X^k = \int_{-\infty}^{\infty} t^k f_X(t) dt$$

для абсолютно непрерывных распределений.

Моменты являются весьма полезными числовыми характеристиками случайных величин. Момент первого порядка — это уже знакомое нам математическое ожидание. Оно имеет смысл среднего значения случайной величины. Мы увидим дальше, что знание моментов второго и первого порядков дает нам определенную информацию о разбросанности значений случайной величины, с помощью моментов можно характеризовать асимметрию распределения и т. д.

Следующая теорема устанавливает связь между существованием моментов разных порядков.

Теорема. Если $\mathbf{E}|X|^k < \infty$, то $\mathbf{E}|X|^m < \infty$ для любого m такого, что $0 < m < k$. Обратное неверно.

Доказательство. Поскольку при $0 < m < k$ всегда верно $|X|^m \leq |X|^k + 1$, то

$$\mathbf{E}|X|^m < \mathbf{E}|X|^k + 1 < \infty.$$

То, что обратное утверждение неверно, показывает следующий пример. Пусть плотность распределения случайной величины X задается формулой

$$f_X(t) = \frac{C}{1 + |t|^{k+2}},$$

где постоянная C выбирается из условия нормировки. Тогда

$$\mathbf{E}|X|^k = C \int_{-\infty}^{\infty} \frac{|t|^k}{1 + |t|^{k+2}} dt < \infty,$$

но $\mathbf{E}|X|^{k+1} = \infty$.

Если при $k > 1$ существует $\mathbf{E}X^k$, то можно рассмотреть также $\mathbf{E}(X - \mathbf{E}X)^k$. Эта величина называется *центральным* моментом k -го порядка. Момент k -го порядка и центральный момент k -го порядка существуют или не существуют одновременно.

3.3. Дисперсия

Дисперсия — это тоже числовая характеристика распределения случайной величины. Она показывает, насколько сильно значения случайной величины отклоняются влево и вправо от среднего значения. Дисперсия определяется только для тех случайных величин, у которых $\mathbf{E}X^2 < \infty$.

Определение. *Дисперсией* случайной величины X называется

$$DX = E(X - EX)^2.$$

Другими словами, дисперсия — это второй центральный момент случайной величины. Она действительно показывает, насколько велик разброс значений случайной величины. Вычитая EX из X , мы получаем всевозможные отклонения от среднего, затем возводим эти разности в квадрат, чтобы не было среди них отрицательных, а потом усредняем, беря математическое ожидание. Таким образом, дисперсия есть среднее квадратическое отклонение значений случайной величины от среднего значения.

Можно предложить альтернативную формулу для дисперсии:

$$\begin{aligned} DX &= E(X - EX)^2 = E(X^2 - 2XEX + (EX)^2) = \\ &= EX^2 - 2EXEX + (EX)^2 = EX^2 - (EX)^2. \end{aligned}$$

Для дискретных распределений дисперсия вычисляется по формулам

$$DX = \sum_{k=1}^{\infty} (y_k - EX)^2 P(X = y_k) = \sum_{k=1}^{\infty} y_k^2 P(X = y_k) - (EX)^2,$$

для распределений абсолютно непрерывного типа имеем

$$DX = \int_{-\infty}^{\infty} (t - EX)^2 f_X(t) dt = \int_{-\infty}^{\infty} t^2 f_X(t) dt - (EX)^2.$$

\sqrt{DX} называется *стандартным отклонением*.

Свойства дисперсии

Здесь и всюду в дальнейшем буквой C будут обозначаться константы.

1. $DX \geq 0$. Свойство очевидно.
2. $DC = 0$. Следует из определения, поскольку $C = EC$.
3. Если $DX = 0$, то $P(X = C) = 1$ для некоторой постоянной C . Действительно, из свойства 6 математических ожиданий вытекает: соотношения $(X - EX)^2 \geq 0$ и $E(X - EX)^2 = 0$ влекут $P(X - EX = 0) = 1$.
4. $D(CX) = C^2 DX$; в частности, $D(-X) = DX$. Это вновь следует из определения: $D(CX) = E(CX - ECX)^2 = C^2 E(X - EX)^2 = C^2 DX$.
5. $D(X + C) = DX$. Следует из определения: $E(X + C - E(X + C))^2 = E(X + C - EX - C)^2 = E(X - EX)^2 = DX$.
6. Если X и Y независимы, то $D(X \pm Y) = DX + DY$.

Доказательство.

$$\begin{aligned} D(X \pm Y) &= E(X \pm Y - E(X \pm Y))^2 = E((X - EX) \pm (Y - EY))^2 \\ &= E(X - EX)^2 + E(Y - EY)^2 \pm 2E((X - EX)(Y - EY)) \\ &= DX + DY \pm 2Cov(X, Y), \end{aligned}$$

где обозначено

$$Cov(X, Y) = E((X - EX)(Y - EY)).$$

Эта величина называется *ковариацией* между случайными величинами X и Y . Если X и Y независимы, то $X - EX$ и $Y - EY$ также независимы и по свойству 4 математических ожиданий имеем

$$Cov(X, Y) = E((X - EX)(Y - EY)) = E(X - EX)E(Y - EY) = 0.$$

Примеры. 1. Пусть $X \in B_p$. Тогда

$$\mathbf{E}X^2 = 1 \cdot p + 0 \cdot (1 - p) = p, \quad \mathbf{E}X = p, \quad \mathbf{D}X = p - p^2 = p(1 - p).$$

2. Если $X \in B_{n,p}$, то можно считать, что X есть число успехов в n испытаниях Бернулли (распределение то же самое). Как мы видели выше, в этом случае X можно представить в виде суммы $X = X_1 + \dots + X_n$, где все X_i распределены по закону Бернулли. Теперь добавим, что они независимы, коль скоро строятся по независимым испытаниям. Поэтому

$$\mathbf{D}X = \mathbf{D}X_1 + \dots + \mathbf{D}X_n = np(1 - p).$$

3. Пусть $X \in \Phi_{\alpha, \sigma^2}$. Нахождение $\mathbf{D}X$ упростится, если мы сведем все к стандартному нормальному закону. Обозначим $Y = (X - \alpha)/\sigma$, тогда $Y \in \Phi_{0,1}$ и $X = \sigma Y + \alpha$. В силу свойств дисперсии $\mathbf{D}X = \sigma^2 \mathbf{D}Y$, поэтому достаточно найти $\mathbf{D}Y$. Интегрируя по частям, получаем

$$\begin{aligned} \mathbf{E}Y^2 &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t^2 e^{-t^2/2} dt = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t d(-e^{-t^2/2}) = \\ &= \frac{1}{\sqrt{2\pi}} \left\{ -te^{-t^2/2} \Big|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} e^{-t^2/2} dt \right\} = \int_{-\infty}^{\infty} \varphi_{0,1}(t) dt = 1. \end{aligned}$$

Поскольку $\mathbf{E}Y = 0$, то $\mathbf{D}Y = 1$ и $\mathbf{D}X = \sigma^2$.

Таким образом, параметры нормального распределения обладают вполне конкретным физическим содержанием: α совпадает с математическим ожиданием, а σ^2 — с дисперсией.

3.4. Коэффициент корреляции

Коэффициент корреляции — это числовая характеристика, которая вводится для пары случайных величин с целью показать, насколько они зависимы.

Определение. Коэффициентом корреляции называется

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\mathbf{D}X \mathbf{D}Y}} = \frac{\mathbf{E}((X - \mathbf{E}X)(Y - \mathbf{E}Y))}{\sqrt{\mathbf{D}X \mathbf{D}Y}} = \frac{\mathbf{E}(XY) - \mathbf{E}X \mathbf{E}Y}{\sqrt{\mathbf{D}X \mathbf{D}Y}}.$$

Коэффициент корреляции вводится не для всякой пары случайных величин: необходимо, чтобы существовали вторые моменты $\mathbf{E}X^2$ и $\mathbf{E}Y^2$, а также, чтобы $\mathbf{D}X > 0$, $\mathbf{D}Y > 0$. Последнее ограничение означает, что X и Y отличны от констант.

Для вычисления коэффициента корреляции необходимо знать совместное распределение пары (X, Y) . Если, к примеру, известна плотность $f_{X,Y}(u, v)$, то смешанный момент вычисляется по формуле

$$\mathbf{E}(XY) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} uv f_{X,Y}(u, v) du dv$$

(мы применяем здесь правило вычисления математического ожидания функции $g(X, Y) = X \cdot Y$). Как ранее установлено, одномерные плотности получаются из двумерной интегрированием:

$$f_X(u) = \int_{-\infty}^{\infty} f_{X,Y}(u, v) dv, \quad f_Y(v) = \int_{-\infty}^{\infty} f_{X,Y}(u, v) du,$$

далее матожидания и дисперсии вычисляются по известным формулам.

Свойства коэффициента корреляции

1. $|\rho(X, Y)| \leq 1$.

Доказательство. Введем случайные величины

$$X_1 = \frac{X - \mathbf{E}X}{\sqrt{\mathbf{D}X}}, \quad Y_1 = \frac{Y - \mathbf{E}Y}{\sqrt{\mathbf{D}Y}}.$$

Здесь к каждой случайной величине применена операция *стандартизации*, которая состоит в вычитании матожидания и делении на корень квадратный из дисперсии. Она производится с единственной целью: добиться, чтобы математическое ожидание стало нулевым, а дисперсия — единичной. Действительно,

$$\mathbf{E}X_1 = \frac{1}{\sqrt{\mathbf{D}X}} \mathbf{E}(X - \mathbf{E}X) = 0, \quad \mathbf{D}X_1 = \left(\frac{1}{\sqrt{\mathbf{D}X}} \right)^2 \mathbf{D}(X - \mathbf{E}X) = \frac{\mathbf{D}X}{\mathbf{D}X} = 1.$$

Кроме того,

$$\rho(X, Y) = \mathbf{E}X_1Y_1 = \text{Cov}(X_1, Y_1).$$

Как уже было установлено,

$$\mathbf{D}(Y_1 \pm X_1) = \mathbf{D}X_1 + \mathbf{D}Y_1 \pm 2\text{Cov}(X_1, Y_1) = 2 \pm 2\rho(X, Y) \geq 0.$$

Последнее неравенство эквивалентно тому, что $-1 \leq \rho(X, Y) \leq 1$.

2. $|\rho(X, Y)| = 1$ тогда и только тогда, когда для некоторых констант $a \neq 0$ и b выполняется $Y = aX + b$.

Доказательство. Если $Y = aX + b$, то

$$\begin{aligned} \rho(X, Y) &= \frac{\mathbf{E}((X - \mathbf{E}X)(aX + b - \mathbf{E}(aX + b)))}{\sqrt{\mathbf{D}X \mathbf{D}(aX + b)}} = \\ &= \frac{\mathbf{E}((X - \mathbf{E}X)(aX + b - a\mathbf{E}X - b))}{\sqrt{\mathbf{D}X a^2 \mathbf{D}X}} = \frac{a\mathbf{E}(X - \mathbf{E}X)^2}{|a|\mathbf{D}X} = \frac{a}{|a|} = \pm 1 \end{aligned}$$

в зависимости от знака числа a .

В другую сторону: пусть, к примеру, $\rho(X, Y) = 1$. Воспользуемся опять соотношением

$$\mathbf{D}(Y_1 - X_1) = 2 - 2\rho(X, Y) = 0.$$

В силу свойств дисперсии это означает, что $Y_1 - X_1 = C$ с вероятностью единица или, что то же самое,

$$Y = \frac{\sqrt{\mathbf{D}Y}}{\sqrt{\mathbf{D}X}} X + \mathbf{E}Y + C\sqrt{\mathbf{D}Y} - \frac{\sqrt{\mathbf{D}Y} \mathbf{E}X}{\sqrt{\mathbf{D}X}}.$$

Если $\rho(X, Y) = -1$, то пользуемся соотношением

$$\mathbf{D}(Y_1 + X_1) = 2 + 2\rho(X, Y) = 0.$$

3. Если X и Y независимы, то $\rho(X, Y) = 0$.

Свойство очевидно.

К сожалению, обратное утверждение не имеет места.

4. Если $\rho(X, Y) = 0$, то X и Y не обязательно независимы.

Пример этому уже приводился: если $X \in U_{-1,1}$ и $Y = X^2$, то $\text{Cov}(X, Y) = \rho(X, Y) = 0$, хотя случайные величины зависимы.

По этой причине случайные величины X и Y называются *некоррелированными*, если $\rho(X, Y) = 0$. Независимость влечет некоррелированность, но не наоборот.

3.5. Многомерный случай: математическое ожидание и матрица ковариаций

В этом разделе мы обобщим понятия математического ожидания и дисперсии на многомерный случай.

Пусть

$$X = \begin{pmatrix} X_{11} & X_{12} & \dots & X_{1n} \\ X_{21} & X_{22} & \dots & X_{2n} \\ \dots & \dots & \dots & \dots \\ X_{m1} & X_{m2} & \dots & X_{mn} \end{pmatrix}$$

— матрица, составленная из случайных величин. Положим, по определению,

$$\mathbf{E}X = \begin{pmatrix} \mathbf{E}X_{11} & \mathbf{E}X_{12} & \dots & \mathbf{E}X_{1n} \\ \mathbf{E}X_{21} & \mathbf{E}X_{22} & \dots & \mathbf{E}X_{2n} \\ \dots & \dots & \dots & \dots \\ \mathbf{E}X_{m1} & \mathbf{E}X_{m2} & \dots & \mathbf{E}X_{mn} \end{pmatrix}.$$

Легко проверить, что при таком определении сохраняются следующие свойства математических ожиданий.

1. Если A и B — матрицы, составленные из констант, то $\mathbf{E}(AX) = A\mathbf{E}X$, $\mathbf{E}(XB) = (\mathbf{E}X)B$.
2. $\mathbf{E}(X + Y) = \mathbf{E}X + \mathbf{E}Y$.
3. Если любой элемент матрицы X не зависит от любого элемента матрицы Y , то $\mathbf{E}(XY) = \mathbf{E}X\mathbf{E}Y$.

Разумеется, мы предполагаем, что размерности участвующих здесь матриц позволяют применять операции сложения и умножения.

Аналог дисперсии вводится только для случайных векторов. На протяжении этого и следующего параграфов мы будем изображать векторы в виде столбцов

$$X = \begin{pmatrix} X_1 \\ X_2 \\ \dots \\ X_n \end{pmatrix}.$$

Определение. Матрицей ковариаций случайного вектора X называется матрица $\mathbf{C}(X)$, у которой на месте с номером (i, j) стоит $c_{i,j} = \text{Cov}(X_i, X_j)$, $i, j = 1, \dots, n$.

Матрица ковариаций есть аналог дисперсии. При $n = 1$ она совпадает с дисперсией. В общем случае на главной диагонали у нее стоят дисперсии $\mathbf{D}X_1, \dots, \mathbf{D}X_n$, матрица симметрична относительно главной диагонали: $c_{i,j} = c_{j,i}$.

Мы знаем, что $\mathbf{D}(AX+B) = A^2\mathbf{D}X$ в одномерном случае, если A и B — константы. Аналогом этого свойства для случайных векторов является следующее утверждение.

Теорема. Пусть A — матрица из констант, имеющая m строк и n столбцов, а B — вектор из констант размерности m . Тогда

$$\mathbf{C}(AX + B) = A\mathbf{C}(X)A^T,$$

где верхний индекс T соответствует транспонированной матрице.

Доказательство. Обозначим для краткости $m_i = \mathbf{E}X_i$, $i = 1, \dots, n$, и воспользуемся следующим свойством произведения матриц: если умножить вектор-столбец на

вектор-строку, то в итоге получим матрицу:

$$\begin{pmatrix} X_1 - m_1 \\ X_2 - m_2 \\ \dots \\ X_n - m_n \end{pmatrix} \cdot (X_1 - m_1, X_2 - m_2, \dots, X_n - m_n) =$$

$$= \begin{pmatrix} (X_1 - m_1)^2 & (X_1 - m_1)(X_2 - m_2) & \dots & (X_1 - m_1)(X_n - m_n) \\ (X_2 - m_2)(X_1 - m_1) & (X_2 - m_2)^2 & \dots & (X_2 - m_2)(X_n - m_n) \\ \dots & \dots & \dots & \dots \\ (X_n - m_n)(X_1 - m_1) & (X_n - m_n)(X_2 - m_2) & \dots & (X_n - m_n)^2 \end{pmatrix}.$$

Взяв теперь математическое ожидание от обеих частей, получим

$$\mathbf{E}(X - \mathbf{E}X)(X - \mathbf{E}X)^T = \mathbf{C}(X),$$

и аналогично

$$\begin{aligned} \mathbf{C}(AX + B) &= \mathbf{E}(AX + B - \mathbf{E}(AX + B))(AX + B - \mathbf{E}(AX + B))^T = \\ &= \mathbf{E}(AX + B - \mathbf{E}AX - B)(AX + B - \mathbf{E}AX - B)^T = \\ &= \mathbf{E}(A(X - \mathbf{E}X)(X - \mathbf{E}X)^T A^T) = \\ &= A(\mathbf{E}(X - \mathbf{E}X)(X - \mathbf{E}X)^T)A^T = A\mathbf{C}(X)A^T. \end{aligned}$$

Теорема доказана.

3.6. Многомерное нормальное распределение

Мы уже рассматривали ранее в качестве примера частный случай плотности многомерного нормального закона, она соответствовала стандартному многомерному нормальному распределению. Сейчас введем этот закон распределения в общей форме.

Будем действовать по аналогии с одномерным случаем.

Пусть Y — одномерная случайная величина, имеющая стандартное нормальное распределение, $Y \in \Phi_{0,1}$. Взяв произвольные числа α и $\sigma > 0$, образуем случайную величину $X = \sigma Y + \alpha$, которая уже будет распределена по закону Φ_{α, σ^2} с плотностью

$$\varphi_{\alpha, \sigma^2}(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(t-\alpha)^2/2\sigma^2}, \quad -\infty < t < \infty.$$

Тем самым мы получили общий вид плотности нормального распределения с помощью линейных преобразований над случайной величиной Y , имеющей стандартное нормальное распределение.

Так же поступим и в многомерном случае. Пусть Y — случайный вектор с координатами Y_1, \dots, Y_n , имеющий многомерное стандартное нормальное распределение с плотностью

$$f_Y(t) = \frac{1}{(2\pi)^{n/2}} \exp \left\{ -\frac{1}{2} \sum_{i=1}^n t_i^2 \right\} = \frac{1}{(2\pi)^{n/2}} \exp \left\{ -\frac{1}{2} t^T t \right\} = \prod_{i=1}^n \varphi_{0,1}(t_i).$$

Последнее означает, что координаты вектора Y_1, \dots, Y_n независимы и одинаково распределены в соответствии со стандартным нормальным законом. Здесь $\mathbf{C}(Y) = E$ — единичная матрица, $t = (t_1, \dots, t_n)^T$ — вектор-столбец.

Возьмем произвольную невырожденную матрицу A размерности $n \times n$, состоящую из констант, и постоянный вектор $\alpha = (\alpha_1, \dots, \alpha_n)^T$ и образуем новый случайный вектор

$$X = AY + \alpha.$$

Распределение получившегося вектора X и будем называть *многомерным нормальным* распределением.

Теорема. *Плотность многомерного нормального распределения задается формулой*

$$f_X(t) = \frac{\sqrt{\det Q}}{(2\pi)^{n/2}} \exp \left\{ -(t - \alpha)^T Q (t - \alpha) / 2 \right\},$$

где $t = (t_1, \dots, t_n)^T$, $Q = (\mathbf{C}(X))^{-1} = (A\mathbf{C}(Y)A^T)^{-1} = (AA^T)^{-1}$.

Заметим, что при $n = 1$ и $A = \sigma$ получаем $Q = 1/\sigma^2$, $\sqrt{\det Q} = 1/\sigma$.

Приведем только схему доказательства.

Какой бы прямоугольник $B \subset \mathbb{R}^n$ ни взять, по основному свойству плотностей должно быть

$$\mathbf{P}(X \in B) = \int_B f_X(t) dt,$$

здесь для краткости обозначено $dt = dt_1 dt_2 \dots dt_n$. В то же время

$$\mathbf{P}(AY + \alpha \in B) = \mathbf{P}(Y \in A^{-1}(B - \alpha)) = \int_{A^{-1}(B - \alpha)} f_Y(u) du,$$

где под множеством $A^{-1}(B - \alpha)$ понимается совокупность всех точек $u : Au + \alpha \in B$. Для того чтобы преобразовать этот интеграл к нужному нам интегралу по множеству B (и тогда стоящая под интегралом функция и будет искомой плотностью), сделаем замену переменных $t = Au + \alpha$. В результате этой замены множество $A^{-1}(B - \alpha)$ перейдет в B , u — в $A^{-1}(t - \alpha)$, $\exp \left\{ -\frac{1}{2} u^T u \right\}$ перейдет в

$$\exp \left\{ -\frac{1}{2} (t - \alpha)^T (A^{-1})^T A^{-1} (t - \alpha) \right\} = \exp \left\{ -\frac{1}{2} (t - \alpha)^T (AA^T)^{-1} (t - \alpha) \right\}.$$

При переходе от du к dt появится якобиан $\sqrt{\det Q}$.

Таким образом, под интегралом появится функция, присутствующая в утверждении теоремы — она и будет плотностью вектора X .

Для нас большую важность представляют следующие два следствия из этой теоремы.

Следствие 1. *Пусть случайный вектор $X = (X_1, X_2, \dots, X_n)^T$ имеет многомерное нормальное распределение и все его компоненты попарно некоррелированы. Тогда они независимы.*

Доказательство. Вследствие некоррелированности компонент заключаем, что матрица ковариаций $\mathbf{C}(X)$ имеет диагональный вид: на главной диагонали стоят дисперсии $\mathbf{D}X_1, \dots, \mathbf{D}X_n$, а все остальные элементы равны нулю. Обозначим для краткости $\sigma_i^2 = \mathbf{D}X_i$, $i = 1, \dots, n$. Тогда матрица $Q = (\mathbf{C}(X))^{-1}$ также будет диагональной, у нее на главной диагонали будут стоять числа $\sigma_1^{-2}, \dots, \sigma_n^{-2}$. По этой причине плотность распределения вектора X приобретает вид

$$f_X(t) = \frac{1}{\sigma_1 \dots \sigma_n (2\pi)^{n/2}} \exp \left\{ -\sum_{i=1}^n \frac{(t_i - \alpha_i)^2}{2\sigma_i^2} \right\} = \prod_{i=1}^n \varphi_{\alpha_i \sigma_i^2}(t_i),$$

что эквивалентно независимости компонент вектора X .

Следствие 2. Пусть случайный вектор $X = (X_1, X_2, \dots, X_n)^T$ имеет многомерное стандартное нормальное распределение (напомним: это соответствует тому, что все компоненты вектора независимы и имеют распределение $\Phi_{0,1}$). Образуем новый вектор $Y = AX$, где A — ортогональная матрица. Тогда вектор Y также будет иметь многомерное стандартное нормальное распределение.

Доказательство. Ортогональная матрица, по определению, обладает свойством $A^T = A^{-1}$. По этой причине $C(Y) = AC(X)A^T = AA^T = E$ и, следовательно,

$$f_Y(t) = \frac{1}{(2\pi)^{n/2}} \exp \left\{ -\frac{1}{2} t^T t \right\} = \prod_{i=1}^n \varphi_{0,1}(t_i) = f_X(t),$$

что и требовалось доказать.

4. Предельные теоремы

4.1. Сходимость по вероятности

В дальнейшем нам предстоит изучить закон больших чисел — теорему о сходимости некоторой последовательности случайных величин. Случайные величины являются функциями, заданными на пространстве элементарных исходов, а сходимость последовательности функций — понятие сложное и ее можно определять по-разному. Мы ограничимся введением понятия *сходимости по вероятности*.

Пусть случайные величины X, X_1, X_2, \dots заданы на одном и том же вероятностном пространстве.

Определение. Последовательность $\{X_n\}$ сходится по вероятности к случайной величине X , если для любого числа $\varepsilon > 0$

$$\mathbf{P}(|X_n - X| \geq \varepsilon) \rightarrow 0$$

при $n \rightarrow \infty$.

Обозначать будем $X_n \xrightarrow{P} X$.

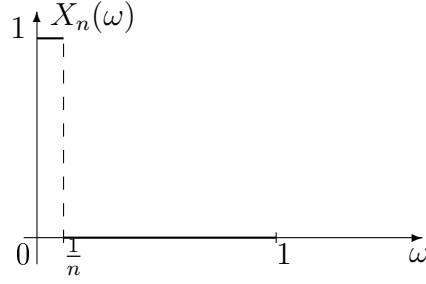
Эквивалентное определение: для любого $\varepsilon > 0$

$$\mathbf{P}(|X_n - X| < \varepsilon) \rightarrow 1.$$

Поясним смысл написанного. При сближении X_n и X расхождение между ними должно в каком-то смысле уменьшаться. То, что написано в определении, означает: большие расхождения (т. е. когда $|X_n - X| \geq \varepsilon$) возможны, но вероятность появления таких расхождений стремится к нулю.

Пример последовательности, сходящейся по вероятности. Пусть, как и ранее, $\Omega = [0, 1]$. Для всякого интервала $A \subset \Omega$ положим $\mathbf{P}(A) = \lambda(A)$, где $\lambda(A)$ — длина интервала. Определим случайные величины $X(\omega) \equiv 0$,

$$X_n(\omega) = \begin{cases} 1, & \omega \in [0, \frac{1}{n}], \\ 0, & \text{иначе.} \end{cases}$$



Ясно, что $\mathbf{P}(|X_n - X| \geq \varepsilon) = 0$ при $\varepsilon > 1$. Если же $\varepsilon \leq 1$, то $\mathbf{P}(|X_n - X| \geq \varepsilon) = 1/n \rightarrow 0$ при $n \rightarrow \infty$.

Некоторые свойства сходимости по вероятности

1. Если $X_n \xrightarrow{P} X$, $Y_n \xrightarrow{P} Y$, то $X_n + Y_n \xrightarrow{P} X + Y$.

Доказательство. Для любого $\varepsilon > 0$

$$\begin{aligned} \mathbf{P}(|X_n + Y_n - X - Y| \geq \varepsilon) &= \mathbf{P}(|(X_n - X) + (Y_n - Y)| \geq \varepsilon) \leq \\ &\leq \mathbf{P}(|X_n - X| + |Y_n - Y| \geq \varepsilon) \leq \\ &\leq \mathbf{P}(\{|X_n - X| \geq \varepsilon/2\} \cup \{|Y_n - Y| \geq \varepsilon/2\}) \leq \\ &\leq \mathbf{P}(|X_n - X| \geq \varepsilon/2) + \mathbf{P}(|Y_n - Y| \geq \varepsilon/2) \rightarrow 0. \end{aligned}$$

2. Пусть при $n \rightarrow \infty$ $X_n^{(1)} \xrightarrow{P} a_1$, $X_n^{(2)} \xrightarrow{P} a_2$, ..., $X_n^{(k)} \xrightarrow{P} a_k$, функция $g: \mathbb{R}^k \rightarrow \mathbb{R}$ непрерывна в точке $a = (a_1, \dots, a_k)$. Тогда

$$g(X_n^{(1)}, X_n^{(2)}, \dots, X_n^{(k)}) \xrightarrow{P} g(a_1, \dots, a_k).$$

при $n \rightarrow \infty$.

Доказательство. По определению непрерывности, для любого $\varepsilon > 0$ найдется число $\delta = \delta(a, \varepsilon)$ такое, что если $|y_i - a_i| < \delta$ для всех i , то

$$|g(y_1, \dots, y_k) - g(a_1, \dots, a_k)| < \varepsilon.$$

Введем события $B_i = \{|X_n^{(i)} - a_i| < \delta\}$, $i = 1, \dots, k$, тогда

$$B_1 B_2 \dots B_k \subset \{|g(X_n^{(1)}, \dots, X_n^{(k)}) - g(a_1, \dots, a_k)| < \varepsilon\},$$

откуда следует неравенство для вероятностей

$$\mathbf{P}(B_1 B_2 \dots B_k) \leq \mathbf{P}\{|g(X_n^{(1)}, \dots, X_n^{(k)}) - g(a_1, \dots, a_k)| < \varepsilon\} \leq 1.$$

По условию $\mathbf{P}(B_i) \rightarrow 1$ при $n \rightarrow \infty$ для каждого i . Покажем, что также $\mathbf{P}(B_1 B_2 \dots B_k) \rightarrow 1$. Для двух событий имеем

$$\mathbf{P}(B_1 B_2) = \mathbf{P}(B_1) + \mathbf{P}(B_2) - \mathbf{P}(B_1 \cup B_2).$$

Каждая вероятность в правой части стремится к единице, следовательно, $\mathbf{P}(B_1 B_2) \rightarrow 1$. Применяя индукцию, получаем утверждение для любого k .

Тем самым мы доказали, что

$$\mathbf{P}\{|g(X_n^{(1)}, \dots, X_n^{(k)}) - g(a_1, \dots, a_k)| < \varepsilon\} \rightarrow 1.$$

В частности, если $X_n^{(1)} \xrightarrow{P} a_1$ и $X_n^{(2)} \xrightarrow{P} a_2$, то $X_n^{(1)} X_n^{(2)} \xrightarrow{P} a_1 a_2$. Это следует из доказанного выше свойства, если при $k = 2$ в качестве функции g взять $g(u, v) = uv$.

Для установления факта сходимости по вероятности часто пользуются следующим утверждением.

Второе неравенство Чебышева. Пусть $\mathbf{E}X^2 < \infty$, тогда для любого $\varepsilon > 0$

$$\mathbf{P}(|X - \mathbf{E}X| \geq \varepsilon) \leq \frac{\mathbf{D}X}{\varepsilon^2}.$$

Доказательство. Достаточно применить первое неравенство Чебышева к случайной величине $(X - \mathbf{E}X)^2$:

$$\mathbf{P}(|X - \mathbf{E}X| \geq \varepsilon) = \mathbf{P}((X - \mathbf{E}X)^2 \geq \varepsilon^2) \leq \frac{\mathbf{E}(X - \mathbf{E}X)^2}{\varepsilon^2}.$$

Неравенство доказано.

4.2. Закон больших чисел

Предположим, что раз за разом повторяется один и тот же случайный эксперимент, и каждый раз в результате него мы измеряем какую-то характеристику. Получаем тем самым последовательность случайных величин X_1, X_2, \dots . Их можно считать взаимно независимыми, если последовательные эксперименты не влияли друг на друга, а также одинаково распределенными (т.е. имеющими одно и то же распределение), если эксперименты по сути повторяют друг друга. Пример тому — повторяющиеся испытания Бернулли. Производя без ограничений один эксперимент за другим, мы можем обнаружить ряд закономерностей в получающейся последовательности случайных величин. Одна из таких закономерностей, возникающая при многократном подбрасывании монеты, уже обсуждалась в начале курса. Она является частным случаем следующего более общего утверждения, которое носит название *закона больших чисел*.

Теорема (закон больших чисел). Пусть случайные величины X_1, X_2, \dots независимы и одинаково распределены, причем $\mathbf{E}X_1^2 < \infty$. Обозначим $a = \mathbf{E}X_1$, $S_n = \sum_{i=1}^n X_i$. Тогда при $n \rightarrow \infty$

$$\frac{S_n}{n} \xrightarrow{P} a.$$

Доказательство. Заметим, что

$$\mathbf{E}(S_n/n) = \frac{\mathbf{E}X_1 + \dots + \mathbf{E}X_n}{n} = \frac{na}{n} = a.$$

Обозначим $\sigma^2 = \mathbf{D}X_1$ и применим второе неравенство Чебышева к случайной величине S_n/n :

$$\mathbf{P}\left\{\left|\frac{S_n}{n} - a\right| \geq \varepsilon\right\} \leq \frac{\mathbf{D}(S_n/n)}{\varepsilon^2} = \frac{n\sigma^2}{n^2\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2} \rightarrow 0$$

при $n \rightarrow \infty$.

Следствие (теорема Бернулли). Пусть S_n — число успехов в n испытаниях схемы Бернулли, p — вероятность успеха в одном испытании. Тогда

$$\frac{S_n}{n} \xrightarrow{P} p$$

при $n \rightarrow \infty$.

Доказательство. Пусть X_i — число успехов в i -м испытании. Тогда $X_i \in B_p$ и все эти случайные величины независимы. Здесь $S_n = X_1 + \dots + X_n$, $\mathbf{E}X_i = p$ и тем самым выполнены все условия теоремы.

Замечания

1. Условие $\mathbf{E}X_1^2 < \infty$ в теореме завышено. Закон больших чисел справедлив, даже если существует только первый момент $\mathbf{E}|X_1| < \infty$. Однако доказательство теоремы при таком условии потребовало бы больших усилий.

2. Число a есть среднее значение каждой из случайных величин X_i , здесь усреднение произведено по пространству значений случайной величины. Параметр $i = 1, \dots, n$ есть номер эксперимента, его значения можно воспринимать как целочисленные моменты времени. Тем самым

$$\frac{S_n}{n} = \frac{X_1 + \dots + X_n}{n}$$

есть усреднение результатов экспериментов по времени.

Закон больших чисел утверждает, что среднее по времени сближается со средним, вычисленным по пространству значений.

4.3. Центральная предельная теорема

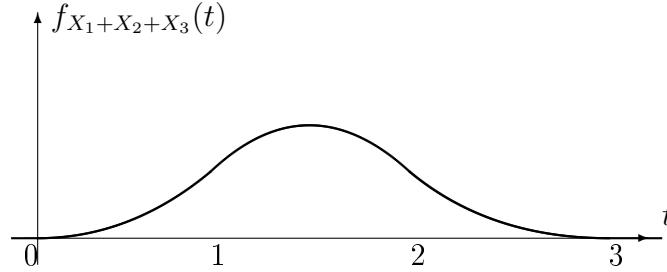
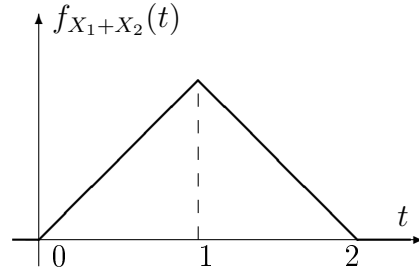
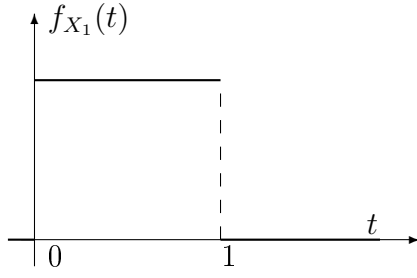
Как и в предыдущем параграфе, будем иметь дело с последовательностью X_1, X_2, \dots независимых одинаково распределенных случайных величин, $S_n = X_1 + \dots + X_n$.

Во многих прикладных задачах возникает необходимость вычислять вероятности вида $\mathbf{P}(A \leq S_n \leq B)$ при больших n . Это происходит, например, при планировании производства, поскольку общая выработка продукции предприятием за смену складывается из случайных объемов продукции, произведенных отдельными рабочими. Вычисление различных средних показателей в экономике, социологии, демографии, статистике также сводится к суммированию случайных величин.

Мы видели, что для нахождения распределения суммы двух независимых случайных величин следует пользоваться формулой свертки. Однако ее применение сопряжено с непростыми вычислениями, в особенности если мы интересуемся распределением суммы большого числа слагаемых. Более продуктивными оказались методы приближенного вычисления указанных вероятностей для сумм.

Мы знаем, что для нахождения вероятности попадания суммы в интервал (или отрезок) достаточно знать ее функцию распределения. Значит, необходимо искать приближения для функций распределения сумм. Оказалось, что в широких условиях функции распределения немного подправленных (а точнее стандартизованных) сумм случайных величин сближаются с функцией распределения стандартного нормального закона, если число слагаемых возрастает.

Этот эффект можно наблюдать на примерах. Пусть все X_i независимы и имеют равномерное на $[0,1]$ распределение. Вычислим с помощью сверток плотности распределения случайных величин $X_1 + X_2$, $X_1 + X_2 + X_3$ (это нетрудно) и увидим, что их графики очень быстро начинают напоминать плотность нормального распределения:



Последний график получается склеиванием трех квадратических парабол. Для $f_{X_1+X_2+X_3+X_4}(t)$ график будет склеиваться из кубических парабол; уже для суммы пяти случайных величин на глаз трудно различить график полученной плотности от гауссовской кривой.

Такую же закономерность мы можем наблюдать, если рисовать графики плотности сумм в том случае, когда все $X_i \in E_\alpha$. Тогда, как мы видели, $S_n \in \Gamma_{\alpha,n}$, и при больших значениях n кривая плотности гамма-распределения, растягиваясь вправо, все больше будет напоминать плотность нормального распределения, только сильно вытянутую и смещенную вправо. Чтобы в пределе получалась плотность стандартного нормального закона, суммы надо подправлять с помощью операции стандартизации.

Эти наблюдения иллюстрируют важную закономерность, о которой пойдет речь ниже.

Центральная предельная теорема (ЦПТ). Пусть X_1, X_2, \dots — независимые одинаково распределенные случайные величины. Предположим, что $\mathbf{E}X_1^2 < \infty$. Обозначим $S_n = X_1 + \dots + X_n$, $a = \mathbf{E}X_1$, $\sigma^2 = \mathbf{D}X_1$, и пусть $\sigma^2 > 0$. Тогда для любого y

$$\mathbf{P}\left(\frac{S_n - na}{\sigma\sqrt{n}} < y\right) = F_{\frac{S_n - na}{\sigma\sqrt{n}}}(y) \rightarrow \Phi_{0,1}(y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^y e^{-t^2/2} dt$$

при $n \rightarrow \infty$.

Доказательство теоремы приводиться не будет. Зато подробно обсудим этот результат.

Замечания

1. Нетрудно видеть, что $\mathbf{E}S_n = na$, $\sqrt{\mathbf{D}S_n} = \sqrt{n\sigma^2} = \sigma\sqrt{n}$, т. е. к случайной величине S_n в теореме применена операция стандартизации;

$$\mathbf{E}\left(\frac{S_n - na}{\sigma\sqrt{n}}\right) = 0, \quad \mathbf{D}\left(\frac{S_n - na}{\sigma\sqrt{n}}\right) = 1.$$

2. Можно показать, что сходимость в ЦПТ является равномерной по всем y , т. е.

$$\sup_y \left| \mathbf{P}\left(\frac{S_n - na}{\sigma\sqrt{n}} < y\right) - \Phi_{0,1}(y) \right| \rightarrow 0$$

при $n \rightarrow \infty$. Доказательство этого факта мы не приводим.

3. Можно сформулировать ЦПТ в эквивалентной форме: для любых $A \leq B$

$$\mathbf{P} \left(A \leq \frac{S_n - na}{\sigma\sqrt{n}} \leq B \right) \rightarrow \frac{1}{\sqrt{2\pi}} \int_A^B e^{-t^2/2} dt.$$

Именно такая форма чаще всего используется при решении задач. Делается это следующим образом. Предположим, что нам необходимо найти вероятность $\mathbf{P}(C \leq S_n \leq D)$ при больших значениях n . Первое, что мы должны сделать, это подогнать наше выражение под формулировку теоремы:

$$\mathbf{P}(C \leq S_n \leq D) = \mathbf{P} \left(\frac{C - na}{\sigma\sqrt{n}} \leq \frac{S_n - na}{\sigma\sqrt{n}} \leq \frac{D - na}{\sigma\sqrt{n}} \right),$$

после чего объявляем эту вероятность почти равной

$$\frac{1}{\sqrt{2\pi}} \int_A^B e^{-t^2/2} dt = \Phi_{0,1}(B) - \Phi_{0,1}(A),$$

где

$$A = \frac{C - na}{\sigma\sqrt{n}}, \quad B = \frac{D - na}{\sigma\sqrt{n}}.$$

Численные значения функции $\Phi_{0,1}(y)$ обычно находятся из таблиц.

4. Коль скоро мы заменяем допредельное выражение в ЦПТ предельным, возникает вопрос о величине погрешности, которую мы допускаем при этом. Это вопрос о скорости сходимости в ЦПТ. Имеет место следующий факт.

Неравенство Берри–Эссеена. Пусть $\mathbf{E}|X_1|^3 < \infty$, тогда

$$\sup_y \left| \mathbf{P} \left(\frac{S_n - na}{\sigma\sqrt{n}} < y \right) - \Phi_{0,1}(y) \right| \leq \frac{\mu}{\sigma^3\sqrt{n}},$$

где $\mu = \mathbf{E}|X_1 - \mathbf{E}X_1|^3$.

5. Условие $\mathbf{E}X_1^2 < \infty$ здесь существенно. А вот требование независимости можно ослабить, допуская небольшую зависимость. Утверждение ЦПТ сохранится в силе при этом. Точно так же можно допустить, что слагаемые могут быть неодинаково распределены, хотя все равно определенные ограничения на их распределения нужно накладывать: нельзя допускать, чтобы одно или несколько слагаемых сильно выделялись на фоне других. Разумеется, точных формулировок мы здесь не даем.

6. Пусть S_n — число успехов в схеме Бернулли, p — вероятность успеха в одном испытании. Тогда при $n \rightarrow \infty$

$$\mathbf{P} \left(A \leq \frac{S_n - np}{\sqrt{np(1-p)}} \leq B \right) \rightarrow \frac{1}{\sqrt{2\pi}} \int_A^B e^{-t^2/2} dt.$$

Это утверждение является частным случаем ЦПТ, поскольку S_n здесь равно сумме независимых случайных величин, распределенных по закону Бернулли; $a = p$, $\sigma^2 = p(1-p)$. Исторически данное утверждение появилось раньше и получило название *теоремы Муавра–Лапласа*.

Пример применения ЦПТ. Предположим, что $n = 1000$ раз бросается игральная кость. Обозначим через S_n сумму выпавших очков. Ясно, что

$$\mathbf{P}(1000 \leq S_n \leq 6000) = 1.$$

С вероятностью единица S_n лежит внутри интервала длиной 5000. Вопрос: насколько ли уменьшится размер интервала, если мы захотим уменьшить вероятность до 0.95? Оказывается, более чем в 20 раз. Этот неожиданный результат невозможно предвидеть, а вот применение ЦПТ сразу же приводит нас к нему.

Действительно, S_n есть сумма независимых случайных величин, каждая из которых принимает значения от 1 до 6 с равными вероятностями. Нетрудно вычислить: $a = \mathbf{E}X_1 = 3.5$, $\mathbf{E}X_1^2 = 91/6$, $\sigma^2 = \mathbf{D}X_1 = 35/12$. В силу ЦПТ случайная величина $(S_n - 3500) / \sqrt{1000 \cdot 35/12}$ имеет почти стандартное нормальное распределение (число n велико!), поэтому

$$\mathbf{P}\left(-1.96 < \frac{S_n - 3500}{\sqrt{1000 \cdot 35/12}} < 1.96\right) \simeq \frac{1}{\sqrt{2\pi}} \int_{-1.96}^{1.96} e^{-t^2/2} dt = 0.95.$$

Последнее мы заранее находим из таблиц. Таким образом,

$$\mathbf{P}(|S_n - 3500| < 1.96\sqrt{1000 \cdot 35/12}) \simeq 0.95, \quad 1.96\sqrt{1000 \cdot 35/12} = 105.85 \dots$$

4.4. Приближение Пуассона в схеме Бернулли

Пусть S_n — число успехов в схеме Бернулли. Мы знаем формулы точного распределения S_n :

$$\mathbf{P}(A \leq S_n \leq B) = \sum_{k=A}^B C_n^k p^k (1-p)^{n-k}.$$

Однако на практике возникают ситуации, когда применение точных формул затруднительно из-за того, что n очень велико. В этом случае можно пользоваться формулами нормального приближения в соответствии с теоремой Муавра–Лапласа. В то же время погрешность при использовании нормального приближения может быть неудовлетворительной несмотря на то что n — очень большое число. Так бывает, когда вероятность успеха p очень мала, т. е. успех появляется в испытаниях Бернулли крайне редко. В этом случае лучшее приближение для распределения S_n обеспечивает теорема Пуассона.

Теорема Пуассона. Пусть в схеме Бернулли $n \rightarrow \infty$ и при этом $p = p(n) \rightarrow 0$ так, что $np(n) \rightarrow \lambda$, где λ — некоторое положительное число. Тогда для любого $k = 0, 1, 2, \dots$

$$\mathbf{P}(S_n = k) = C_n^k p^k (1-p)^{n-k} \rightarrow \frac{\lambda^k}{k!} e^{-\lambda}.$$

Доказательство. Обозначим $\lambda_n = np(n)$, тогда $p = \lambda_n/n$ и

$$\begin{aligned} C_n^k p^k (1-p)^{n-k} &= \frac{n(n-1)\dots(n-k+1)}{k!} \left(\frac{\lambda_n}{n}\right)^k \left(1 - \frac{\lambda_n}{n}\right)^{n-k} = \\ &= \frac{n}{n} \frac{n-1}{n} \dots \frac{n-k+1}{n} \frac{\lambda_n^k}{k!} \left(1 - \frac{\lambda_n}{n}\right)^n \left(1 - \frac{\lambda_n}{n}\right)^{-k}. \end{aligned}$$

Выясним, к чему стремятся отдельные выражения из правой части.

$$\frac{n}{n} \frac{n-1}{n} \dots \frac{n-k+1}{n} = \left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \dots \left(1 - \frac{k-1}{n}\right) \rightarrow 1,$$

поскольку каждый множитель стремится к единице, а их фиксированное число. По условию $\lambda_n^k \rightarrow \lambda^k$. Далее, воспользовавшись разложением в окрестности нуля $\ln(1-x) = -x + o(x)$, получим

$$\ln \left(1 - \frac{\lambda_n}{n}\right)^n = n \ln \left(1 - \frac{\lambda_n}{n}\right) = n \left(-\frac{\lambda_n}{n} + o\left(\frac{\lambda_n}{n}\right)\right) = -\lambda_n + o(1) \rightarrow -\lambda,$$

т. е.

$$\left(1 - \frac{\lambda_n}{n}\right)^n \rightarrow e^{-\lambda}.$$

И наконец,

$$\left(1 - \frac{\lambda_n}{n}\right)^{-k} \rightarrow 1,$$

в силу того что $\lambda_n/n \rightarrow 0$. Теорема доказана.

Эта теорема используется при решении задач следующим образом. Поскольку при $n \rightarrow \infty$ и $np \rightarrow \lambda$

$$\mathbf{P}(S_n = k) \rightarrow \frac{\lambda^k}{k!} e^{-\lambda}$$

и одновременно

$$\frac{(np)^k}{k!} e^{-np} \rightarrow \frac{\lambda^k}{k!} e^{-\lambda},$$

то

$$\mathbf{P}(S_n = k) \simeq \frac{(np)^k}{k!} e^{-np}.$$

Этим приближением обычно и пользуются. Несмотря на то что теорема доказана при условии, что число k фиксировано, сумма левых частей по любому множеству индексов может быть приближена суммой правых частей по тому же множеству индексов. Точность приближения характеризуется следующей оценкой (дается без доказательства).

Теорема. Для любого подмножества $B \subset \{0, 1, 2, \dots\}$

$$\left| \sum_{k \in B} \mathbf{P}(S_n = k) - \sum_{k \in B} \frac{(np)^k}{k!} e^{-np} \right| \leq \min(p, np^2).$$

Пример. Имеется производство спичек. Каждая спичка независимо от других с вероятностью 0.015 является бракованной и при употреблении не возгорается. В соответствии с требованиями стандарта спички должны расфасовываться в коробки по 100 штук в каждую. Ясно, что при этом в каждой коробке с большой вероятностью годных спичек окажется меньше 100. Чтобы избежать претензий со стороны потребителей, руководство решает класть в каждую коробку добавочно некоторое число x спичек так, чтобы с вероятностью не менее 0.95 годных спичек там оказалось не менее 100.

Какое наименьшее число x спичек нужно для этого положить в коробку?

Мы имеем здесь схему Бернулли с числом испытаний $n = 100 + x$ и вероятностью успеха 0.015. Обозначим число бракованных спичек S_n . Тогда годных спичек будет

в коробке не менее 100, если $S_n \leq x$. Из приведенной выше оценки заключаем, что приближение Пуассона дает в нашем случае вполне удовлетворительную точность. Считая для простоты, что $np = (100 + x)0.015 \simeq 1.5$, получаем соотношение

$$\mathbf{P}(S_n \leq x) = \sum_{k=0}^x \mathbf{P}(S_n = k) \simeq e^{-1.5} \left(1 + 1.5 + \frac{1.5^2}{2} + \dots + \frac{1.5^x}{x!} \right).$$

Требуется, чтобы эта вероятность была не менее 0.95. Нетрудно вычислить, что для этого достаточно взять $x = 4$ в правой части.

Часть II.

Математическая статистика

5. Введение

5.1. Основные понятия

Представим себе, что n раз производится некоторый случайный эксперимент и каждый раз по результатам эксперимента мы измеряем какую-то характеристику. Результатом наших измерений (или наблюдений) будет совокупность из n случайных величин X_1, X_2, \dots, X_n , которые будем называть *наблюдениями*. Мы предполагаем, что результаты отдельных экспериментов не влияют друг на друга, поэтому считаем наблюдения независимыми. Кроме того, они будут предполагаться одинаково распределенными в соответствии с некоторой функцией распределения F , поскольку повторяется один и тот же эксперимент. Случайный вектор $X = (X_1, X_2, \dots, X_n)$ называется *выборкой*. Будем использовать обозначение $X \in F$. Это означает, что каждое наблюдение $X_i \in F$. При этом обычно употребляются слова «выборка из распределения F », хотя на самом деле F относится к отдельным наблюдениям.

Значения случайного вектора X принадлежат \mathbb{R}^n — это пространство здесь будет называться *выборочным*.

В задачах теории вероятностей, как правило, мы оперировали известными функциями распределения случайных величин, строили по ним и изучали различные числовые характеристики, исследовали предельное поведение распределения сумм большого числа слагаемых и т. д.

В математической статистике будут рассматриваться другие задачи. В качестве исходного материала мы располагаем выборкой $X = (X_1, X_2, \dots, X_n)$ — она получена в результате n -кратного повторения эксперимента. Однако распределение F , которому подчиняются наблюдения, неизвестно полностью или частично, в разных задачах по-разному. Задача состоит в том, чтобы уменьшить неопределенность наших знаний о функции распределения F , основываясь на информации, заложенной в выборке.

Разумеется, если бы мы знали в полной мере, как устроена, скажем, случайная величина X_1 , то и ее распределение нашли бы в соответствии с определением $F(y) = \mathbf{P}(X_1 < y)$. Однако проблема в том, что на практике результатом наблюдений является всего лишь набор чисел, называемый, кстати, тоже выборкой. К нему надо относиться как к одному конкретному значению вектора X (или как к n независимым реализациям одной случайной величины). По одному значению вектора X (или по n значениям одной случайной величины) полностью восстановить распределение F невозможно. Любые выводы будут носить вероятностный характер.

В этой ситуации математическая статистика предлагает целый ряд процедур и методов, позволяющих с максимально возможной точностью восстановить недостающие знания о функции распределения F .

Традиционно рассматриваются три блока задач.

Первый из них посвящен *оцениванию неизвестных параметров*. Здесь предполагается, что распределение выборки зависит от неизвестного параметра θ . Наша единственная цель состоит в определении этого параметра. Коль скоро найти значение параметра в точном виде нам не удастся, мы будем довольствоваться его оценками, т. е. приближениями, которые строятся на основе имеющихся наблюдений.

Здесь будут предложены методы, позволяющие сразу получать весьма точные оценки неизвестного параметра.

Другой тип задач состоит в нахождении интервала, в котором с большой вероятностью содержится неизвестный параметр. Этот блок задач называется *интервальным оцениванием*.

И наконец, третий блок — это задачи *проверки гипотез* о неизвестном распределении выборки.

Более подробно эти задачи будут обсуждаться ниже.

5.2. Выборочные характеристики

Начнем с понятия *вариационного ряда*.

Полученные нами наблюдения не обязаны располагаться в возрастающем порядке, хотя это часто бывает полезно. Расположив наблюдения в порядке неубывания, получим упорядоченную выборку

$$X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)},$$

которая называется вариационным рядом. Элементы вариационного ряда называются *порядковыми статистиками*: $X_{(1)}$ — первая порядковая статистика, $X_{(2)}$ — вторая порядковая статистика и т. д. Вообще, в математической статистике принято функции от выборки называть *статистиками*.

Если известна функция распределения F выборки, то без труда можно найти распределение порядковых статистик. Например:

$$\mathbf{P}(X_{(n)} < y) = \mathbf{P}(X_1 < y, X_2 < y, \dots, X_n < y) = F^n(y).$$

Далее введем понятие *эмпирической функции распределения*.

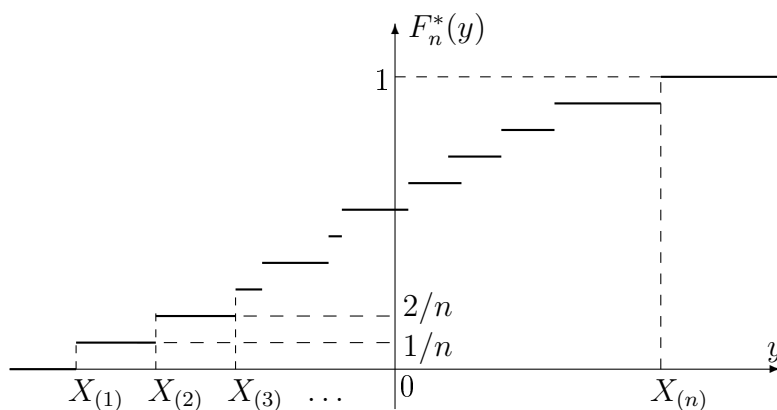
Определение. Эмпирической функцией распределения называется

$$F_n^*(y) = \frac{\nu(y)}{n},$$

где $\nu(y)$ — число наблюдений X_i таких, что $X_i < y$. Другими словами,

$$F_n^*(y) = \begin{cases} 0, & y \leq X_{(1)}, \\ \frac{k}{n}, & X_{(k)} < y \leq X_{(k+1)}, \quad k = 1, \dots, n-1, \\ 1, & y > X_{(n)}. \end{cases}$$

Эмпирическая функция распределения по своему типу является дискретной, она имеет скачки, равные $1/n$, во всех точках X_i .



Таким образом, эмпирическая функция распределения сама является случайной, поскольку ее скачки происходят в случайных точках.

Пусть $X \in F$. Оказывается тогда, что если увеличивать число наблюдений, то эмпирическая функция распределения $F_n^*(y)$ будет приближаться к $F(y)$. Об этом говорится в следующем утверждении.

Теорема Гливенко–Кантелли. Пусть $X \in F$, тогда при $n \rightarrow \infty$

$$\sup_y |F_n^*(y) - F(y)| \xrightarrow{P} 0.$$

Доказательство. Мы докажем это утверждение в ослабленном варианте, без знака \sup , т. е. покажем, что для любого y при $n \rightarrow \infty$

$$F_n^*(y) \xrightarrow{P} F(y).$$

Введем вспомогательные случайные величины

$$Z_i = \begin{cases} 1, & X_i < y, \\ 0, & \text{иначе,} \end{cases}, \quad i = 1, \dots, n.$$

Случайные величины Z_1, \dots, Z_n независимы, поскольку являются функциями от независимых наблюдений, и одинаково распределены по закону Бернулли:

$$\mathbf{P}(Z_i = 1) = \mathbf{P}(X_i < y) = F(y), \quad \mathbf{P}(Z_i = 0) = 1 - F(y), \quad i = 1, \dots, n.$$

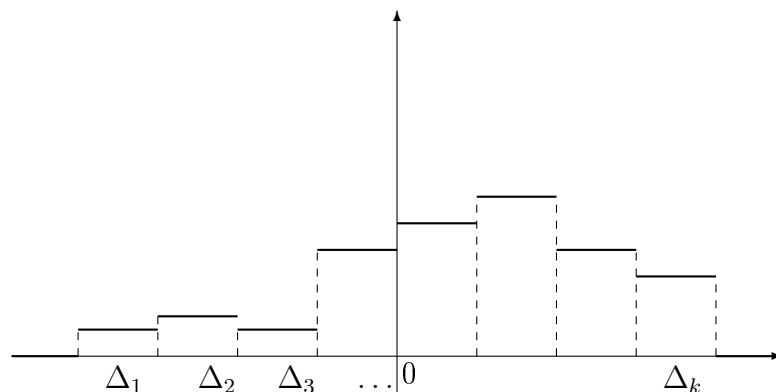
Кроме того, $Z_1 + \dots + Z_n = \nu(y)$. Поэтому по закону больших чисел

$$F_n^*(y) = \frac{Z_1 + \dots + Z_n}{n} \xrightarrow{P} \mathbf{E}Z_1 = F(y),$$

что и требовалось доказать.

Эта теорема лежит в основе многих статистических выводов. Она дает нам право использовать при больших n эмпирическую функцию распределения F_n^* вместо неизвестной теоретической (или истинной) функции распределения F .

По выборке можно также строить приближения для неизвестной плотности распределения, если она существует. Простейшее из них называется *гистограммой*. Для ее построения возьмем какой-нибудь отрезок $[a, b]$, содержащий все наблюдения, и разобьем его на k равных по длине отрезков $\Delta_1, \Delta_2, \dots, \Delta_k$. Пусть длина каждого из этих малых отрезков равна h . На каждом отрезке Δ_i , как на основании, построим прямоугольник с высотой, равной ν_i/nh , где ν_i — число наблюдений, попавших в Δ_i , $i = 1, \dots, k$, $\nu_1 + \dots + \nu_k = n$.



Верхний контур этой фигуры, составленный из горизонтальных отрезков, и есть гистограмма. Попробуем понять, что она в некотором смысле приближает неизвестную плотность f распределения наблюдений. Если функция f непрерывна, то при больших n в силу закона больших чисел

$$\frac{\nu_i}{n} \simeq \mathbf{P}(X_1 \in \Delta_i) = \int_{\Delta_i} f(t) dt \simeq f(t')h,$$

где $t' \in \Delta_i$ — некоторая средняя точка. Иначе говоря, ν_i/nh приближенно равняется значению плотности f в некоторой внутренней точке отрезка Δ_i .

Гистограмма тем самым является ступенчатой функцией. Если же плотность f непрерывна, то, как известно, лучший результат дает приближение ее непрерывными функциями. Мы можем модифицировать гистограмму, соединив отрезками прямых середины горизонтальных отрезков. Полученная таким способом ломаная будет уже графиком непрерывной функции, она называется *полигоном частот*. Здесь середины крайних отрезков (соответствующих Δ_1 и Δ_k) соединяются с осью абсцисс так, чтобы по-прежнему суммарная площадь под графиком равнялась единице.

Отметим, что при построении гистограммы допускалось много произвола: ничего не сказано об исходном отрезке $[a, b]$, о количестве и длине малых отрезков Δ_i . Совершенно не используется информация о том, как наблюдения располагаются внутри отрезков Δ_i . Поэтому гистограмма является весьма грубым приближением для плотности распределения выборки и ее рекомендуется использовать только на предварительных этапах обработки статистической информации с последующим применением более точных методов.

Далее введем понятие *выборочных моментов*.

Обозначим сначала $a_k = \mathbf{E}X_1^k$, $k = 1, \dots$, моменты порядка k для распределения F выборки (разумеется, если они существуют). Если F неизвестно, то, значит, и его моменты нам недоступны. Однако, как мы видели, распределение F можно приблизить эмпирическим распределением F_n^* .

Возникает вопрос: нельзя ли неизвестные нам моменты a_1, a_2, \dots теоретического распределения хорошо приблизить моментами, вычисленными по эмпирической функции распределения F_n^* ? Оказывается, можно.

При фиксированной выборке эмпирическое распределение табличным способом задается так:

Значения	X_1	X_2	X_3	\dots
Вероятности	$1/n$	$1/n$	$1/n$	\dots

Следовательно, для эмпирического распределения момент порядка k должен начинаться по формуле

$$a_k^* = \frac{1}{n} \sum_{i=1}^n X_i^k.$$

Мы будем называть его *выборочным моментом* в отличие от теоретического. Разумеется, выборочные моменты являются случайными величинами (так как строятся из наблюдений) и существуют для любого k . Выборочный момент первого порядка обозначается чаще всего

$$\bar{X} = \frac{X_1 + \dots + X_n}{n}$$

и называется *выборочным средним*. Центральный выборочный момент второго порядка называется *выборочной дисперсией* и обозначается

$$S^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - (\bar{X})^2 = a_2^* - (a_1^*)^2.$$

Эти обозначения будут часто использоваться в дальнейшем.

6. Оценивание неизвестных параметров

6.1. Постановка задачи. Несмещенность и состоятельность

Пусть имеется выборка X (с этого начинается любая статистическая задача). В этом разделе мы будем предполагать, что $X \in F_\theta$, т. е. распределение выборки зависит от некоторого параметра θ , который нам неизвестен и который мы хотим оценить по выборке. Параметр может быть как одномерным, так и многомерным.

Определение. *Оценкой* неизвестного параметра θ называется любая функция от выборки $\theta^* = g(X_1, \dots, X_n)$, в том или ином смысле приближающая θ .

Если $\theta \in \mathbb{R}^k$, то и $g: \mathbb{R}^n \rightarrow \mathbb{R}^k$.

Среди распределений, рассматривавшихся нами ранее в качестве примеров, почти все обладали одним или двумя параметрами.

Разумеется, в одной и той же ситуации можно построить бесконечно много различных оценок. Нам же хочется иметь хорошую оценку. Что это значит?

Рассмотрим несколько желательных свойств оценок.

Определение. Оценка θ^* параметра θ называется несмещенной, если $\mathbf{E}\theta^* = \theta$.

Это значит, что в среднем значение оценки совпадает со значением параметра, который она и призвана оценивать.

Может возникнуть вопрос: как же мы будем вычислять $\mathbf{E}\theta^* = \mathbf{E}g(X_1, \dots, X_n)$, если распределение наблюдений зависит от неизвестного нам параметра?

Мы здесь рассуждаем так: предположим, что истинное значение параметра равно θ , после чего начинаем вычислять $\mathbf{E}\theta^*$ (чтобы подчеркнуть это предположение, символы математического ожидания, дисперсии и вероятности часто снабжают индексом θ : $\mathbf{E}_\theta\theta^*$). Если в итоге этих вычислений мы вновь получим θ , это и будет означать несмещенность оценки.

Определение. Оценка θ^* параметра θ называется асимптотически несмещенной, если $\mathbf{E}\theta^* \rightarrow \theta$ при $n \rightarrow \infty$.

Величина $\mathbf{E}\theta^* - \theta$ называется *смещением* оценки.

Асимптотической несмещенностью довольствуются, как правило, в тех случаях, когда обычной несмещенности достичь не удастся или же если смещение настолько мало при больших n , что им можно пренебречь.

Моменты распределения тоже могут рассматриваться как неизвестные параметры. Для их оценивания мы предполагали пользоваться выборочными моментами. Нетрудно видеть, что выборочные моменты являются несмещенными оценками для моментов теоретических:

$$\mathbf{E}a_k^* = \frac{1}{n}(\mathbf{E}X_1^k + \dots + \mathbf{E}X_n^k) = \frac{na_k}{n} = a_k.$$

В том числе $\mathbf{E}\bar{X} = a_1$, запомним это. Заодно вычислим дисперсию для \bar{X} :

$$\mathbf{D}\bar{X} = \frac{1}{n^2}(\mathbf{D}X_1 + \dots + \mathbf{D}X_n) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n},$$

где $\sigma^2 = \mathbf{D}X_1 = a_2 - a_1^2$.

Оказывается, выборочная дисперсия S^2 не будет являться несмещенной оценкой для σ^2 . Действительно,

$$\mathbf{E}S^2 = \mathbf{E}a_2^* - \mathbf{E}(\bar{X})^2 = a_2 - (\mathbf{D}\bar{X} + (\mathbf{E}\bar{X})^2) = a_2 - \left(\frac{\sigma^2}{n} + a_1^2\right) = \sigma^2 - \frac{\sigma^2}{n}.$$

Оценка оказалась асимптотически несмещенной. Если n велико, то смещением $-\sigma^2/n$ можно пренебречь. Можно поступить по-другому: вместо S^2 использовать оценку

$$S_0^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{n}{n-1} S^2.$$

В этом случае

$$\mathbf{E}S_0^2 = \frac{n}{n-1} \mathbf{E}S^2 = \sigma^2,$$

т. е. оценка S_0^2 является несмещенной.

Определение. Оценка θ^* одномерного параметра θ называется состоятельной, если $\theta^* \xrightarrow{P} \theta$ при $n \rightarrow \infty$.

Состоятельность означает, что при увеличении объема выборки (т. е. при накоплении все большего объема информации) значение оценки должно сближаться с оцениваемым значением параметра. Так и должно быть по логике вещей. Если этого не происходит, то оценка плоха, неразумна. Не рекомендуется пользоваться несостоятельными оценками!

В силу закона больших чисел, выборочные моменты являются состоятельными оценками моментов настоящих:

$$a_k^* = \frac{1}{n} \sum_{i=1}^n X_i^k \xrightarrow{P} \mathbf{E}X_1^k = a_k.$$

Из изученных ранее свойств сходимости по вероятности вытекает, что S^2 и S_0^2 обе являются состоятельными оценками для дисперсии. В самом деле, положим $g(a_1, a_2) = a_2 - a_1^2$; эта функция непрерывна всюду на плоскости, поэтому

$$S^2 = a_2^* - (a_1^*)^2 = g(a_1^*, a_2^*) \xrightarrow{P} g(a_1, a_2) = \sigma^2,$$

$$S_0^2 = \frac{n}{n-1} S^2 = S^2 + \frac{1}{n-1} S^2 \xrightarrow{P} \sigma^2.$$

Далее мы изучим два метода построения весьма точных оценок для неизвестных параметров распределения выборки.

6.2. Метод моментов

Пусть $X \in F_\theta$ и $\theta = (\theta_1, \dots, \theta_k)$ — неизвестный векторный параметр распределения. Применение метода моментов сводится к двум этапам.

Первый этап. Выражаем $\theta_1, \dots, \theta_k$ через моменты a_1, a_2, \dots распределения. В итоге получаем, например, такие соотношения:

$$\begin{aligned} \theta_1 &= g_1(a_1, a_2, \dots, a_k), \\ \theta_2 &= g_2(a_1, a_2, \dots, a_k), \\ &\dots \\ \theta_k &= g_k(a_1, a_2, \dots, a_k). \end{aligned}$$

Чаще всего именно через первые k моментов можно выразить все неизвестные параметры. Если это не удастся сделать, то берутся любые другие моменты, лишь бы через них выражались все $\theta_1, \dots, \theta_k$.

Поскольку распределение выборки зависит от неизвестных параметров, то и моменты a_1, a_2, \dots неизбежно будут от них зависеть. Другими словами, пока что мы

выразили одни неизвестные величины через другие. Однако для моментов нам уже известны хорошие оценки — выборочные моменты. Поэтому переходим ко второму этапу.

Второй этап. Заменяем в полученных соотношениях моменты a_1, a_2, \dots, a_k на выборочные моменты $a_1^*, a_2^*, \dots, a_k^*$. Тем самым получим оценки по методу моментов (ММ-оценки):

$$\begin{aligned}\theta_1^* &= g_1(a_1^*, a_2^*, \dots, a_k^*), \\ \theta_2^* &= g_2(a_1^*, a_2^*, \dots, a_k^*), \\ &\dots \\ \theta_k^* &= g_k(a_1^*, a_2^*, \dots, a_k^*).\end{aligned}$$

Замечания

1. В одной и той же ситуации методом моментов можно получать разные оценки, потому что первый этап можно реализовывать по-разному. Например, если $X \subset \Pi_\lambda$, то, с одной стороны, $\lambda = a_1$, поэтому $\lambda^* = a_1^* = \bar{X}$. Если же на первом этапе воспользоваться формулой $\lambda = a_2 - a_1^2$, то придем к другой оценке: $\lambda_1^* = a_2^* - (a_1^*)^2 = S^2$.

2. Если возникают затруднения при реализации первого этапа, то можно сначала выполнить действия, скажем, на нулевом этапе: найти моменты распределения a_1, a_2, \dots, a_k . Получится набор соотношений вида

$$\begin{cases} a_1 = h_1(\theta_1, \theta_2, \dots, \theta_k), \\ a_2 = h_2(\theta_1, \theta_2, \dots, \theta_k), \\ \dots \\ a_k = h_k(\theta_1, \theta_2, \dots, \theta_k). \end{cases}$$

После чего нужно разрешить эту систему уравнений относительно $\theta_1, \dots, \theta_k$ — тем самым получим нужные нам формулы для первого этапа.

Пример. Пусть $X \in \Gamma_{\alpha, \lambda}$. Найдём ММ-оценки α^*, λ^* .

Начнём с нулевого этапа. Для моментов гамма-распределения имеем

$$\begin{aligned}a_k &= \frac{\alpha^\lambda}{\Gamma(\lambda)} \int_0^\infty t^k t^{\lambda-1} e^{-\alpha t} dt = \frac{1}{\alpha^k \Gamma(\lambda)} \int_0^\infty y^{k+\lambda-1} e^{-y} dy = \\ &= \frac{\Gamma(k+\lambda)}{\alpha^k \Gamma(\lambda)} = \frac{(k+\lambda-1)(k+\lambda-2) \dots (\lambda+1)\lambda \Gamma(\lambda)}{\alpha^k \Gamma(\lambda)} = \\ &= \frac{(k+\lambda-1) \dots (\lambda+1)\lambda}{\alpha^k}.\end{aligned}$$

Поэтому

$$\begin{cases} a_1 = \frac{\lambda}{\alpha}, \\ a_2 = \frac{\lambda(\lambda+1)}{\alpha^2}. \end{cases}$$

Выражаем из первого уравнения $\alpha = \frac{\lambda}{a_1}$ и подставляем во второе:

$$a_2 = \frac{\lambda(\lambda+1)}{\lambda^2} a_1^2 = \left(1 + \frac{1}{\lambda}\right) a_1^2.$$

Отсюда получаем соотношения первого этапа

$$\lambda = \frac{a_1^2}{a_2 - a_1^2}, \quad \alpha = \frac{a_1}{a_2 - a_1^2}.$$

Следовательно,

$$\lambda^* = \frac{(\bar{X})^2}{S^2}, \quad \alpha^* = \frac{\bar{X}}{S^2}.$$

Теорема. Пусть $\theta = g(a_1, \dots, a_k)$ — одномерный параметр распределения выборки. Предположим, что функция g непрерывна в точке (a_1, \dots, a_k) . Тогда $\theta^* = g(a_1^*, \dots, a_k^*)$ является состоятельной оценкой для θ .

Доказательство. В силу сходимости $a_i^* \xrightarrow{P} a_i$, $i = 1, \dots, k$, данное утверждение вытекает из свойства 2 сходимости по вероятности.

Вернемся к рассмотренному примеру, где $X \in \Gamma_{\alpha, \lambda}$. Поскольку $a_2 - a_1^2 = \mathbf{D}X_1 > 0$, то функции

$$g_1(y_1, y_2) = \frac{y_1}{y_2 - y_1^2}, \quad g_2(y_1, y_2) = \frac{y_1^2}{y_2 - y_1^2}$$

непрерывны в точке (a_1, a_2) , $\alpha = g_1(a_1, a_2)$, $\lambda = g_2(a_1, a_2)$. Значит, полученные нами ММ-оценки состоятельны.

Свойство несмещенности проверяется в каждом случае по-своему, обычно метод моментов приводит к несмещенным или асимптотически несмещенным оценкам.

6.3. Метод максимального правдоподобия

Пусть, как и ранее, $X \in F_\theta$ и $\theta \in \mathbb{R}$ — неизвестный параметр, подлежащий оценке.

1. Дискретный случай. Попробуем пояснить основную идею метода на примере. Пусть, стреляя 10 раз по мишени в тире, мы трижды попали и 7 раз промахнулись. Мы не знаем, какова вероятность p попадания при одном выстреле, можем строить лишь различные предположения об этом. Рассмотрим три из них:

$$1) p = 0.01; \quad 2) p = 0.3; \quad 3) p = 0.9.$$

Какое из них выглядит более правдоподобным после того, как стрельба завершена? Разумеется, второе. Конечно, при каждом из этих предположений мы могли бы 7 раз промахнуться и 3 раза попасть, но вероятность такого результата стрельбы будет наибольшей при $p = 0.3$.

Эти соображения и легли в основу метода максимального правдоподобия.

Предположим сначала, что распределение F_θ дискретно, и обозначим $f(\theta, t) = \mathbf{P}(X_1 = t)$. Имеет смысл рассматривать здесь только те значения t , для которых эти вероятности положительны. Пусть, далее, для $\mathbf{t} = (t_1, \dots, t_n)$

$$f(\theta, \mathbf{t}) = \mathbf{P}(X_1 = t_1, \dots, X_n = t_n) = \prod_{i=1}^n f(\theta, t_i)$$

— вероятность того, что выборка примет конкретное значение (t_1, \dots, t_n) . Коль скоро в результате наших экспериментов реализовалась выборка X , то, подставив ее в функцию f , получим $f(\theta, X)$, что при фиксированном значении выборки равняется вероятности ее появления.

Функция $f(\theta, X)$ называется *функцией правдоподобия*.

Идея метода состоит в следующем: мы подбираем такое значение θ , при котором вероятность получить нашу выборку максимальна. Другими словами, мы подбираем наиболее правдоподобное с точки зрения полученного результата значение параметра.

Аналитически это означает, что мы должны исследовать на максимум функцию правдоподобия и взять в качестве оценки метода максимального правдоподобия (ММП-оценки) то значение θ^* , при котором

$$f(\theta^*, X) = \max_{\theta} f(\theta, X).$$

Поскольку $f(\theta, X) = \prod_{i=1}^n f(\theta, X_i)$, то в ряде случаев исследовать эту функцию на максимум удобнее, предварительно взяв от нее логарифм:

$$l(\theta, X) = \ln f(\theta, X) = \sum_{i=1}^n \ln f(\theta, X_i).$$

Функция $l(\theta, X)$ называется *логарифмической функцией правдоподобия*. Точки максимума у $l(\theta, X)$ и $f(\theta, X)$ совпадают, а с суммой работать удобнее, чем с произведением.

Если производная по θ существует и непрерывна, то точку экстремума можно найти из уравнения

$$\frac{\partial l(\theta, X)}{\partial \theta} = 0.$$

Убедившись, что в найденной точке действительно достигается максимум, а не минимум, мы тем самым находим ММП-оценку как решение данного уравнения.

Пример. Пусть $X \in \Pi_{\lambda}$. Тогда $f(\lambda, t) = \frac{\lambda^t}{t!} e^{-\lambda}$ и для функции правдоподобия имеем

$$f(\lambda, X) = \prod_{i=1}^n \frac{\lambda^{X_i}}{X_i!} e^{-\lambda} = \frac{\lambda^{X_1 + \dots + X_n}}{X_1! \dots X_n!} e^{-n\lambda}.$$

Как функция переменной $\lambda > 0$, это выражение равно степенной функции, умноженной на экспоненту в отрицательной степени. Эта функция дифференцируема по λ сколь угодно раз, и равенство нулю первой производной приведет нас к точке максимума. Для удобства найдем

$$l(\lambda, X) = (X_1 + \dots + X_n) \ln \lambda - n\lambda - \ln(X_1! \dots X_n!).$$

Далее находим точку максимума:

$$\frac{\partial l(\lambda, X)}{\partial \lambda} = \frac{X_1 + \dots + X_n}{\lambda} - n = 0, \quad \lambda^* = \frac{X_1 + \dots + X_n}{n} = \bar{X}.$$

Пусть теперь функция распределения F_{θ} абсолютно непрерывна. Обозначим $f(\theta, t)$ соответствующую ей плотность распределения. Пусть для $\mathbf{t} = (t_1, \dots, t_n)$

$$f(\theta, \mathbf{t}) = \prod_{i=1}^n f(\theta, t_i)$$

— плотность распределения случайного вектора X .

По аналогии с дискретным случаем функцией правдоподобия будем называть $f(\theta, X)$. ММП-оценкой называется то значение $\theta = \theta^*$, которое максимизирует функцию правдоподобия:

$$f(\theta^*, X) = \max_{\theta} f(\theta, X).$$

Как и ранее, можно ввести логарифмическую функцию правдоподобия $l(\theta, X) = \ln f(\theta, X)$ и работать с ней.

Пример. Пусть $X \in E_\alpha$. Поскольку все наблюдения при таком условии положительны, то имеем

$$f(\alpha, X) = \prod_{i=1}^n \alpha e^{-\alpha X_i} = \alpha^n e^{-\alpha(X_1 + \dots + X_n)}.$$

Ясно, что дифференцирование приведет нас к точке максимума.

$$l(\alpha, X) = n \ln \alpha - \alpha(X_1 + \dots + X_n);$$

$$\frac{\partial l(\alpha, X)}{\partial \alpha} = \frac{n}{\alpha} - (X_1 + \dots + X_n) = 0, \quad \alpha^* = \frac{n}{X_1 + \dots + X_n} = \frac{1}{\bar{X}}.$$

Замечания

1. Если $\theta = (\theta_1, \dots, \theta_k)$, то все остается по-прежнему, только исследовать на максимум функцию правдоподобия нужно будет как функцию k переменных. Например, поиск максимума с помощью дифференцирования приведет к системе уравнений

$$\begin{cases} \frac{\partial l(\theta, X)}{\partial \theta_1} = 0, \\ \dots \\ \frac{\partial l(\theta, X)}{\partial \theta_k} = 0. \end{cases}$$

2. Если функция правдоподобия достигает максимального значения в нескольких точках, то все они, по определению, считаются ММП-оценками.

3. ММП-оценки, как правило, являются асимптотически несмещенными и состоятельными.

4. Во многих случаях ММ-оценки совпадают с ММП-оценками, но это происходит не всегда.

Пример. Пусть $X \in U_{0,\theta}$, и θ неизвестно. Поскольку

$$a_k = \int_0^\theta t^k \frac{1}{\theta} dt = \frac{\theta^k}{k+1},$$

то мы получаем целую последовательность ММ-оценок:

$$\theta = ((k+1)a_k)^{1/k}, \quad \theta_k^* = ((k+1)a_k^*)^{1/k}, \quad k = 1, 2, \dots$$

В то же время ММП-оценка будет иной. Построим функцию правдоподобия. Для $t \in \mathbb{R}$

$$f(\theta, t) = \begin{cases} \frac{1}{\theta}, & \text{если } 0 \leq t \leq \theta, \\ 0, & \text{иначе.} \end{cases}$$

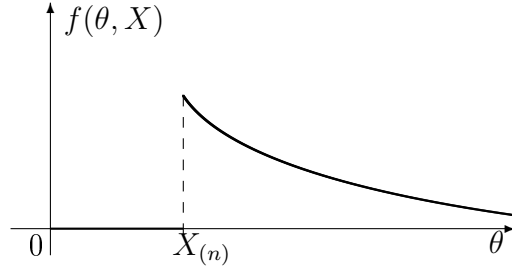
Поэтому

$$f(\theta, X) = \begin{cases} \frac{1}{\theta^n}, & \text{если } 0 \leq X_i \leq \theta \text{ при всех } i = 1, \dots, n, \\ 0, & \text{иначе.} \end{cases}$$

Так как все $X_i \geq 0$, то можно переписать

$$f(\theta, X) = \begin{cases} \frac{1}{\theta^n}, & \text{если } \max(X_1, \dots, X_n) = X_{(n)} \leq \theta, \\ 0, & \text{если } X_{(n)} > \theta. \end{cases}$$

Построим график зависимости функции правдоподобия от θ .



Ясно, что максимум достигается в точке $\theta^* = X_{(n)}$ — это и будет ММП-оценкой.

Данный пример служит напоминанием о том, что для нахождения точки максимума не стоит спешить с дифференцированием функции. Переход к логарифмической функции правдоподобия здесь также неуместен.

6.4. Сравнение оценок

Пусть $X \in F_\theta$, $\theta \in \mathbb{R}$ — неизвестный параметр, и пусть мы уже построили две различные оценки θ_1^* и θ_2^* . Обе оказались хорошими: например, несмещенными (асимптотически несмещенными) и состоятельными. Какую из них предпочесть?

Мы до сих пор не учитывали еще одно важное свойство оценок. Чем меньше разброс значений оценки относительно неизвестного параметра, тем она точнее. Естественно из двух оценок выбирать ту, у которой этот разброс меньше. Разброс можно характеризовать по-разному. Мы будем использовать среднее квадратическое отклонение.

Определение. Будем считать, что оценка θ_1^* лучше, чем θ_2^* , если при всех значениях θ

$$\mathbf{E}_\theta(\theta_1^* - \theta)^2 \leq \mathbf{E}_\theta(\theta_2^* - \theta)^2$$

и хотя бы при одном значении θ неравенство является строгим.

Если оценка θ^* несмещенная, то $\mathbf{E}_\theta(\theta^* - \theta)^2 = \mathbf{D}_\theta \theta^*$. Следовательно, из двух несмещенных оценок лучше та, у которой дисперсия меньше.

Если среди всех несмещенных оценок (а их бесконечно много) найдется та, у которой дисперсия минимальна, то она называется *эффективной*.

Эффективные оценки являются наиболее точными, к их отысканию и нужно стремиться. Делать это непросто, хотя соответствующие методы разработаны. Из-за дефицита времени мы упомянем только один подход, позволяющий во многих случаях определять, является ли данная оценка эффективной. В математической статистике известно *неравенство Рао–Крамера*, которое гласит: если $f(\theta, t)$ как функция переменной θ обладает определенными свойствами регулярности (мы не уточняем здесь — какими), то для любой несмещенной оценки θ^*

$$\mathbf{D}_\theta \theta^* \geq C(\theta, n).$$

Константа $C(\theta, n)$ легко вычисляется, после чего с ней можно сравнивать дисперсии оценок. Если для некоторой оценки в неравенстве Рао–Крамера выполняется равенство, значит, эта оценка обладает наименьшей дисперсией, в силу неравенства меньше уже не бывает.

Пример. Вернемся к ситуации, рассмотренной выше: $X \in U_{0,\theta}$, и θ неизвестно. Сравним ММ-оценку $\theta_1^* = 2\bar{X}$ и ММП-оценку $\theta^* = X_{(n)}$. Первая из них является несмещенной:

$$\mathbf{E}(2\bar{X}) = 2\mathbf{E} X_1 = 2 \frac{\theta}{2} = \theta.$$

Для нее

$$\mathbf{D}\theta_1^* = 4 \frac{\mathbf{D}X_1}{n} = \frac{\theta^2}{3n}.$$

Исследуем вторую оценку.

$$\begin{aligned} F_{\theta^*}(t) &= \mathbf{P}(X_{(n)} < t) = \mathbf{P}(X_1 < t, \dots, X_n < t) = \prod_{i=1}^n \mathbf{P}(X_i < t) = \\ &= (U_{0,\theta}(t))^n = \begin{cases} 0, & \text{если } t \leq 0, \\ \frac{t^n}{\theta^n}, & \text{если } 0 < t \leq \theta, \\ 1, & \text{если } t > \theta. \end{cases} \end{aligned}$$

Отсюда, кстати, следует состоятельность оценки θ^* . Поскольку всегда $X_{(n)} \leq \theta$, то для любого $\varepsilon > 0$

$$\mathbf{P}(|X_{(n)} - \theta| \geq \varepsilon) = \mathbf{P}(\theta - X_{(n)} \geq \varepsilon) = \mathbf{P}(X_{(n)} \leq \theta - \varepsilon) = \frac{(\theta - \varepsilon)^n}{\theta^n} \rightarrow 0$$

при $n \rightarrow \infty$.

Далее находим плотность

$$f(\theta, t) = \begin{cases} \frac{nt^{n-1}}{\theta^n}, & \text{если } 0 < t \leq \theta, \\ 0, & \text{иначе} \end{cases}$$

и моменты

$$\mathbf{E}\theta^* = \int_0^\theta t \frac{nt^{n-1}}{\theta^n} dt = \frac{n\theta}{n+1}, \quad \mathbf{E}(\theta^*)^2 = \int_0^\theta t^2 \frac{nt^{n-1}}{\theta^n} dt = \frac{n\theta^2}{n+2}.$$

Оценка θ^* оказалась смещенной; чтобы сравнение было справедливым, подправим ее, сделав, как и θ_1^* , несмещенной. Пусть

$$\theta_2^* = \frac{n+1}{n} \theta^*, \quad \mathbf{E}\theta_2^* = \theta,$$

и дальше сравниваем θ_1^* и θ_2^* . Найдем дисперсию новой оценки.

$$\mathbf{D}\theta_2^* = \frac{(n+1)^2}{n^2} \mathbf{D}\theta^* = \frac{(n+1)^2}{n^2} \left(\frac{n\theta^2}{n+2} - \left(\frac{n\theta}{n+1} \right)^2 \right) = \frac{\theta^2}{n(n+2)}.$$

Мы видим, что $\mathbf{D}\theta_2^* < \mathbf{D}\theta_1^*$ при $n > 1$, причем $\mathbf{D}\theta_2^*$ стремится к нулю с ростом n на порядок быстрее, чем $\mathbf{D}\theta_1^*$.

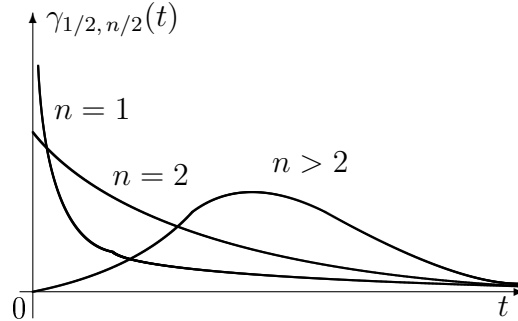
7. Доверительные интервалы

7.1. Некоторые распределения, связанные с нормальным

Нам потребуется ввести некоторые распределения, играющие большую роль в математической статистике.

I. Распределение хи-квадрат. По определению, распределением хи-квадрат с n степенями свободы называется $\chi_n^2 = \Gamma_{1/2, n/2}$. Таким образом, это распределение с плотностью

$$\gamma_{1/2, n/2}(t) = \begin{cases} \frac{1}{2^{n/2}\Gamma(n/2)} t^{\frac{n}{2}-1} e^{-\frac{t}{2}}, & t > 0, \\ 0, & t \leq 0. \end{cases}$$



Свойства распределения хи-квадрат

1. Если случайные величины Z_1 и Z_2 независимы, $Z_1 \in \chi_{n_1}^2$, $Z_2 \in \chi_{n_2}^2$, то $Z_1 + Z_2 \in \chi_{n_1+n_2}^2$.

Это свойство в свое время было доказано в общем виде для свертки плотностей гамма-распределения.

2. Пусть случайные величины Y_1, \dots, Y_n независимы и $Y_i \in \Phi_{0,1}$ при всех $i = 1, \dots, n$. Тогда

$$Y_1^2 + \dots + Y_n^2 \in \chi_n^2.$$

Доказательство. Достаточно проверить утверждение для $n = 1$ и затем воспользоваться предыдущим свойством. Имеем $F_{Y_1^2}(y) = \mathbf{P}(Y_1^2 < y) = 0$ для $y \leq 0$. Если $y > 0$, то

$$\begin{aligned} F_{Y_1^2}(y) &= \mathbf{P}(Y_1^2 < y) = \mathbf{P}(-\sqrt{y} < Y_1 < \sqrt{y}) = \frac{1}{\sqrt{2\pi}} \int_{-\sqrt{y}}^{\sqrt{y}} \exp\left\{-\frac{u^2}{2}\right\} du = \\ &= \frac{2}{\sqrt{2\pi}} \int_0^{\sqrt{y}} \exp\left\{-\frac{u^2}{2}\right\} du. \end{aligned}$$

Сделав замену $t = u^2$, получим

$$F_{Y_1^2}(y) = \frac{1}{\sqrt{2\pi}} \int_0^y t^{-1/2} \exp\left\{-\frac{t}{2}\right\} dt = \Gamma_{1/2, 1/2}(y).$$

3. При больших значениях n распределение $\chi_n^2(y)$ можно аппроксимировать нормальным.

Действительно, пользуясь обозначениями предыдущего утверждения, находим

$$\chi_n^2(y) = \mathbf{P}(Y_1^2 + \dots + Y_n^2 < y) = \mathbf{P}\left(\frac{Y_1^2 + \dots + Y_n^2 - n\mathbf{E}Y_1^2}{\sqrt{n\mathbf{D}Y_1^2}} < \frac{y - n\mathbf{E}Y_1^2}{\sqrt{n\mathbf{D}Y_1^2}}\right).$$

Если n велико, то можно применить ЦПТ. Здесь $\mathbf{E}Y_1^2 = 1$, $\mathbf{E}Y_1^4 = 3$ (последнее предлагается проверить в качестве самостоятельного упражнения), поэтому $\mathbf{D}Y_1^2 = 2$. Получаем

$$\chi_n^2(y) \simeq \Phi_{0,1} \left(\frac{y - n}{\sqrt{2n}} \right).$$

II. Распределение Стьюдента. Пусть случайные величины Y и Z_n независимы, $Y \in \Phi_{0,1}$, $Z_n \in \chi_n^2$. Тогда распределение T_n случайной величины

$$\frac{Y}{\sqrt{\frac{1}{n} Z_n}}$$

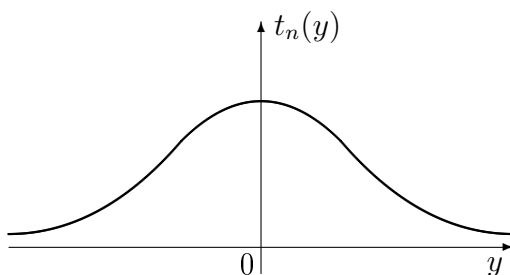
называется распределением Стьюдента с n степенями свободы.

Вместо случайной величины Z_n в этом определении можно поставить $Y_1^2 + \dots + Y_n^2$, где Y, Y_1, \dots, Y_n независимы и $Y_i \in \Phi_{0,1}$ при всех $i = 1, \dots, n$.

Плотность распределения Стьюдента равна (приводится без доказательства)

$$t_n(y) = \frac{\Gamma((n+1)/2)}{\sqrt{\pi n} \Gamma(n/2)} \left(1 + \frac{y^2}{n} \right)^{-(n+1)/2}.$$

Это симметричная кривая, по своему виду напоминающая график плотности распределения Коши (кстати, распределение Коши совпадает с распределением Стьюдента с одной степенью свободы).



Данное распределение введено английским математиком В. С. Госсетом; он подписывал свои работы псевдонимом Student, что и закрепилось в названии распределения.

Если $n \rightarrow \infty$, то $t_n(y) \rightarrow \varphi_{0,1}(y)$. Этот результат нетрудно получить из явной формулы для $t_n(y)$, воспользовавшись замечательным пределом. Кстати, из закона больших чисел и свойств сходимости по вероятности следует, что

$$\frac{Y}{\sqrt{\frac{Y_1^2 + \dots + Y_n^2}{n}}} \xrightarrow{P} Y \in \Phi_{0,1},$$

поскольку

$$\frac{Y_1^2 + \dots + Y_n^2}{n} \xrightarrow{P} \mathbf{E}Y_1^2 = 1.$$

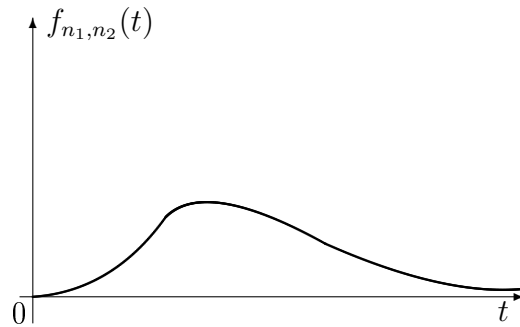
Здесь, напомним, случайные величины Y, Y_1, \dots, Y_n независимы и все имеют стандартное нормальное распределение.

III. Распределение Фишера. Пусть Z_1 и Z_2 независимы, $Z_1 \in \chi_{n_1}^2$, $Z_2 \in \chi_{n_2}^2$. Тогда распределение случайной величины

$$\frac{Z_1/n_1}{Z_2/n_2}$$

называется распределением Фишера с числом степеней свободы (n_1, n_2) и обозначается F_{n_1, n_2} .

Мы не будем приводить формулу для плотности этого распределения, ее график в своих общих чертах выглядит так:



Все три введенных распределения широко используются в статистических вычислениях, поэтому почти во всех пособиях по математической статистике можно встретить таблицы значений этих функций распределения. Наиболее полно таблицы представлены в работе Большев Л.Н., Смирнов Н.В. Таблицы математической статистики. М., 1965.

7.2. Свойства выборок из нормального распределения

Следующая лемма лежит в основе многих статистических выводов.

Лемма Фишера. Пусть случайные величины X_1, \dots, X_n независимы, $X_i \in \Phi_{0,1}$, $i = 1, \dots, n$, и

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \dots \\ Y_n \end{pmatrix} = A \begin{pmatrix} X_1 \\ X_2 \\ \dots \\ X_n \end{pmatrix},$$

где A — ортогональная матрица. Тогда для любого $r = 1, \dots, n-1$

$$\sum_{i=1}^n X_i^2 - Y_1^2 - \dots - Y_r^2 \in \chi_{n-r}^2$$

и эта случайная величина не зависит от Y_1, \dots, Y_r .

Доказательство. Ранее было установлено, что в наших условиях случайные величины Y_1, \dots, Y_n независимы и имеют распределение $\Phi_{0,1}$. Ортогональное преобразование не меняет длины вектора: $\sum_{i=1}^n X_i^2 = \sum_{i=1}^n Y_i^2$, поэтому

$$\sum_{i=1}^n X_i^2 - Y_1^2 - \dots - Y_r^2 = Y_{r+1}^2 + \dots + Y_n^2 \in \chi_{n-r}^2,$$

и эта случайная величина не зависит от Y_1, \dots, Y_r . Лемма доказана.

Теорема о свойствах выборок из нормального распределения. Пусть $X = (X_1, \dots, X_n) \in \Phi_{\alpha, \sigma^2}$. Тогда

$$1) \frac{\bar{X} - \alpha}{\sigma} \sqrt{n} \in \Phi_{0,1};$$

$$2) \frac{nS^2}{\sigma^2} \in \chi_{n-1}^2;$$

3) \bar{X} и S^2 независимы.

Доказательство. 1. Ранее установлено, что $\bar{X} \in \Phi_{\alpha, \sigma^2/n}$. Применяем операцию стандартизации:

$$\frac{\bar{X} - \mathbf{E}\bar{X}}{\sqrt{\mathbf{D}\bar{X}}} = \frac{\bar{X} - \alpha}{\sigma} \sqrt{n} \in \Phi_{0,1}.$$

2. Обозначим $Z_i = \frac{X_i - \alpha}{\sigma}$, $i = 1, \dots, n$. Тогда $Z_i \in \Phi_{0,1}$, $\frac{\bar{X} - \alpha}{\sigma} = \bar{Z}$ и

$$\begin{aligned} \frac{S^2}{\sigma^2} &= \frac{1}{n} \sum_{i=1}^n \left(\frac{X_i - \bar{X}}{\sigma} \right)^2 = \frac{1}{n} \sum_{i=1}^n \left(\frac{X_i - \alpha}{\sigma} - \frac{\bar{X} - \alpha}{\sigma} \right)^2 = \\ &= \frac{1}{n} \sum_{i=1}^n (Z_i - \bar{Z})^2 = \frac{1}{n} \sum_{i=1}^n Z_i^2 - (\bar{Z})^2. \end{aligned}$$

Поэтому

$$\frac{nS^2}{\sigma^2} = \sum_{i=1}^n Z_i^2 - (\sqrt{n} \bar{Z})^2.$$

В свою очередь,

$$\sqrt{n} \bar{Z} = \left(\frac{1}{\sqrt{n}}, \dots, \frac{1}{\sqrt{n}} \right) \begin{pmatrix} Z_1 \\ Z_2 \\ \dots \\ Z_n \end{pmatrix}.$$

Вектор $\left(\frac{1}{\sqrt{n}}, \dots, \frac{1}{\sqrt{n}} \right)$ имеет единичную длину. Его всегда можно достроить до ортогональной матрицы A , в которой он будет являться первой строкой. Тогда $\sqrt{n} \bar{Z}$ будет совпадать с первой компонентой вектора $A(Z_1, \dots, Z_n)^T$ и по лемме Фишера

$$\sum_{i=1}^n Z_i^2 - (\sqrt{n} \bar{Z})^2 \in \chi_{n-1}^2.$$

Из леммы Фишера следует также, что $\frac{nS^2}{\sigma^2}$ и $\sqrt{n} \bar{Z} = \frac{\bar{X} - \alpha}{\sigma} \sqrt{n}$ независимы. Следовательно, независимы S^2 и \bar{X} как функции этих величин.

Теорема доказана.

Следствие. В условиях теоремы

$$\frac{\bar{X} - \alpha}{S} \sqrt{n-1} \in T_{n-1}.$$

Доказательство. Согласно определению, T_{n-1} — это распределение дроби $Y/\sqrt{Z/(n-1)}$, где Y и Z независимы, $Y \in \Phi_{0,1}$ и $Z \in \chi_{n-1}^2$. В силу теоремы мы можем взять

$$Y = \frac{\bar{X} - \alpha}{\sigma} \sqrt{n}, \quad Z = \frac{nS^2}{\sigma^2}.$$

Следствие доказано.

7.3. Доверительные интервалы для параметров нормального распределения

Пусть $X \in F_\theta$, $\theta \in \mathbb{R}$ — неизвестный параметр. Ранее мы занимались поиском подходящих оценок для θ , что можно назвать также точечным оцениванием, поскольку вместо неизвестной точки θ на прямой предлагалось использовать другую, случайную точку θ^* . В этом разделе мы будем поступать по-другому: постараемся указать интервал, содержащий точку θ с большой вероятностью.

Определение. Доверительным интервалом уровня $1 - \varepsilon$ для неизвестного параметра θ называется интервал $(A(X_1, \dots, X_n), B(X_1, \dots, X_n))$ такой, что

$$\mathbf{P}(A(X_1, \dots, X_n) < \theta < B(X_1, \dots, X_n)) \geq 1 - \varepsilon.$$

Обычно в качестве ε выбирают достаточно малое число.

Доверительный интервал называется асимптотическим, если

$$\lim_{n \rightarrow \infty} \mathbf{P}(A(X_1, \dots, X_n) < \theta < B(X_1, \dots, X_n)) \geq 1 - \varepsilon.$$

Разумеется, пользоваться асимптотическим доверительным интервалом следует только при больших объемах выборки.

Отметим, что доверительный интервал — это интервал со случайными концами, коль скоро они строятся по выборке. Ясно, что интервал тем лучше, чем он уже.

Далее мы займемся построением доверительных интервалов для неизвестных параметров в случае выборки из нормального распределения.

Пусть $X = (X_1, \dots, X_n) \in \Phi_{\alpha, \sigma^2}$.

1. Доверительный интервал для α при условии, что σ^2 известно. Мы установили ранее, что $\frac{\bar{X} - \alpha}{\sigma} \sqrt{n} \in \Phi_{0,1}$. С помощью таблиц стандартного нормального распределения можно найти число $q > 0$ такое, что $\Phi_{0,1}(-q) = \varepsilon/2$. Это значит, что

$$\mathbf{P}\left(-q < \frac{\bar{X} - \alpha}{\sigma} \sqrt{n} < q\right) = \Phi_{0,1}(q) - \Phi_{0,1}(-q) = 1 - \varepsilon$$

или после преобразований

$$\mathbf{P}\left(\bar{X} - \frac{q\sigma}{\sqrt{n}} < \alpha < \bar{X} + \frac{q\sigma}{\sqrt{n}}\right) = 1 - \varepsilon.$$

Тем самым мы построили доверительный интервал $\left(\bar{X} - \frac{q\sigma}{\sqrt{n}}, \bar{X} + \frac{q\sigma}{\sqrt{n}}\right)$, его длина равна $2q\sigma/\sqrt{n}$. Это значит, что при больших n мы можем довольно точно локализовать значение неизвестного параметра α .

2. Доверительный интервал для α при условии, что σ^2 неизвестно. Предыдущая конструкция не годится, поскольку в ней участвует неизвестный параметр σ . Здесь поможет следствие из теоремы предыдущего раздела, согласно которого случайная величина $\frac{\bar{X} - \alpha}{S} \sqrt{n-1}$ (уже не зависящая от σ) распределена по закону Стьюдента с $n-1$ степенью свободы. Теперь мы воспользуемся таблицами распределения T_{n-1} и найдем число q такое, что $T_{n-1}(-q) = \varepsilon/2$. Тогда

$$\mathbf{P}\left(-q < \frac{\bar{X} - \alpha}{S} \sqrt{n-1} < q\right) = T_{n-1}(q) - T_{n-1}(-q) = 1 - \varepsilon,$$

и после преобразований получаем доверительный интервал

$$\mathbf{P} \left(\bar{X} - \frac{qS}{\sqrt{n-1}} < \alpha < \bar{X} + \frac{qS}{\sqrt{n-1}} \right) = 1 - \varepsilon.$$

3. Доверительный интервал для σ^2 при условии, что α известно. Случайные величины $\frac{X_i - \alpha}{\sigma}$, $i = 1, \dots, n$, независимы и имеют стандартное нормальное распределение, поэтому

$$\sum_{i=1}^n \left(\frac{X_i - \alpha}{\sigma} \right)^2 \in \chi_n^2.$$

Из таблиц распределения χ_n^2 найдем числа q_1 и q_2 такие, что $\chi_n^2(q_1) = \varepsilon/2$, $\chi_n^2(q_2) = 1 - \varepsilon/2$. Тогда

$$\mathbf{P} \left(q_1 < \sum_{i=1}^n \frac{(X_i - \alpha)^2}{\sigma^2} < q_2 \right) = \chi_n^2(q_2) - \chi_n^2(q_1) = 1 - \varepsilon.$$

Это соотношение эквивалентно тому, что

$$\mathbf{P} \left(\frac{\sum_{i=1}^n (X_i - \alpha)^2}{q_2} < \sigma^2 < \frac{\sum_{i=1}^n (X_i - \alpha)^2}{q_1} \right) = 1 - \varepsilon.$$

4. Доверительный интервал для σ^2 при условии, что α неизвестно. Воспользуемся тем, что $\frac{nS^2}{\sigma^2} \in \chi_{n-1}^2$. Из таблиц распределения χ_{n-1}^2 находим числа q_1 и q_2 такие, что $\chi_{n-1}^2(q_1) = \varepsilon/2$, $\chi_{n-1}^2(q_2) = 1 - \varepsilon/2$. Тогда

$$\mathbf{P} \left(q_1 < \frac{nS^2}{\sigma^2} < q_2 \right) = \chi_{n-1}^2(q_2) - \chi_{n-1}^2(q_1) = 1 - \varepsilon$$

и после преобразований получаем доверительный интервал

$$\mathbf{P} \left(\frac{nS^2}{q_2} < \sigma^2 < \frac{nS^2}{q_1} \right) = 1 - \varepsilon.$$

7.4. Построение доверительных интервалов с помощью нормального приближения

В предыдущем разделе нам удалось построить точные доверительные интервалы для параметров нормального распределения, пользоваться которыми можно при любых значениях n . К сожалению, в общем случае для параметров других распределений таких хороших конструкций нет. Однако для параметров многих распределений удастся построить асимптотические доверительные интервалы. Мы приведем здесь краткое описание этой конструкции, опуская строгое обоснование некоторых выводов.

Пусть, как и ранее, $X \in F_\theta$, $\theta \in \mathbb{R}$ — неизвестный параметр, для которого будет строиться доверительный интервал. Предположим, что этот параметр можно выразить через один из моментов распределения: $\theta = g(a_k)$, $k \geq 1$, и пусть функция g дифференцируема и $g'(a_k) \neq 0$. Рассмотрим ММ-оценку $\theta^* = g(a_k^*)$. В силу близости

точек a_k^* и a_k при больших n , можно воспользоваться линейным приближением в соответствии с формулой Тейлора:

$$\theta^* = g(a_k^*) \simeq g(a_k) + (a_k^* - a_k)g'(a_k)$$

или

$$\frac{\theta^* - \theta}{g'(a_k)} \simeq \frac{1}{n} \sum_{i=1}^n X_i^k - a_k = \frac{\sum_{i=1}^n X_i^k - na_k}{n}.$$

Предположим дополнительно, что существует a_{2k} , и обозначим $\sigma^2 = \mathbf{D}X_1^k = a_{2k} - a_k^2$. Умножим обе части полученного соотношения на \sqrt{n} и поделим на σ . Тогда

$$\frac{\theta^* - \theta}{\sigma g'(a_k)} \sqrt{n} \simeq \frac{\sum_{i=1}^n X_i^k - na_k}{\sigma \sqrt{n}}.$$

При больших значениях n распределение правой части близко к $\Phi_{0,1}$ в силу ЦПТ. Пусть число q таково, что $\Phi_{0,1}(-q) = \varepsilon/2$. Тогда

$$\mathbf{P} \left(\left| \frac{\theta^* - \theta}{\sigma g'(a_k)} \sqrt{n} \right| < q \right) \simeq \Phi_{0,1}(q) - \Phi_{0,1}(-q) = 1 - \varepsilon.$$

Как и ранее, нам нужно разрешить это неравенство относительно θ . Однако существует трудность: стоящая в знаменателе величина $\sigma|g'(a_k)|$ также неизбежно зависит от θ : $\sigma|g'(a_k)| = h(\theta)$. Предположим дополнительно, что h — непрерывная функция, тогда $h(\theta^*) \simeq h(\theta)$. Поэтому при больших n

$$\mathbf{P} \left(\frac{|\theta^* - \theta|}{h(\theta^*)} \sqrt{n} < q \right) \simeq 1 - \varepsilon.$$

Тем самым мы получаем асимптотический доверительный интервал

$$\mathbf{P} \left(\theta^* - \frac{qh(\theta^*)}{\sqrt{n}} < \theta < \theta^* + \frac{qh(\theta^*)}{\sqrt{n}} \right) \simeq 1 - \varepsilon.$$

Пример. Пусть $X \in \Pi_\lambda$. Здесь $\lambda = a_1$, то есть можно взять $k = 1$, $g(y) = y$. Тогда $h(\lambda) = \sigma g'(a_k) = \sqrt{\lambda}$, и мы получаем для параметра λ асимптотический доверительный интервал

$$\left(\bar{X} - \frac{q\sqrt{\bar{X}}}{\sqrt{n}}, \bar{X} + \frac{q\sqrt{\bar{X}}}{\sqrt{n}} \right).$$

8. Проверка гипотез

8.1. Постановка задачи, основные понятия

Пусть $X \in F$ и распределение F неизвестно. В этой ситуации естественно строить различные предположения, или гипотезы, относительно F . Гипотезы будем обозначать H_1, H_2, \dots . Гипотеза называется *простой*, если она однозначно определяет распределение выборки. Все остальные гипотезы называются *сложными*.

Например, $H_1 : X \in \Phi_{0,1}$ — простая гипотеза, $H_2 : X \in \Phi_{\alpha, \sigma^2}$ — сложная, если значения α и σ^2 не конкретизированы.

Чаще всего выдвигаются две взаимоисключающие друг друга гипотезы H_1 и H_2 , одна из которых, по нашему предположению, верна, только мы не знаем, какая именно. Первую из них, H_1 , называют *основной* гипотезой, а вторую — *конкурирующей*

гипотезой или *альтернативой*. Мы должны одну из гипотез принять и тем самым отвергнуть другую — в этом состоит наше решение. В дальнейшем решение будем формулировать относительно основной гипотезы H_1 , поскольку это однозначно определяет наши действия относительно альтернативы.

Нам необходимо вооружиться правилом, в соответствии с которым по выборке сразу же можно было бы определить, принимается H_1 или нет. Такое правило называется *критерием*. Построение критерия означает, что все возможные значения выборки разбиваются на две категории или, что то же самое, выборочное пространство \mathbb{R}^n нужно разбить на две части:

$$\mathbb{R}^n = \mathbf{K} \cup \overline{\mathbf{K}}.$$

Если $X \in \mathbf{K}$, то гипотеза H_1 отвергается, если $X \in \overline{\mathbf{K}}$, то принимается. Множество \mathbf{K} называется *критическим*, его задание полностью определяет критерий.

Ситуации, которые могут возникнуть при принятии нами решения, отражены в представленной ниже таблице.

	Верна H_1	Верна H_2
Принимаем H_1	Хорошо	Плохо
Принимаем H_2	Плохо	Хорошо

Мы видим, что существуют две нежелательные ситуации, когда верна одна гипотеза, а мы принимаем другую в соответствии с выбранным критерием. Как правило, избежать подобных ошибок не удастся. Выход в следующем: нужно использовать такие критерии, для которых вероятности принятия ошибочных решений малы.

В дальнейшем будем использовать обозначение $\mathbf{P}_i(A)$, если вычисляется вероятность при условии, что верна гипотеза H_i .

Предположим, что проверяется простая гипотеза $H_1 : F = F_1$ против простой альтернативы $H_2 : F = F_2$. Тогда вероятность отвергнуть верную (основную) гипотезу $\beta_1 = \beta_1(\mathbf{K}) = \mathbf{P}_1(X \in \mathbf{K})$ называется *вероятностью ошибки первого рода*. Аналогично, вероятность принять неверную гипотезу $\beta_2 = \mathbf{P}_2(X \in \overline{\mathbf{K}})$ называется *вероятностью ошибки второго рода*. Число $1 - \beta_2$ называется *мощностью* критерия.

Вычисление вероятности ошибочного решения при справедливости сложной гипотезы, как правило, невозможно: мы ведь не знаем, каким конкретно является распределение выборки.

Далее мы рассмотрим некоторые *критерии согласия*. Они строятся для проверки гипотез вида

$$H_1 : F = F_1 \quad \text{против} \quad H_2 : F \neq F_1$$

(т. е. мы должны проверить, согласуются ли данные наблюдений с предположением о том, что $X \in F_1$). Будем требовать, чтобы для рассматриваемых критериев вероятность ошибки первого рода была мала: $\beta_1 \leq \varepsilon$ для заранее выбранного малого числа ε . В таких случаях говорят, что критерий имеет уровень $1 - \varepsilon$. Часто приходится довольствоваться асимптотическим критерием уровня $1 - \varepsilon$, то есть если $\lim_{n \rightarrow \infty} \beta_1 \leq \varepsilon$.

Поскольку конкурирующая гипотеза является сложной, то вероятность ошибки второго рода мы рассматривать не будем.

8.2. Критерий Колмогорова

Критерий основывается на следующей теореме (приводится без доказательства).

Теорема Колмогорова. Пусть $X \in F$ и F непрерывна. Обозначим

$$D_n = \sup_y |F_n^*(y) - F(y)|.$$

Тогда для любого $y > 0$ при $n \rightarrow \infty$

$$\mathbf{P}(\sqrt{n}D_n < y) \rightarrow \mathcal{K}(y) = \sum_{m=-\infty}^{\infty} (-1)^m e^{-2m^2 y^2}.$$

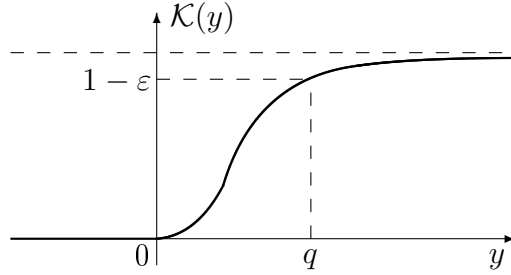
Функция распределения $\mathcal{K}(y)$ называется *функцией Колмогорова*, она абсолютно непрерывна; для нахождения ее значений имеются таблицы.

Перейдем к построению критерия.

Пусть $X \in F$ и проверяются гипотезы $H_1 : F = F_1$ против $H_2 : F \neq F_1$, где F_1 непрерывна. Наша задача — построить асимптотический критерий уровня $1 - \varepsilon$. Для начала вычислим величину D_n в предположении, что верна гипотеза H_1 , т. е. $F = F_1$:

$$D_n = \sup_y |F_n^*(y) - F_1(y)|.$$

В силу теоремы Колмогорова, при больших n функция распределения случайной величины $\sqrt{n}D_n$ мало отличается от $\mathcal{K}(y)$, поэтому заранее по таблицам функции Колмогорова мы можем найти такое число $q > 0$, что $\mathcal{K}(q) = 1 - \varepsilon$.



Следовательно, если верна H_1 , то $\mathbf{P}_1(\sqrt{n}D_n < q) \simeq \mathcal{K}(q) = 1 - \varepsilon$. Поэтому мы будем отвергать гипотезу H_1 , если окажется, что $\sqrt{n}D_n \geq q$, т. е. если расхождение между эмпирической и гипотетической функциями распределения достаточно велико.

Ясно, что при этом

$$\beta_1 = \mathbf{P}_1(\sqrt{n}D_n \geq q) = 1 - \mathbf{P}_1(\sqrt{n}D_n < q) \simeq 1 - \mathcal{K}(q) = \varepsilon.$$

Критическое множество для построенного нами критерия выглядит так:

$$\mathbf{K} = \{(X_1, \dots, X_n) \in \mathbb{R}^n : \sqrt{n}D_n \geq q\}.$$

8.3. Критерий хи-квадрат Пирсона

Пусть $X \in F$ и проверяются гипотезы $H_1 : F = F_1$ против $H_2 : F \neq F_1$. По-прежнему наша задача состоит в построении асимптотического критерия уровня $1 - \varepsilon$. В предположении, что $X \in F_1$, разобьем область возможных значений X_1 на некоторое количество непересекающихся промежутков:

$$\mathbf{P}_1(X_1 \in \Delta_1 \cup \dots \cup \Delta_k) = 1,$$

где Δ_i имеет вид $\Delta_i = [a_i, b_i)$, $i = 1, \dots, k$.

Пусть ν_i — число наблюдений, попавших в Δ_i , $i = 1, \dots, k$, $\nu_1 + \dots + \nu_k = n$. Обозначим также

$$p_i = \mathbf{P}_1(X_1 \in \Delta_i) = F_1(b_i) - F_1(a_i), \quad i = 1, \dots, k.$$

Из закона больших чисел следует, что

$$\frac{\nu_i}{n} \xrightarrow{P} p_i, \quad n \rightarrow \infty,$$

при каждом i , если верна H_1 . В качестве меры близости совокупностей $\{\nu_1/n, \dots, \nu_k/n\}$ и $\{p_1, \dots, p_k\}$ предлагается использовать величину

$$\Psi_n = n \sum_{i=1}^k \frac{1}{p_i} \left(\frac{\nu_i}{n} - p_i \right)^2 = \sum_{i=1}^k \frac{(\nu_i - np_i)^2}{np_i}.$$

Теорема Пирсона. Если $0 < p_i < 1$ при всех $i = 1, \dots, k$, то для любого $y > 0$

$$\mathbf{P}_1(\Psi_n < y) \rightarrow \chi_{k-1}^2(y), \quad n \rightarrow \infty.$$

Доказательство этой теоремы весьма сложно, и по этой причине мы его не приводим.

Займемся построением критерия. Найдем число q такое, что $\chi_{k-1}^2(q) = 1 - \varepsilon$. Если верна гипотеза H_1 , то с вероятностью, близкой к $1 - \varepsilon$, значение случайной величины Ψ_n должно быть меньше q . Поэтому мы отвергаем гипотезу, если $\Psi_n \geq q$, и принимаем ее в противном случае. Это значит, что мы принимаем H_1 , если нет явного противоречия этой гипотезы с наблюдаемыми значениями. Критическое множество выглядит следующим образом:

$$\mathbf{K} = \{(X_1, \dots, X_n) : \Psi_n \geq q\}.$$

Для вероятности ошибки первого рода имеем

$$\beta_1 = \mathbf{P}_1(\Psi_n \geq q) = 1 - \mathbf{P}_1(\Psi_n < q) \simeq 1 - \chi_{k-1}^2(q) = \varepsilon.$$

Замечание. Приближение $\mathbf{P}_1(\Psi_n < q) \simeq \chi_{k-1}^2(q)$ является вполне удовлетворительным для практических целей, если $np_i \geq 10$ для всех i . В противном случае следует укрупнить разбиение (например, объединить два соседних интервала в один).

Пример (данные взяты из книги: Крамер Г. Математические методы статистики. М., Мир, 1975). В Швеции в 1935 году родились 88273 человека. Известны их дни рождения. Нужно проверить гипотезу о том, что день рождения произвольно взятого человека с равными вероятностями может приходиться на любой день года.

Перенумеруем от 1 до 365 все дни 1935-го года и пусть X_i — номер дня рождения i -го человека в соответствии с этой шкалой. Мы имеем выборку X_1, \dots, X_n , где $n = 88\,273$, и тогда в соответствии с основной гипотезой H_1

$$\mathbf{P}_1(X_1 = k) = \frac{1}{365}, \quad k = 1, \dots, 365.$$

Чтобы применить критерий χ^2 , воспользуемся естественным разбиением года по месяцам ($k = 12$):

$$\Delta_1 = [1, 31] \text{ (январь)}, \quad \Delta_2 = [32, 59] \text{ (февраль)}, \quad \dots$$

Данные по месяцам приведены в таблице.

i	1	2	3	4	5	6	7	8	9	10	11	12
ν_i	7280	6957	7883	7884	7892	7609	7585	7393	7203	6903	6552	7132
p_i	$\frac{31}{365}$	$\frac{28}{365}$	\dots									
s_i	6.29	12.27	19.85	54.48	20.79	17.24	1.03	1.45	0.38	47.09	68.18	17.79

Здесь обозначено $s_i = \frac{(\nu_i - np_i)^2}{np_i}$. Получаем

$$\Psi_n = \sum_{i=1}^k s_i = 266.84.$$

Если взять $\varepsilon = 0.05$, то из таблиц находим $\chi_{11}^2(19.7) = 0.95$, т. е. $\Psi_n > q = 19.7$ с большим запасом. Гипотезу о равных вероятностях следует отвергнуть. Кстати, тот же вывод следует и для меньших значений ε , так как из тех же таблиц следует, к примеру, что $\chi_{11}^2(45) = 0.99999\dots$

8.4. Построение критерия с помощью доверительного интервала

Предположим, что $X \in F_\theta$, $\theta \in \mathbb{R}$ — неизвестный параметр. Наша задача состоит в проверке основной гипотезы $H_1: \theta = \theta_1$ против $H_2: \theta \neq \theta_1$.

Если мы имеем доверительный интервал для θ уровня $1 - \varepsilon$ (точный или асимптотический), то с его помощью можно построить критерий согласия (также точный или асимптотический) уровня $1 - \varepsilon$. Действительно, если при всех значениях θ

$$\mathbf{P}_\theta(A(X_1, \dots, X_n) < \theta < B(X_1, \dots, X_n)) \geq 1 - \varepsilon,$$

то и при $\theta = \theta_1$ должно быть

$$\mathbf{P}_{\theta_1}(A(X_1, \dots, X_n) < \theta_1 < B(X_1, \dots, X_n)) \geq 1 - \varepsilon.$$

Поэтому мы отвергаем H_1 , если $\theta_1 \notin (A(X_1, \dots, X_n), B(X_1, \dots, X_n))$, поскольку такое событие имеет малую вероятность (не больше ε) при справедливости H_1 . Критическое множество выглядит так:

$$\mathbf{K} = \{(X_1, \dots, X_n) : \theta_1 \notin (A(X_1, \dots, X_n), B(X_1, \dots, X_n))\}.$$

8.5. Проверка гипотез в случае двух выборок

В этом разделе мы будем предполагать, что проведены две серии независимых испытаний, в результате которых имеем две независимые выборки

$$X = (X_1, \dots, X_n) \in F$$

и

$$Y = (Y_1, \dots, Y_m) \in G.$$

Чаще всего проверяется основная гипотеза о совпадении распределений $F = G$. В этом случае критерии называются *критериями однородности*. В других ситуациях проверяется гипотеза о совпадении только некоторых параметров распределений F и G . С таких задач мы и начнем.

Заметим предварительно, что теперь мы имеем $n+m$ наблюдений, следовательно, выборочным пространством будет \mathbb{R}^{n+m} и критическое множество \mathbf{K} будет $n+m$ -мерным.

Итак, пусть сначала

$$X = (X_1, \dots, X_n) \in \Phi_{\alpha_1, \sigma_1^2}, \quad Y = (Y_1, \dots, Y_m) \in \Phi_{\alpha_2, \sigma_2^2}.$$

Все четыре параметра неизвестны.

1. Проверка гипотезы о совпадении дисперсий. Здесь мы проверяем основную гипотезу $H_1: \sigma_1^2 = \sigma_2^2$ против $H_2: \sigma_1^2 \neq \sigma_2^2$. Заранее выберем малое число $\varepsilon > 0$, и пусть

$$\begin{aligned} \bar{X} &= \frac{1}{n} \sum_{i=1}^n X_i, & S_X^2 &= \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2, \\ \bar{Y} &= \frac{1}{m} \sum_{i=1}^m Y_i, & S_Y^2 &= \frac{1}{m} \sum_{i=1}^m (Y_i - \bar{Y})^2. \end{aligned}$$

По теореме о свойствах выборок из нормального распределения

$$\frac{nS_X^2}{\sigma_1^2} \in \chi_{n-1}^2, \quad \frac{mS_Y^2}{\sigma_2^2} \in \chi_{m-1}^2,$$

причем эти случайные величины независимы, поскольку построены по независимым выборкам. Из них можно построить случайную величину, имеющую распределение Фишера:

$$\frac{1}{n-1} \frac{nS_X^2}{\sigma_1^2} : \frac{1}{m-1} \frac{mS_Y^2}{\sigma_2^2} = \frac{n(m-1)\sigma_2^2 S_X^2}{m(n-1)\sigma_1^2 S_Y^2} \in F_{n-1, m-1}.$$

Если верна гипотеза H_1 , т. е. $\sigma_1^2 = \sigma_2^2$, то

$$\eta = \frac{n(m-1)S_X^2}{m(n-1)S_Y^2} \in F_{n-1, m-1}.$$

С помощью таблиц распределения $F_{n-1, m-1}$ можно найти числа q_1 и q_2 такие, что $F_{n-1, m-1}(q_1) = \varepsilon/2$, $F_{n-1, m-1}(q_2) = 1 - \varepsilon/2$. Тогда

$$\mathbf{P}_1(q_1 < \eta < q_2) = F_{n-1, m-1}(q_2) - F_{n-1, m-1}(q_1) = 1 - \varepsilon.$$

Поэтому логично отвергать H_1 , если $\eta \notin (q_1, q_2)$; вероятность этого события равна в точности ε , если верна H_1 . Здесь

$$\mathbf{K} = \{(X_1, \dots, X_n, Y_1, \dots, Y_m) : \eta \notin (q_1, q_2)\}.$$

2. Проверка гипотезы о совпадении средних. Мы будем это делать в предположении, что дисперсии совпадают: $\sigma_1^2 = \sigma_2^2 = \sigma^2$; σ^2 по-прежнему неизвестно. Проверяется гипотеза $H_1: \alpha_1 = \alpha_2$ против $H_2: \alpha_1 \neq \alpha_2$.

Здесь будет использоваться распределение Стьюдента. В силу того что \bar{X} и \bar{Y} независимы и

$$\bar{X} \in \Phi_{\alpha_1, \sigma^2/n}, \quad \bar{Y} \in \Phi_{\alpha_2, \sigma^2/m},$$

имеем

$$\bar{X} - \bar{Y} \in \Phi_{\alpha_1 - \alpha_2, \sigma^2(1/n + 1/m)}$$

и после стандартизации

$$\frac{\bar{X} - \bar{Y} - (\alpha_1 - \alpha_2)}{\sqrt{\sigma^2 \left(\frac{1}{n} + \frac{1}{m} \right)}} \in \Phi_{0,1}.$$

Далее, по свойству распределения хи-квадрат

$$\frac{nS_X^2}{\sigma^2} + \frac{mS_Y^2}{\sigma^2} \in \chi_{n+m-2}^2;$$

эта случайная величина не зависит от $\bar{X} - \bar{Y}$. Таким образом,

$$\frac{\bar{X} - \bar{Y} - (\alpha_1 - \alpha_2)}{\sigma \sqrt{\frac{1}{n} + \frac{1}{m}}} : \sqrt{\frac{1}{n+m-2} \frac{nS_X^2 + mS_Y^2}{\sigma^2}} \in T_{n+m-2}.$$

Если верна гипотеза H_1 , то $\alpha_1 - \alpha_2 = 0$ и

$$\psi = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{1}{n} + \frac{1}{m}} \sqrt{\frac{nS_X^2 + mS_Y^2}{n+m-2}}} \in T_{n+m-2}.$$

Из таблиц распределения T_{n+m-2} находим число q такое, что $T_{n+m-2}(-q) = \varepsilon/2$. Тогда

$$\mathbf{P}_1(-q < \psi < q) = T_{n+m-2}(q) - T_{n+m-2}(-q) = 1 - \varepsilon.$$

Следовательно, выбрав

$$\mathbf{K} = \{(X_1, \dots, X_n, Y_1, \dots, Y_m) : |\psi| \geq q\},$$

мы будем иметь $\beta_1 = \mathbf{P}_1((X_1, \dots, X_n, Y_1, \dots, Y_m) \in \mathbf{K}) = \varepsilon$.

3. Критерий Колмогорова–Смирнова однородности двух выборок. Пусть

$$X = (X_1, \dots, X_n) \in F, \quad Y = (Y_1, \dots, Y_m) \in G,$$

где F и G — непрерывные функции распределения. Проверяется гипотеза $H_1 : F = G$ против $H_2 : F \neq G$. Мы построим асимптотический критерий уровня $1 - \varepsilon$.

Пусть F_n^* и G_m^* — эмпирические функции распределения, построенные по выборкам X и Y соответственно. Введем

$$D_{n,m} = \sup_y |F_n^*(y) - G_m^*(y)|.$$

Если верна H_1 , то при увеличении объемов выборок эмпирические функции распределения сходятся по вероятности к общему пределу, т. е. $D_{n,m} \xrightarrow{P} 0$. Следующая теорема показывает, с какой скоростью это происходит (приводится без доказательства).

Теорема Колмогорова–Смирнова. Пусть верна гипотеза H_1 и общая функция распределения выборок непрерывна. Тогда для любого $y > 0$ при $n \rightarrow \infty$, $m \rightarrow \infty$

$$\mathbf{P}_1 \left(\sqrt{\frac{nm}{n+m}} D_{n,m} < y \right) \rightarrow \mathcal{K}(y) = \sum_{i=-\infty}^{\infty} (-1)^i e^{-2i^2 y^2}.$$

Пусть q таково, что $\mathcal{K}(q) = 1 - \varepsilon$. Положим

$$\mathbf{K} = \left\{ (X_1, \dots, X_n, Y_1, \dots, Y_m) : \sqrt{\frac{nm}{n+m}} D_{n,m} \geq q \right\},$$

т. е. мы отвергаем гипотезу об однородности, если расхождение между двумя эмпирическими функциями распределения достаточно велико. Тогда при больших n

$$\beta_1 = \mathbf{P}_1 \left(\sqrt{\frac{nm}{n+m}} D_{n,m} \geq q \right) \simeq 1 - \mathcal{K}(q) = \varepsilon.$$

8.6. Дисперсионный анализ: однофакторная модель

Дисперсионный анализ объединяет значительное число задач математической статистики, в которых анализируется влияние тех или иных факторов на конечный результат. Мы рассмотрим здесь простейшую модель, в которой проверяется гипотеза о влиянии одного фактора.

Пусть имеется k независимых выборок

$$\begin{aligned} (X_{11}, X_{12}, \dots, X_{1n_1}) &\in \Phi_{\alpha_1, \sigma^2}, \\ (X_{21}, X_{22}, \dots, X_{2n_2}) &\in \Phi_{\alpha_2, \sigma^2}, \\ &\dots \quad \dots \quad \dots \quad \dots \quad \dots \\ (X_{k1}, X_{k2}, \dots, X_{kn_k}) &\in \Phi_{\alpha_k, \sigma^2}. \end{aligned}$$

Все параметры $\alpha_1, \dots, \alpha_k, \sigma^2$ неизвестны. Проверяются гипотезы

$$H_1 : \alpha_1 = \alpha_2 = \dots = \alpha_k,$$

$$H_2 : \text{существуют индексы } i \neq j \text{ такие, что } \alpha_i \neq \alpha_j.$$

Такая задача может возникнуть, к примеру, в следующей ситуации. Пусть на k станках производится изготовление (или обработка) одинаковых деталей. У каждой изготовленной детали замеряется некий параметр, скажем диаметр. Он является случайной величиной вследствие неизбежных отклонений от стандарта. Мы получаем тем самым k выборок, предполагается, что на i -м станке изготовлено n_i деталей. Гипотеза H_1 утверждает, что не важно, на каком станке изготовлена деталь, фактор станка не играет роли. Это соответствует тому, что средние значения y всех выборок совпадают. В то же время конкурирующая гипотеза объявляет о наличии систематических отклонений для некоторых станков.

Схема наших действий такова: мы будем строить из наблюдений случайную величину, которая при справедливости H_1 распределена по закону Фишера с известным числом степеней свободы. Это и определит в итоге наше решение.

Обозначим

$$N = \sum_{i=1}^k n_i, \quad \bar{X}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij}, \quad \bar{X} = \frac{1}{N} \sum_{i=1}^k \sum_{j=1}^{n_i} X_{ij}.$$

Теорема. Если верна гипотеза H_1 , то

$$\frac{(N-k) \sum_{i=1}^k n_i (\bar{X}_i - \bar{X})^2}{(k-1) \sum_{i=1}^k \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2} \in F_{k-1, N-k}.$$

Доказательство. Предположим на время, что нам известны все параметры $\alpha_1, \dots, \alpha_k, \sigma^2$, и применим к каждому наблюдению стандартизацию. Пусть

$$Y_{ij} = \frac{X_{ij} - \alpha_i}{\sigma} \in \Phi_{0,1}, \quad i = 1, \dots, k, \quad j = 1, \dots, n_i, \quad \bar{Y}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} Y_{ij}.$$

Запишем выражение для выборочной дисперсии, построенной по i -й выборке из стандартизованных наблюдений:

$$\frac{1}{n_i} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2 = \frac{1}{n_i} \sum_{j=1}^{n_i} Y_{ij}^2 - (\bar{Y}_i)^2.$$

Как и при доказательстве теоремы о свойствах выборок из нормального распределения, с помощью леммы Фишера устанавливаем, что

$$\sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2 = \sum_{j=1}^{n_i} Y_{ij}^2 - (\sqrt{n_i} \bar{Y}_i)^2 \in \chi_{n_i-1}^2$$

и эта величина не зависит от \bar{Y}_i . Суммируя левые части по i , получаем

$$Q_1 = \sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2 \in \chi_{N-k}^2.$$

Отметим, что Q_1 не зависит от $\bar{Y}_1, \dots, \bar{Y}_k$.

Введем далее

$$\bar{Y} = \frac{1}{N} \sum_{i=1}^k \sum_{j=1}^{n_i} Y_{ij} = \frac{1}{N} \sum_{i=1}^k n_i \bar{Y}_i.$$

Тогда

$$\begin{aligned} Q_2 &= \sum_{i=1}^k n_i (\bar{Y}_i - \bar{Y})^2 = \sum_{i=1}^k n_i \bar{Y}_i^2 - 2\bar{Y} \sum_{i=1}^k n_i \bar{Y}_i + (\bar{Y})^2 \sum_{i=1}^k n_i = \\ &= \sum_{i=1}^k (\sqrt{n_i} \bar{Y}_i)^2 - 2\bar{Y} N \bar{Y} + N \bar{Y}^2 = \\ &= \sum_{i=1}^k (\sqrt{n_i} \bar{Y}_i)^2 - (\sqrt{N} \bar{Y})^2. \end{aligned}$$

Мы знаем, что $\bar{Y}_i \in \Phi_{0,1/n_i}$, поэтому $\sqrt{n_i} \bar{Y}_i \in \Phi_{0,1}$. Далее,

$$\begin{aligned} \sqrt{N} \bar{Y} &= \frac{\sqrt{N}}{N} \sum_{i=1}^k n_i \bar{Y}_i = \sum_{i=1}^k \frac{\sqrt{n_i}}{\sqrt{N}} \sqrt{n_i} \bar{Y}_i = \\ &= \left(\frac{\sqrt{n_1}}{\sqrt{N}}, \dots, \frac{\sqrt{n_k}}{\sqrt{N}} \right) (\sqrt{n_1} \bar{Y}_1, \dots, \sqrt{n_k} \bar{Y}_k)^T. \end{aligned}$$

Вектор $\left(\frac{\sqrt{n_1}}{\sqrt{N}}, \dots, \frac{\sqrt{n_k}}{\sqrt{N}}\right)$ имеет единичную длину, поэтому его всегда можно достроить до ортогональной матрицы, в которой он будет первой строкой. Воспользовавшись леммой Фишера, получим, что

$$Q_2 = \sum_{i=1}^k (\sqrt{n_i} \bar{Y}_i)^2 - (\sqrt{N} \bar{Y})^2 \in \chi_{k-1}^2.$$

Напомним, что Q_1 и Q_2 независимы, поэтому случайная величина

$$\frac{Q_2/(k-1)}{Q_1/(N-k)}$$

распределена по закону Фишера $F_{k-1, N-k}$.

Вернемся к исходным наблюдениям.

$$Y_{ij} - \bar{Y}_i = \frac{X_{ij} - \alpha_i}{\sigma} - \frac{\bar{X}_i - \alpha_i}{\sigma} = \frac{X_{ij} - \bar{X}_i}{\sigma},$$

поэтому

$$Q_1 = \frac{1}{\sigma^2} \sum_{i=1}^k \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2.$$

Предположим теперь, что H_1 верна, т. е. $\alpha_1 = \alpha_2 = \dots = \alpha_k = \alpha$. Тогда

$$\bar{Y} = \frac{1}{N} \sum_{i=1}^k \sum_{j=1}^{n_i} \left(\frac{X_{ij} - \alpha_i}{\sigma} \right) = \frac{\bar{X} - \alpha}{\sigma}.$$

Поэтому

$$Q_2 = \sum_{i=1}^k n_i \left(\frac{\bar{X}_i - \alpha}{\sigma} - \frac{\bar{X} - \alpha}{\sigma} \right)^2 = \frac{1}{\sigma^2} \sum_{i=1}^k n_i (\bar{X}_i - \bar{X})^2.$$

Таким образом, если верна гипотеза H_1 , то

$$\xi = \frac{Q_2/(k-1)}{Q_1/(N-k)} = \frac{(N-k) \sum_{i=1}^k n_i (\bar{X}_i - \bar{X})^2}{(k-1) \sum_{i=1}^k \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2} \in F_{k-1, N-k}.$$

Теорема доказана.

Перейдем к построению критерия. В полученном выражении для случайной величины ξ именно числитель чувствителен к систематическим отклонениям между выборками, поэтому мы будем реагировать на большие значения ξ . По таблицам распределения $F_{k-1, N-k}$ находим число $q > 0$ такое, что $F_{k-1, N-k}(q) = 1 - \varepsilon$. Иными словами, если верна гипотеза H_1 , то событие $\{\xi \geq q\}$ маловероятно. Поэтому отвергаем H_1 , если $\xi \geq q$, и принимаем ее в противном случае. При этом $\beta_1 = \mathbf{P}_1(\xi \geq q) = \varepsilon$.

9. Задачи линейной регрессии

9.1. Постановка задачи

Предположим, что в результате n -кратного повторения эксперимента мы получаем выборку (Y_1, \dots, Y_n) . Причем известно, что значение наблюдаемой величины Y

линейно зависит от некоторых известных неслучайных числовых факторов x_1, \dots, x_k и еще от некоторого случайного фактора, наличие которого объясняется случайными погрешностями в работе измерительных инструментов или же его присутствие заложено в основе эксперимента. Другими словами,

$$Y = \theta_1 x_1 + \dots + \theta_k x_k + \varepsilon,$$

назовем это основным соотношением. Здесь величины x_1, \dots, x_k могут принимать известные нам значения в каждом эксперименте, неизвестны только коэффициенты зависимости $\theta_1, \dots, \theta_k$. Их определение и составляет основную задачу, и она была бы проста, если бы не мешали случайные отклонения. Проводя эксперименты при тех или иных значениях x_1, \dots, x_k , мы получаем наблюдения

$$Y_1 = \theta_1 x_{11} + \dots + \theta_k x_{1k} + \varepsilon_1,$$

$$Y_2 = \theta_1 x_{21} + \dots + \theta_k x_{2k} + \varepsilon_2,$$

...

$$Y_n = \theta_1 x_{n1} + \dots + \theta_k x_{nk} + \varepsilon_n.$$

Следует проводить $n > k$ наблюдений, иначе мы не сможем хорошо оценить все коэффициенты. Случайные величины $\varepsilon_1, \dots, \varepsilon_n$ предполагаются независимыми и одинаково распределенными, при этом $\mathbf{E}\varepsilon_i = 0$, дисперсия $\mathbf{D}\varepsilon_i = \sigma^2$ чаще всего также предполагается неизвестной.

Запишем полученные соотношения в векторном виде. Пусть

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \dots \\ Y_n \end{pmatrix}, \quad X = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1k} \\ x_{21} & x_{22} & \dots & x_{2k} \\ \dots & \dots & \dots & \dots \\ x_{n1} & x_{n2} & \dots & x_{nk} \end{pmatrix},$$

$$\theta = \begin{pmatrix} \theta_1 \\ \theta_2 \\ \dots \\ \theta_k \end{pmatrix}, \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \dots \\ \varepsilon_n \end{pmatrix},$$

тогда

$$Y = X\theta + \varepsilon.$$

Матрица X называется *регрессором*, она состоит из известных нам чисел, которые мы задаем в ходе проведения эксперимента. Регрессор имеет n строк и k столбцов; его элементы выбираются так, чтобы столбцы были линейно независимыми. Случайный вектор ε неизбежно присутствует в наших соотношениях, но его значения нам неизвестны. Вектор Y называется *откликом*, он состоит из наблюдаемых нами случайных величин. И наконец, θ — вектор неизвестных параметров, которые подлежат оцениванию.

Отметим, что, в отличие от предыдущих рассмотрений, здесь мы имеем дело с выборкой, состоящей из разнораспределенных наблюдений, поскольку

$$\mathbf{E}Y_i = \theta_1 x_{i1} + \dots + \theta_k x_{ik}$$

зависит от i .

Правая часть основного соотношения линейно зависит от неизвестных параметров $\theta_1, \dots, \theta_k$, по этой причине мы говорим о задаче *линейной* регрессии.

Исторически сложившийся термин «регрессия» не отражает сути проблемы, здесь более подошло бы название «статистическое исследование зависимостей».

Частным случаем является следующая постановка задачи. Пусть имеется набор функций $\psi_1(t), \dots, \psi_k(t)$ и основное соотношение выглядит так:

$$Y = \theta_1 \psi_1(t) + \dots + \theta_k \psi_k(t) + \varepsilon.$$

Переменная t может интерпретироваться как время или температура; проводя эксперименты при $t = t_1, \dots, t_n$, мы получаем наблюдения

$$Y_i = \theta_1 \psi_1(t_i) + \dots + \theta_k \psi_k(t_i) + \varepsilon_i, \quad i = 1, \dots, n,$$

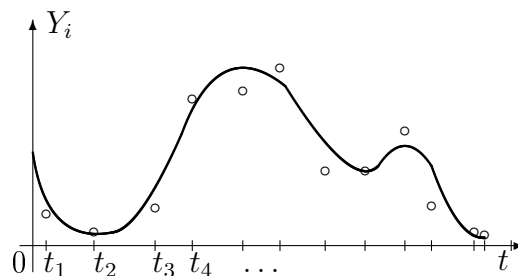
т. е. $x_{ij} = \psi_j(t_i)$. Например, можно взять

$$\psi_1(t) = 1, \quad \psi_2(t) = t, \quad \dots, \quad \psi_k(t) = t^{k-1},$$

и тогда основное соотношение примет вид

$$Y = \theta_1 + \theta_2 t + \dots + \theta_k t^{k-1} + \varepsilon.$$

В этом случае задача получает простую геометрическую интерпретацию.

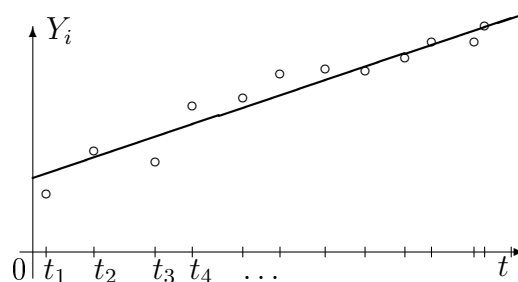


Числа $\theta_1, \dots, \theta_k$ являются коэффициентами полинома; мы должны подобрать их так, чтобы график полинома наилучшим образом приближал полученную совокупность точек $(t_1, Y_1), (t_2, Y_2), \dots, (t_n, Y_n)$.

В частном случае, когда $k = 2$, мы имеем дело с *простой* регрессией, в остальных случаях регрессия называется *множественной*.

Пример. Предположим, что мы изучаем зависимость растворимости вещества в некоторой жидкости от температуры этой жидкости. Обозначим температуру буквой t и проведем измерения растворимости при разных температурах. Полученные данные (см. график ниже) наводят на мысль о линейной зависимости

$$Y_i = \theta_1 + \theta_2 t_i + \varepsilon_i, \quad i = 1, \dots, n.$$



Задача состоит в оценивании неизвестных параметров θ_1 и θ_2 , определяющих эту зависимость.

9.2. Метод наименьших квадратов

Оценки неизвестных параметров $\theta_1, \dots, \theta_k$ будем находить *методом наименьших квадратов*.

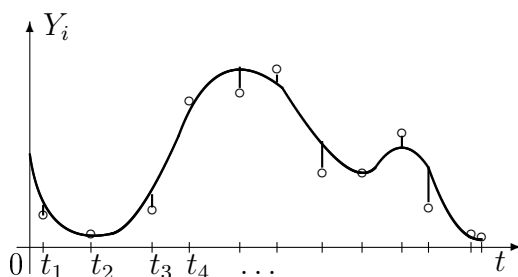
Введем

$$S(\theta) = \sum_{i=1}^n (Y_i - \theta_1 x_{i1} - \dots - \theta_k x_{ik})^2 = |Y - X\theta|^2.$$

Оценкой метода наименьших квадратов (МНК-оценкой) называется то значение $\theta = \theta^*$, при котором $S(\theta)$ достигает минимального значения:

$$S(\theta^*) = \min_{\theta} S(\theta).$$

Возвращаясь к рассмотренной выше графической иллюстрации мы видим, что числа $\theta_1, \dots, \theta_k$ подбираются так, чтобы минимальной была сумма квадратов длин вертикальных отрезков, соединяющих точки (t_i, Y_i) с соответствующими точками на кривой.



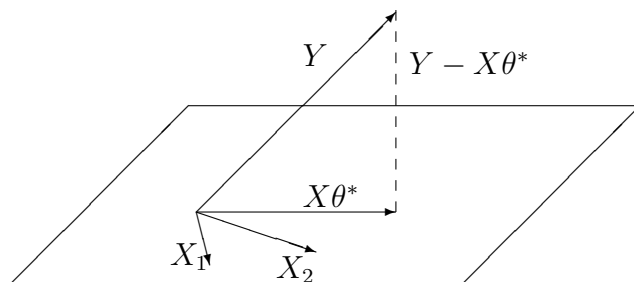
Находить МНК-оценку можно по-разному. Один из способов состоит в решении системы так называемых *нормальных уравнений*

$$\frac{\partial S(\theta)}{\partial \theta_j} = 0, \quad j = 1, \dots, k,$$

что достаточно громоздко. Мы попробуем найти оценку из геометрических соображений. Обозначим через X_1, \dots, X_k столбцы матрицы X . Это линейно независимые векторы в \mathbb{R}^n . Поскольку $k < n$, то эти векторы порождают в \mathbb{R}^n подпространство \mathbb{R}^k . Любая линейная комбинация этих векторов вновь принадлежит тому же \mathbb{R}^k , значит, для любого θ

$$X\theta = X_1\theta_1 + \dots + X_k\theta_k \in \mathbb{R}^k.$$

В том числе $X\theta^* \in \mathbb{R}^k$. Проиллюстрируем все на рисунке при $n = 3$ и $k = 2$.



В соответствии с методом наименьших квадратов нужно найти такое значение $\theta = \theta^*$, при котором длина вектора $Y - X\theta^*$ будет минимальной. Этот вектор на чертеже

показан пунктиром. Ясно, что его длина минимальна, если он ортогонален плоскости, а значит, и векторам, ее порождающим. Запишем этот вывод уже для общего случая:

$$Y - X\theta^* \perp X_j, \quad j = 1, \dots, k.$$

По-другому это можно записать так:

$$X^T(Y - X\theta^*) = \mathbf{0},$$

где $\mathbf{0}$ — нулевой вектор размерности k . Выводим отсюда

$$X^T Y = X^T X \theta^*.$$

Квадратная матрица $X^T X$ является невырожденной. Умножив полученное равенство слева на $(X^T X)^{-1}$, получим МНК-оценку

$$\theta^* = (X^T X)^{-1} X^T Y.$$

Предположим дополнительно, что векторы X_1, \dots, X_k ортогональны. Мы увидим, что в этом случае многое упрощается в задачах линейной регрессии. В частности, матрица $X^T X$ становится диагональной:

$$X^T X = \begin{pmatrix} (X_1, X_1) & 0 & \dots & 0 \\ 0 & (X_2, X_2) & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & (X_k, X_k) \end{pmatrix};$$

$(X^T X)^{-1}$ также будет диагональной:

$$(X^T X)^{-1} = \begin{pmatrix} (X_1, X_1)^{-1} & 0 & \dots & 0 \\ 0 & (X_2, X_2)^{-1} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & (X_k, X_k)^{-1} \end{pmatrix}.$$

Умножение этой матрицы на вектор

$$X^T Y = \begin{pmatrix} (X_1, Y) \\ (X_2, Y) \\ \dots \\ (X_k, Y) \end{pmatrix}$$

приводит к простым выражениям для компонент МНК-оценки θ^* :

$$\theta_i^* = \frac{(X_i, Y)}{(X_i, X_i)}, \quad i = 1, \dots, k.$$

9.3. Доверительные интервалы и проверка гипотез

После нахождения оценок неизвестных параметров естественно поставить вопрос о возможности построения доверительных интервалов и проверки гипотез. Для этого потребуется информация о распределении случайных величин ε_i .

Начиная с этого места мы предполагаем, что случайные величины $\varepsilon_1, \dots, \varepsilon_n$ независимы и распределены по закону Φ_{0, σ^2} . В таких случаях говорят о задачах *нормальной* регрессии.

Теорема. Пусть все ε_i независимы и распределены по закону Φ_{0,σ^2} . Тогда ММП-оценка и МНК-оценка для параметра θ совпадают.

Доказательство. Запишем функцию правдоподобия выборки (несмотря на то что наблюдения здесь распределены не одинаково, принцип остается тем же самым):

$$\begin{aligned} f(\theta, \sigma^2, Y) &= (2\pi\sigma^2)^{-n/2} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (Y_i - \theta_1 x_{i1} - \dots - \theta_k x_{ik})^2 \right\} = \\ &= (2\pi\sigma^2)^{-n/2} \exp \left\{ -\frac{1}{2\sigma^2} S(\theta) \right\}. \end{aligned}$$

Из этой записи видно, что исследовать функцию правдоподобия на максимум — это то же самое, что исследовать $S(\theta)$ на минимум. Теорема доказана.

Коль скоро выписана функция правдоподобия, найдем попутно ММП-оценку для σ^2 . Имеем

$$\begin{aligned} l(\theta, \sigma^2, Y) &= -\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} |Y - X\theta|^2, \\ \frac{\partial l(\theta, \sigma^2, Y)}{\partial(\sigma^2)} &= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} |Y - X\theta|^2 = 0, \end{aligned}$$

откуда получаем ММП-оценку

$$(\sigma^2)^* = \frac{|Y - X\theta^*|^2}{n}.$$

Вероятностные свойства оценок неизвестных параметров в задачах нормальной регрессии устанавливаются следующей теоремой (приводится без доказательства).

Теорема. Пусть все ε_i независимы и распределены по закону Φ_{0,σ^2} . Тогда:

1) МНК-оценка θ^* имеет многомерное нормальное распределение, при этом

$$\mathbf{E}\theta^* = \theta, \quad \mathbf{C}(\theta^*) = \sigma^2(X^T X)^{-1};$$

$$2) \frac{|Y - X\theta^*|^2}{\sigma^2} \in \chi_{n-k}^2;$$

3) θ^* и $|Y - X\theta^*|^2$ независимы.

Следствие 1. Оценка

$$\hat{\sigma}^2 = \frac{1}{n-k} |Y - X\theta^*|^2$$

является несмещенной для σ^2 .

Доказательство. В силу второго утверждения теоремы, случайная величина $\frac{|Y - X\theta^*|^2}{\sigma^2}$ распределена так же, как и $Z_1^2 + \dots + Z_{n-k}^2$, где все Z_i независимы и распределены по закону $\Phi_{0,1}$, поэтому

$$\mathbf{E} \left(\frac{|Y - X\theta^*|^2}{\sigma^2} \right) = \mathbf{E}(Z_1^2 + \dots + Z_{n-k}^2) = n - k,$$

т. е.

$$\mathbf{E} \left(\frac{|Y - X\theta^*|^2}{n-k} \right) = \sigma^2.$$

Отсюда вытекает, между прочим, что найденная ранее ММП-оценка $(\sigma^2)^*$ является смещенной.

Следствие 2. Если столбцы регрессора X_1, \dots, X_k ортогональны, то матрица $C(\theta^*) = \sigma^2(X^T X)^{-1}$ диагональна и компоненты оценки $\theta_1^*, \dots, \theta_k^*$ независимы. При этом

$$\theta_i^* \in \Phi_{\theta_i, \sigma^2/|X_i|^2}.$$

Это утверждение сразу же следует из свойств многомерного нормального распределения.

Следствие 3. Если столбцы регрессора X_1, \dots, X_k ортогональны, то

$$\frac{(\theta_i^* - \theta_i)|X_i|}{\sigma} : \sqrt{\frac{1}{n-k} \frac{|Y - X\theta^*|^2}{\sigma^2}} = \frac{(\theta_i^* - \theta_i)|X_i|\sqrt{n-k}}{|Y - X\theta^*|} \in T_{n-k}.$$

Теперь мы можем перейти к построению доверительных интервалов.

1. Доверительный интервал для σ^2 . Из таблиц распределения χ_{n-k}^2 находим числа q_1 и q_2 такие, что $\chi_{n-k}^2(q_1) = \varepsilon/2$, $\chi_{n-k}^2(q_2) = 1 - \varepsilon/2$. Тогда, в силу следствия 1,

$$\mathbf{P} \left(q_1 < \frac{|Y - X\theta^*|^2}{\sigma^2} < q_2 \right) = \chi_{n-k}^2(q_2) - \chi_{n-k}^2(q_1) = 1 - \varepsilon,$$

откуда следует

$$\mathbf{P} \left(\frac{|Y - X\theta^*|^2}{q_2} < \sigma^2 < \frac{|Y - X\theta^*|^2}{q_1} \right) = 1 - \varepsilon.$$

Далее предполагаем, что столбцы регрессора X_1, \dots, X_k ортогональны.

2. Доверительный интервал для θ_i при условии, что σ^2 известно. Здесь используется тот факт, что $\frac{(\theta_i^* - \theta_i)|X_i|}{\sigma} \in \Phi_{0,1}$. Пусть число q таково, что $\Phi_{0,1}(-q) = \varepsilon/2$, тогда

$$\mathbf{P} \left(-q < \frac{(\theta_i^* - \theta_i)|X_i|}{\sigma} < q \right) = \Phi_{0,1}(q) - \Phi_{0,1}(-q) = 1 - \varepsilon,$$

поэтому

$$\mathbf{P} \left(\theta_i^* - \frac{q\sigma}{|X_i|} < \theta_i < \theta_i^* + \frac{q\sigma}{|X_i|} \right) = 1 - \varepsilon.$$

3. Доверительный интервал для θ_i при условии, что σ^2 неизвестно. Воспользуемся следствием 3 и найдем из таблиц число q такое, что $T_{n-k}(-q) = \varepsilon/2$. Тогда

$$\mathbf{P} \left(-q < \frac{(\theta_i^* - \theta_i)|X_i|\sqrt{n-k}}{|Y - X\theta^*|} < q \right) = T_{n-k}(q) - T_{n-k}(-q) = 1 - \varepsilon$$

и после очевидных преобразований получаем доверительный интервал

$$\mathbf{P} \left(\theta_i^* - \frac{q|Y - X\theta^*|}{|X_i|\sqrt{n-k}} < \theta_i < \theta_i^* + \frac{q|Y - X\theta^*|}{|X_i|\sqrt{n-k}} \right) = 1 - \varepsilon.$$

Построенные доверительные интервалы позволяют проверять гипотезы в соответствии с изложенной ранее конструкцией. Например, для проверки гипотезы $H_1 : \theta_i = C$ против $H_2 : \theta_i \neq C$ при неизвестной дисперсии σ^2 берем критическое множество

$$\mathbf{K} = \left\{ (Y_1, \dots, Y_n) : C \notin \left(\theta_i^* - \frac{q|Y - X\theta^*|}{|X_i|\sqrt{n-k}}, \theta_i^* + \frac{q|Y - X\theta^*|}{|X_i|\sqrt{n-k}} \right) \right\}.$$

Тогда

$$\beta_1 = \mathbf{P}_1((Y_1, \dots, Y_n) \in \mathbf{K}) = \varepsilon.$$

Список использованной литературы

Гнеденко Б. В. Курс теории вероятностей. М.: Наука, 1988. 448 с.

Смирнов Н. В., Дунин-Барковский И. В. Курс теории вероятностей и математической статистики. М., Наука, 1965. 512 с.

Чистяков В. П. Курс теории вероятностей. М.: Наука, 1982. 256 с.