

Министерство образования и науки Российской Федерации
Новосибирский государственный университет

В.И. Лотов

ЛЕКЦИИ ПО ТЕОРИИ ВЕРОЯТНОСТЕЙ

*Учебное пособие для студентов механико-математического
факультета НГУ*

Новосибирск – 2011

Содержание

1. Вероятностные пространства. Основные формулы	5
1.1. Дискретные пространства	5
1.2. Вероятностное пространство общего вида	11
1.3. Континуальные пространства	15
1.4. Формула для вероятности объединения событий	17
1.5. Независимые события	18
1.6. Схема Бернулли	19
1.7. Приближение Пуассона в схеме Бернулли	21
1.8. Локальная предельная теорема о нормальном приближении	23
1.9. Полиномиальное распределение	24
1.10. Условные вероятности	25
1.11. Формула полной вероятности	26
1.12. Формула Байеса	27
2. Распределения	28
2.1. Случайные величины. Функции распределения	28
2.2. Типы распределений. Примеры	31
2.3. Независимость случайных величин и классов событий	40
2.4. Многомерные распределения и плотности	42
2.5. Преобразования случайных величин	45
3. Числовые характеристики распределений	50
3.1. Интеграл по вероятностной мере. Математическое ожидание	50
3.2. Моменты	59
3.3. Дисперсия	60
3.4. Коэффициент корреляции	62
3.5. Многомерный случай: математическое ожидание и матрица ковариаций	63
3.6. Многомерное нормальное распределение	65
3.7. Математическое ожидание суммы случайного числа слагаемых	66
3.8. Условное математическое ожидание	70
3.9. Задача о наилучшем приближении	72
4. Сходимость случайных величин и распределений. Предельные теоремы	73
4.1. Сходимость последовательностей случайных величин	73
4.2. О сходимости математических ожиданий	77
4.3. Законы больших чисел	79
4.4. Слабая сходимость распределений	81
4.5. Характеристические функции	84
4.6. Центральная предельная теорема	90
4.7. Оценка точности в теореме Пуассона.	94
5. Цепи Маркова	96
5.1. Основные определения	96
5.2. Возвратность состояний	98
5.3. Эргодическая теорема	101

6. Ветвящиеся процессы	104
7. Случайные процессы с непрерывным временем	107
7.1. Общие определения. Винеровский процесс	107
7.2. Процесс Пуассона	109

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

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

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:

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

Свойства вероятности легко запомнить, если провести аналогию с массой тела. Представим себе, что каждый элементарный исход ω имеет массу $\mathbf{P}(\omega)$. Тогда масса всего пространства Ω равна единице, и вероятность каждого события тоже может восприниматься как его масса. Перечисленные свойства вероятности вполне соответствуют нашим представлениям о свойствах массы тела.

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), $k = 1, 2, \dots, n$. Сколькими различными выборками 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. Скольким числом 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.

Задачи о размещении шаров по ящикам широко используются в статистической физике. Обычно там говорят о размещении частиц по ячейкам. Если k различных частиц произвольным образом размещаются по n ячейкам, и все n^k полученных размещений равновероятны, то такую схему физики называют *статистикой Максвелла-Больцмана*.

4. Пусть теперь шары неразличимы, их по-прежнему k штук, и они без каких-либо ограничений распределяются по n ящикам. Сколько существует различных размещений в этой схеме?

Эта задача сложна. Предварительно зададимся вопросом: сколько можно построить двоичных последовательностей, если в нашем распоряжении имеется m единиц и r нулей? Последовательность имеет длину $m + r$, и из $m + r$ мест в ней любые m могут быть заняты единицами. Следовательно, мы выбираем любые m мест из $m + r$ имеющихся — а это можно сделать C_{m+r}^m способами.

Вернемся теперь к исходной задаче. Представим себе, что у нас есть узкое длинное корыто, на дне которого в один ряд разложены шары, и имеются перегородки, вставляя которые поперек корыта, мы получим n ящиков. Ясно, что перегородок потребуется $n - 1$.



Становится очевидной аналогия с предыдущей задачей: перегородки можно считать единицами, а шары — нулями. Число размещений шаров по ящикам будет совпадать с числом последовательностей из $n - 1$ единиц и k нулей. В итоге получаем $C_{n-1+k}^{m-1} = C_{n-1+k}^k$ размещений.

В связи с тем, что многие задачи на подсчет вероятностей с помощью классического определения могут быть сведены к размещению шаров по ящикам, возникает вопрос: какой схемой пользоваться. Если считать шары различными, то получается один результат, неразличимыми — другой.

Наша рекомендация состоит в следующем. Нужно всегда считать шары различными. Как правило, разные размещения в этой схеме имеют одинаковую вероятность (если только задача не связана с частицами из микромира). Если же шары объявить неразличимыми, то тогда некоторые размещения перестанут различаться, то есть произойдет "укрупнение" элементарных исходов. Например, если в схеме с различными шарами k шаров размещаются по одному в разных ящиках, то все $k!$ перестановок этих шаров приведут к разным элементарным исходам. Если же эти шары считать неразличимыми, то $k!$ элементарных исходов сольются в один исход. Однако, если изначально k шаров находились в одном ящике, то их перестановки ничего не дадут ни в той, ни в другой схеме. Это значит, что укрупнение элементарных исходов происходит неравномерно, и полученные таким образом новые элементарные исходы уже не будут иметь одинаковые вероятности, что исключает возможность пользования классическим определением.

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. Вероятностное пространство общего вида

Мы подробно изучили дискретную вероятностную модель. В ней вероятностная масса распределялась по дискретному набору элементарных исходов. Однако, как мы видели, такая модель годится не для всех случайных экспериментов. Пространство элементарных исходов может быть более богатым и вероятностная масса может непрерывным образом «размазываться» по пространству или по его части. Поэтому мы приходим к необходимости построить более общую конструкцию вероятностного пространства. Она задается системой аксиом, которые были предложены в 30-х годах прошлого столетия А.Н. Колмогоровым.

Итак, мы по-прежнему называем вероятностным пространством тройку $\langle \Omega, \mathcal{S}, \mathbf{P} \rangle$, где про Ω уже все сказано — это множество возможных исходов эксперимента, \mathcal{S} — совокупность подмножеств Ω , называемых событиями. В отличие от дискретной модели, в общем случае \mathcal{S} может включать в себя не все подмножества Ω . Однако для \mathcal{S} всегда должны выполняться следующие требования.

S1. $\Omega \in \mathcal{S}$.

S2. Если $\{A_i\}$ — последовательность множеств из \mathcal{S} , то и $\bigcup_{i=1}^{\infty} A_i \in \mathcal{S}$.

S3. Если $A \in \mathcal{S}$, то и $\bar{A} \in \mathcal{S}$.

Из соотношения $\bigcap_{i=1}^{\infty} A_i = \overline{\bigcup_{i=1}^{\infty} \bar{A}_i}$ следует, что если $\{A_i\}$ — последовательность множеств из \mathcal{S} , то также $\bigcap_{i=1}^{\infty} A_i \in \mathcal{S}$.

Совокупность подмножеств \mathcal{S} , удовлетворяющая требованиям S1 — S3, называется σ -алгеброй. Разумеется, множество всех подмножеств Ω является σ -алгеброй. Это самая богатая σ -алгебра. Для контраста можно привести пример самой бедной σ -алгебры, состоящей всего из двух множеств $\{\emptyset, \Omega\}$, или, к примеру, $\{\emptyset, A, \bar{A}, \Omega\}$, где A — произвольное подмножество Ω .

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:

P1. $\mathbf{P}(\Omega) = 1$.

P2. $\mathbf{P}(A) \geq 0$ для любого $A \in \mathcal{S}$.

Р3. Счётная аддитивность: если события A_1, A_2, \dots таковы, что $A_i A_j = \emptyset$ ($i \neq j$) (т. е. попарно несовместны), то

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

Аксиомы Р2 — Р3 задают *меру*, определенную на \mathcal{S} . Мера, для которой выполняется еще и свойство Р1, называется *нормированной* или *вероятностной*. Из приведенных аксиом вытекает ряд полезных свойств вероятности. Некоторые из них мы имели возможность наблюдать в дискретных 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 Р3

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

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 Р2 получаем $\mathbf{P}(B) = \mathbf{P}(A) + \mathbf{P}(B \setminus A) \geq \mathbf{P}(A)$.

6. Для любой последовательности событий $\{A_i\}$ имеет место

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

Доказательство.

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

7. 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.

При построении вероятностного пространства могут возникнуть трудности, связанные со способом задания вероятности так, чтобы выполнялись требования Р1 — Р3. Мы видели, что в дискретной модели вероятность любого подмножества Ω без труда могла быть определена через вероятности отдельных элементарных исходов. Эта конструкция не годится, если пространство элементарных исходов несчетно. Выясняется, что в этом случае мы не можем задать удовлетворительным способом вероятность на всех без исключения подмножествах Ω . С другой стороны, хотелось бы,

чтобы класс событий был как можно более широким, это диктуется многочисленными приложениями модели. Но чем шире класс множеств, тем труднее определить на нем вероятность с соблюдением необходимых правил P1 — P3. Гораздо проще задать вероятность на совокупности подмножеств Ω , образующих алгебру. По определению, алгеброй \mathcal{A} называется совокупность подмножеств Ω , удовлетворяющая следующим условиям.

A1. $\Omega \in \mathcal{A}$.

A2. Если A_1, A_2, \dots, A_n — множества из \mathcal{A} , то и $\bigcup_{i=1}^n A_i \in \mathcal{A}$.

A3. Если $A \in \mathcal{A}$, то и $\bar{A} \in \mathcal{A}$.

Таким образом, в отличие от σ -алгебры алгебра замкнута относительно объединения и пересечения только лишь конечного числа своих элементов.

Для каждой алгебры найдется σ -алгебра, которая ее содержит. Для этих целей можно взять, к примеру, σ -алгебру всех подмножеств. Рассмотрим все σ -алгебры, содержащие алгебру \mathcal{A} . Их пересечение тоже будет являться σ -алгеброй, это легко проверить, причем оно также будет содержать \mathcal{A} . Обозначим это пересечение $\sigma(\mathcal{A})$ и назовем его σ -алгеброй, порожденной \mathcal{A} . Оказывается, что если сначала задать вероятность на алгебре \mathcal{A} (что не так трудно), то потом ее можно единственным образом продолжить на σ -алгебру $\sigma(\mathcal{A})$. Об этом — следующая теорема Каратеодори о продолжении меры (приводится без доказательства).

Теорема. Пусть вероятность \mathbf{P} задана на алгебре событий \mathcal{A} , то есть выполнены условия

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

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

3. Если события A_1, A_2, \dots таковы, что $A_i A_j = \emptyset$ ($i \neq j$) и $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$, то

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

Тогда существует и притом единственная вероятностная мера $\bar{\mathbf{P}}$, определенная на $\sigma(\mathcal{A})$ такая, что $\mathbf{P}(A) = \bar{\mathbf{P}}(A)$ для всех $A \in \mathcal{A}$.

Пусть, для примера, $\Omega = \mathbb{R}$. Разумеется, проще всего определить вероятность с соблюдением аддитивности сначала на множестве \mathcal{K} всевозможных промежутков вида $[a, b)$, $a < b$. Например, можно положить $\mathbf{P}([a, b)) = F(b) - F(a)$, где F — некоторая неубывающая непрерывная слева функция, обладающая свойствами $\lim_{x \rightarrow -\infty} F(x) = 0$, $\lim_{x \rightarrow \infty} F(x) = 1$. Ясно, что \mathcal{K} алгеброй не является. Расширим \mathcal{K} , добавив в него всевозможные объединения и пересечения конечного числа множеств из \mathcal{K} . Полученная таким образом совокупность множеств \mathcal{A} будет алгеброй. Вероятностную меру объединения конечного числа непересекающихся промежутков определим как сумму мер отдельных промежутков. Пользуясь свойствами функции F , можно показать, что на \mathcal{A} будет иметь место и свойство счетной аддитивности. Теперь можем рассмотреть σ -алгебру $\sigma(\mathcal{A})$, порожденную в конечном итоге всевозможными промежутками, она называется σ -алгеброй борелевских множеств и обозначается $\mathcal{B}(\mathbb{R})$. В силу теоремы о продолжении меры вероятность, заданная первоначально на промежутках, единственным образом продолжается на множество $\mathcal{B}(\mathbb{R})$.

В случае $\Omega = \mathbb{R}^n$ мы вначале задаем вероятность на всевозможных прямоугольниках вида $[a_1, b_1) \times \dots \times [a_n, b_n)$, а затем по той же схеме продолжаем ее на σ -алгебру борелевских множеств $\mathcal{B}(\mathbb{R}^n)$.

1.3. 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 будем рассматривать множества из $\mathcal{B}(\mathbb{R}^n)$, а для задания вероятности достаточно будет определить ее на всевозможных прямоугольниках.

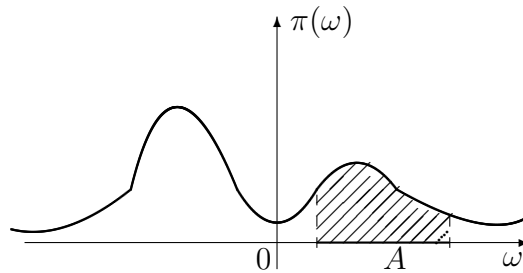
Для лучшего понимания остановимся сначала более подробно на случае $\Omega = \mathbb{R}^1$. Предположим, что у нас имеется некоторая интегрируемая функция $\pi : \Omega \rightarrow \mathbb{R}$ такая, что:

- 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, для любого промежутка A на числовой оси

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

Это определение имеет простой геометрический смысл: вероятность того или иного промежутка на прямой вычисляется как площадь криволинейной трапеции, имеющей данный промежуток 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 $\lambda(D) > 0$.

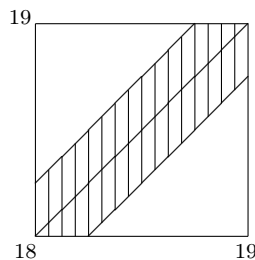
Рассмотрим в качестве примера *задачу о встрече*.

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.

Дальнейшее изложение материала будет относиться к вероятностным пространствам общего вида.

1.4. Формула для вероятности объединения событий

Мы видели, что для любых событий A_1 и A_2 имеет место

$$\mathbf{P}(A_1 \cup A_2) = \mathbf{P}(A_1) + \mathbf{P}(A_2) - \mathbf{P}(A_1 A_2).$$

Следующая теорема дает обобщение этой формулы для любого числа событий.

Теорема. Для любых событий A_1, \dots, A_n

$$\mathbf{P}\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n \mathbf{P}(A_i) - \sum_{i<j} \mathbf{P}(A_i A_j) + \sum_{i<j<k} \mathbf{P}(A_i A_j A_k) - \dots + (-1)^{n-1} \mathbf{P}(A_1 \dots A_n).$$

Доказательство. Воспользуемся индукцией. При $n = 2$ утверждение верно. Предположим, что оно верно для вероятности объединения произвольных $n - 1$ событий.

Обозначим $B = \bigcup_{i=1}^{n-1} A_i$, тогда

$$\begin{aligned} \mathbf{P}\left(\bigcup_{i=1}^n A_i\right) &= \mathbf{P}(B \cup A_n) = \mathbf{P}(B) + \mathbf{P}(A_n) - \mathbf{P}(B A_n) = \\ &= \sum_{i=1}^{n-1} \mathbf{P}(A_i) + \mathbf{P}(A_n) - \sum_{i<j \leq n-1} \mathbf{P}(A_i A_j) + \dots + (-1)^{n-2} \mathbf{P}(A_1 \dots A_{n-1}) - \mathbf{P}\left(\bigcup_{i=1}^{n-1} A_i A_n\right) = \end{aligned}$$

$$= \sum_{i=1}^n \mathbf{P}(A_i) - \sum_{i < j \leq n-1} \mathbf{P}(A_i A_j) - \sum_{i=1}^n \mathbf{P}(A_i A_n) + \dots + (-1)^{n-1} \mathbf{P}(A_1 \dots A_n).$$

Теорема доказана.

Пример. Некто написал n писем и подготовил для них n конвертов с адресами. Однако секретарь этого человека решил подшутить, разложив письма по конвертам случайным образом. Какова вероятность, что при этом хотя бы одно письмо дойдет по назначению?

Здесь пространство элементарных исходов состоит из $n!$ перестановок, вероятность каждого элементарного исхода равна $1/n!$. Пусть событие A_i означает, что i -е письмо дойдет по своему назначению, нас интересует $\mathbf{P}(\bigcup_{i=1}^n A_i)$. Каждое событие A_i состоит из $(n-1)!$ элементарных исходов,

$$\begin{aligned} \mathbf{P}(A_i) &= \frac{(n-1)!}{n!} = \frac{1}{n}, & \mathbf{P}(A_i A_j) &= \frac{(n-2)!}{n!}, \dots, \\ \mathbf{P}\left(\bigcup_{i=1}^n A_i\right) &= n \frac{1}{n} - C_n^2 \frac{(n-2)!}{n!} + C_n^3 \frac{(n-3)!}{n!} - \dots + (-1)^{n-1} \frac{1}{n!} = \\ &= 1 - \frac{1}{2!} + \frac{1}{3!} - \dots + (-1)^{n-1} \frac{1}{n!}. \end{aligned}$$

Последнее выражение стремится к $1 - e^{-1} = 0.632\dots$ при $n \rightarrow \infty$.

1.5. 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:

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

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 $k > 2$.

Пример Бернштейна. На плоскость бросается тетраэдр, три грани которого окрашены соответственно в красный, зеленый и синий цвета, а на четвертую грань нанесены все три эти цвета. Обозначим R, G и B события, означающие, что на выпавшей книзу грани присутствует соответственно красный, зеленый или синий цвет. Поскольку каждый цвет присутствует на двух гранях из четырех, то $\mathbf{P}(R) = \mathbf{P}(G) = \mathbf{P}(B) = 1/2$. Очевидно, попарная независимость событий имеет место, $\mathbf{P}(RG) = \mathbf{P}(GB) = \mathbf{P}(RB) = 1/4$, однако $\mathbf{P}(RGB) = 1/4 \neq \mathbf{P}(R)\mathbf{P}(G)\mathbf{P}(B)$.

1.6. 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 $k = [np - q] + 1 = [(n + 1)p]$.

1.7. Приближение Пуассона в схеме Бернулли

Пусть по-прежнему S_n — число успехов в схеме Бернулли. Мы знаем точные формулы для вероятностей вида

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

Однако на практике возникают ситуации, когда применение точных формул затруднительно из-за того, что n очень велико. В этом случае можно пользоваться формулами приближенного вычисления, если возникающая при этом погрешность пренебрежимо мала при больших n . Одно из возможных приближений обеспечивается теоремой Пуассона, которой пользуются в тех случаях, когда вероятность успеха p очень мала, т. е. успех появляется в испытаниях Бернулли крайне редко.

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.

Теорема. Для любого множества $A \subset \{0, 1, 2, \dots\}$

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

Доказательство этой теоремы будет проведено позже.

Семейство вероятностей вида $\left\{ \frac{\lambda^k}{k!} e^{-\lambda}, k = 0, 1, \dots \right\}$, где $\lambda > 0$ — фиксированное число, называется *распределением Пуассона*. Здесь, очевидно, $\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} = 1$. Ясно, что данное приближение обеспечивает хорошую точность в схеме Бернулли с настолько малой вероятностью успеха p , чтобы и число np^2 также было малым.

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.

1.8. Локальная предельная теорема о нормальном приближении

В этом разделе будет предложена другая формула для приближенного вычисления вероятностей в схеме Бернулли.

Мы будем писать $a_n \sim b_n$, если $\frac{a_n}{b_n} \rightarrow 1$ при $n \rightarrow \infty$.

Теорема (локальная теорема Муавра). Пусть S_n — число успехов в схеме Бернулли, и пусть $k = np + o(n^{2/3})$ при $n \rightarrow \infty$. Обозначим $x_k = \frac{k - np}{\sqrt{npq}}$. Тогда при $n \rightarrow \infty$

$$\mathbf{P}(S_n = k) = \mathbf{P}\left(\frac{S_n - np}{\sqrt{npq}} = x_k\right) \sim \frac{1}{\sqrt{2\pi npq}} \exp\left\{-\frac{x_k^2}{2}\right\}.$$

Доказательство. Сначала докажем вспомогательное утверждение.

Лемма. Если $k \rightarrow \infty$ и $n - k \rightarrow \infty$, то

$$\mathbf{P}(S_n = k) = \mathbf{P}\left(\frac{S_n}{n} = p^*\right) \sim \frac{1}{\sqrt{2\pi np^*(1-p^*)}} \exp\{-nH(p^*)\},$$

где

$$H(x) = x \ln \frac{x}{p} + (1-x) \ln \frac{1-x}{1-p}, \quad p^* = \frac{k}{n}.$$

Доказательство. Воспользуемся формулой Стирлинга, согласно которой

$$n! \sim \sqrt{2\pi n} \left(\frac{n}{e}\right)^n$$

при $n \rightarrow \infty$. Имеем

$$\begin{aligned} \mathbf{P}(S_n = k) &= C_n^k p^k q^{n-k} \sim \sqrt{\frac{n}{2\pi k(n-k)}} \frac{n^n}{k^k (n-k)^{n-k}} p^k q^{n-k} = \\ &= \frac{1}{\sqrt{2\pi np^*(1-p^*)}} \exp\left\{-k \ln \frac{k}{n} - (n-k) \ln \frac{n-k}{n} + k \ln p + (n-k) \ln(1-p)\right\} = \end{aligned}$$

$$= \frac{1}{\sqrt{2\pi np^*(1-p^*)}} \exp \left\{ -n [p^* \ln p^* + (1-p^*) \ln(1-p^*) - p^* \ln p - (1-p^*) \ln(1-p)] \right\}$$

$$= \frac{1}{\sqrt{2\pi np^*(1-p^*)}} \exp \{-nH(p^*)\}.$$

Лемма доказана.

Приступим теперь к доказательству теоремы. Имеем при $x \in (0, 1)$

$$H'(x) = \ln \frac{x}{p} - \ln \frac{1-x}{1-p}, \quad H''(x) = \frac{1}{x} + \frac{1}{1-x}, \quad H(p) = H'(p) = 0.$$

Разлагая $H(p^*)$ в ряд в окрестности точки p , получим при $p^* - p \rightarrow 0$

$$H(p^*) = \frac{1}{2} \left(\frac{1}{p} + \frac{1}{q} \right) (p^* - p)^2 + O(|p^* - p|^3).$$

По условию теоремы $p^* = \frac{k}{n} = p + o(n^{-1/3})$. Значит, $p^* - p \rightarrow 0$, $n(p^* - p)^3 \rightarrow 0$ и

$$nH(p^*) = \frac{n(p^* - p)^2}{2pq} + o(1),$$

т. е.

$$\mathbf{P}(S_n = k) \sim \frac{1}{\sqrt{2\pi npq}} \exp \left\{ -\frac{n}{2pq} (p^* - p)^2 \right\} = \frac{1}{\sqrt{2\pi npq}} \exp \left\{ -\frac{x_k^2}{2} \right\}$$

(здесь также использовалось свойство $p^*(1-p^*) \sim p(1-p)$). Утверждение доказано.

Используя локальную теорему, можно было бы доказать *интегральную теорему Муавра-Лапласа*, состоящую в следующем. Для любых чисел $a < b$

$$\lim_{n \rightarrow \infty} \mathbf{P} \left(a < \frac{S_n - np}{\sqrt{npq}} < b \right) = \frac{1}{\sqrt{2\pi}} \int_a^b \exp \left\{ -\frac{x^2}{2} \right\} dx.$$

Мы докажем впоследствии так называемую *центральную предельную теорему*, по отношению к которой интегральная теорема Муавра-Лапласа будет являться частным случаем.

1.9. Полиномиальное распределение

Полиномиальное распределение является обобщением биномиального. Пусть производится n независимых испытаний, каждое из которых может завершиться одним из исходов A_1, A_2, \dots, A_k с вероятностями p_1, p_2, \dots, p_k соответственно, $p_1 + \dots + p_k = 1$. Очевидно, в случае схемы Бернулли имеем $k = 2$. Какова вероятность того, что в n испытаниях исход A_1 произойдет n_1 раз, исход A_2 — n_2 раз, и т. д., исход A_k произойдет n_k раз?

В этой задаче элементарный исход эксперимента можно представлять себе как упорядоченную последовательность вида $A_{i_1} A_{i_2} \dots A_{i_n}$, где $1 \leq i_j \leq k$. Всего в пространстве k^n исходов. Каждому благоприятному исходу припишем вероятность $p_1^{n_1} \dots p_k^{n_k}$, всего таких исходов

$$C_n^{n_1} C_{n-n_1}^{n_2} \dots C_{n_k}^{n_k} = \frac{n!}{n_1! \dots n_k!}.$$

Получаем в итоге, что искомая вероятность равна

$$\frac{n!}{n_1! \dots n_k!} p_1^{n_1} \dots p_k^{n_k}.$$

Совокупность вероятностей такого вида, где $n_i \geq 0$ и $n_1 + \dots + n_k = n$, называется *полиномиальным* распределением. Обобщением бинома Ньютона является полиномиальная формула

$$(p_1 + \dots + p_k)^n = \sum_{\substack{n_1 \geq 0, \dots, n_k \geq 0 \\ n_1 + \dots + n_k = n}} \frac{n!}{n_1! \dots n_k!} p_1^{n_1} \dots p_k^{n_k}.$$

1.10. 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 .

1.11. 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 изделий, 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

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

1.12. 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 рассмотрении случайных явлений или экспериментов нас интересует, как правило, не сам реализовавшийся исход, 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 пробоины. Для нас важно, сколько очков мы выбили при стрельбе.

Это наводит на необходимость введения понятия случайной величины.

Пусть имеется некоторое вероятностное пространство $\langle \Omega, \mathcal{S}, \mathbf{P} \rangle$.

Определение. Случайной величиной X называется произвольная измеримая функция, заданная на пространстве элементарных исходов Ω и принимающая значения в \mathbb{R} (значения $\pm\infty$ исключаются). Измеримость означает, что $X^{-1}(B) \in \mathcal{S}$ для всякого борелевского множества $B \in \mathcal{B}(\mathbb{R})$.

Таким образом, 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.

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.

Каждая случайная величина X порождает на σ -алгебре борелевских множеств $\mathcal{B}(\mathbb{R})$ вероятностную меру

$$\mathbf{P}_X(B) = \mathbf{P}\{\omega : X(\omega) \in B\} = \mathbf{P}(X^{-1}(B)).$$

Определение. Вероятностная мера \mathbf{P}_X называется *распределением* случайной величины X .

В дальнейшем будем использовать краткую запись:

$$\mathbf{P}\{\omega : X(\omega) \in B\} = \mathbf{P}(X \in B).$$

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.

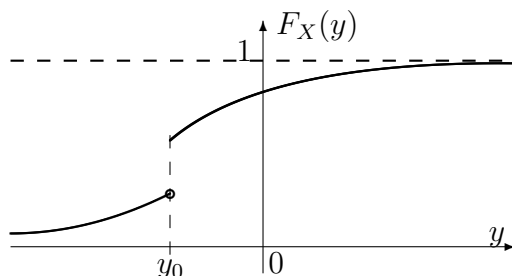
Для доказательства второго предельного соотношения рассмотрим последовательность чисел y_k , монотонно стремящуюся к бесконечности (например, $y_k = k$), и введем события $A_k = \{X < y_k\}$, $k = 1, 2, \dots$. Очевидно, $A_1 \subset A_2 \subset \dots$; и опять в силу свойства непрерывности вероятности

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

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 являются 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. $\omega \in A_{k_0} \subset \bigcup A_k$.

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 $\{X < b\} = \{X < a\} \cup \{a \leq X < b\}$, откуда следует

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

поэтому

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

Точно так же

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

Свойства 2–4, доказанные нами, являются характеристическими для функций распределения в том смысле, что любая функция F , ими обладающая, является функцией распределения какой-то случайной величины в подходящем вероятностном пространстве.

Мы не будем приводить полного доказательства этого факта, хотя его идея проста. Нужно взять $\Omega = \mathbb{R}$, $\mathcal{S} = \mathcal{B}(\mathbb{R})$, и положить $\mathbf{P}([a, b)) = F(b) - F(a)$. Введенная таким образом функция \mathbf{P} обладает свойством счетной аддитивности на алгебре множеств, составленных из конечного числа полуинтервалов — это проверяется с

использованием свойств 2-4. По теореме Каратеодори эта функция продолжается единственным образом на σ -алгебру борелевских множеств $\mathcal{B}(\mathbb{R})$.

После того как построили вероятностное пространство, введем случайную величину $X(\omega) \equiv \omega$. Нетрудно видеть, что ее функцией распределения и будет являться F .

Таким образом, распределение полностью определяется его заданием на всевозможных промежутках, а для этого достаточно знать всего одну функцию — функцию распределения этой случайной величины. В связи с этим термины «распределение» и «функция распределения» (а также «закон распределения») иногда 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

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

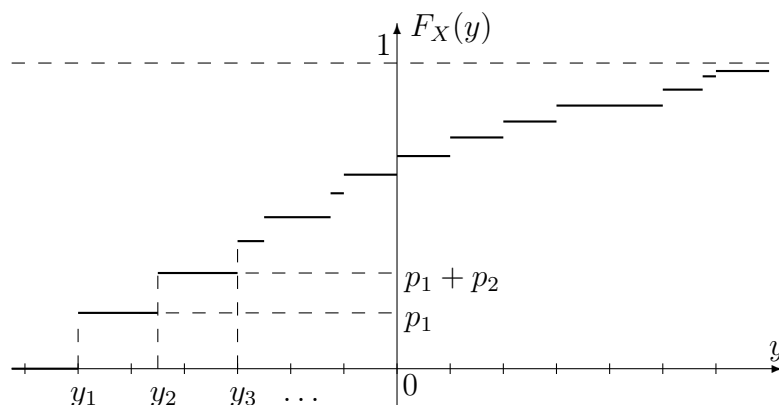
тогда приведенная ниже таблица полностью характеризует распределение.

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:

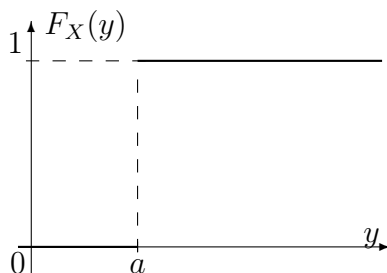


Действительно, если $y < y_1$, то $F_X(y) = \mathbf{P}(X < y) = 0$; если $y_1 < y < y_2$, то $F_X(y) = \mathbf{P}(X < y) = \mathbf{P}(X = y_1) = p_1$, и т. д.

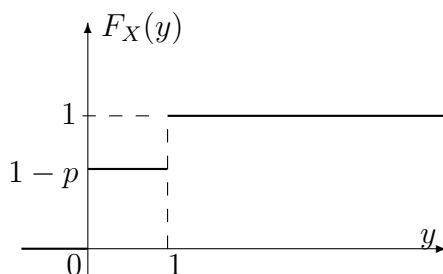
В дальнейшем мы будем писать $X \in F$, если X имеет функцию распределения F .

Примеры дискретных распределений

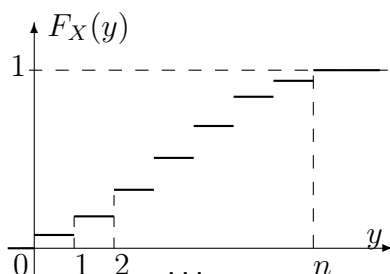
1. *Вырожденное распределение* I_a : $X \in I_a$, если $\mathbf{P}(X = a) = 1$.



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

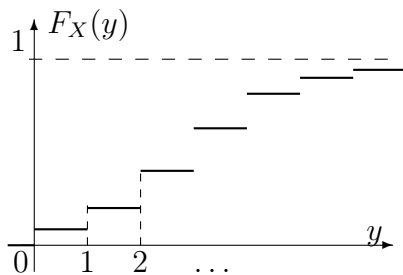


3. *Биномиальное распределение* $B_{n,p}$: $X \in B_{n,p}$, если $\mathbf{P}(X = k) = C_n^k p^k (1 - p)^{n-k}$, $k = 0, 1, \dots, n$ (в частности, $B_{1,p} = B_p$).



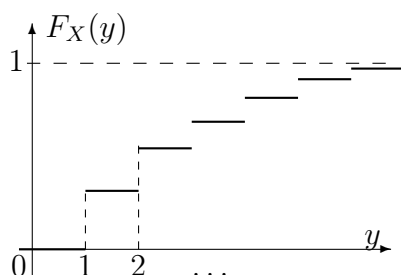
Биномиальное распределение, как мы уже видели, возникает при рассмотрении схемы Бернулли — это распределение числа успехов в n испытаниях.

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;

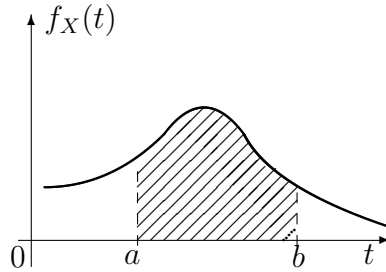
2) $\int_{-\infty}^{\infty} f_X(t) dt = 1$.

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

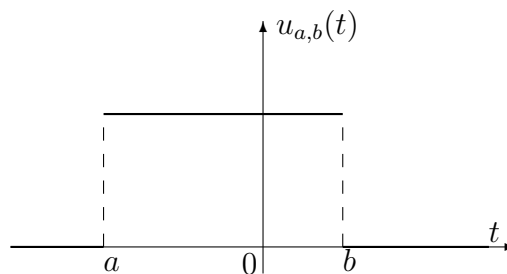
$$\mathbf{P}(X \in B) = \int_B f_X(t) dt.$$

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

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



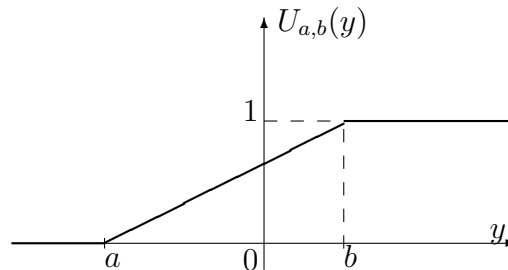
Ясно, что в данном случае все значения случайной величины располагаются на отрезке $[a, b]$ и равномерно там разбросаны; вероятность попадания в любой промежуток $[c, d] \subset [a, b]$ равна отношению длин

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

что уже встречалось нам в задачах на геометрические вероятности.

Для функции распределения имеем формулу

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

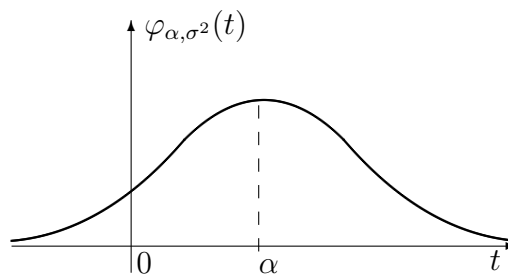


Как видим, в двух точках эта функция производной не имеет.

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

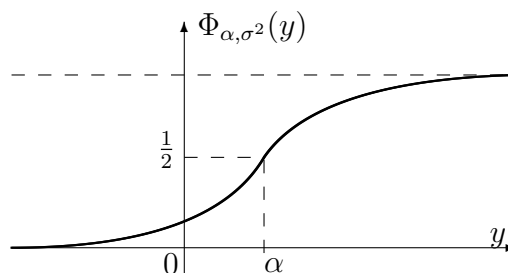
$$\varphi_{\alpha, \sigma^2}(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(t-\alpha)^2/2\sigma^2}, \quad -\infty < t < \infty.$$

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

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

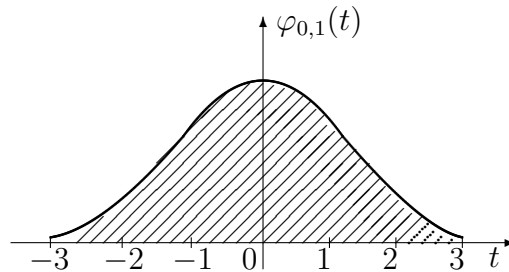
и с функцией распределения

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

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,

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

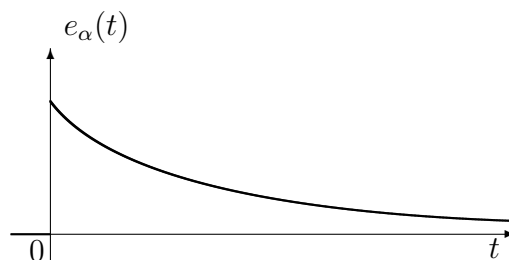
Последнее известно как *правило трех сигм*.

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

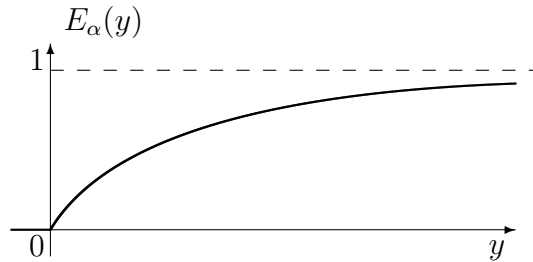
$$e_\alpha(t) = \begin{cases} \alpha e^{-\alpha t}, & t > 0, \\ 0, & t \leq 0. \end{cases}$$

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. Каково будет распределение у оставшейся продолжительности разговора? Оказывается, в точности такое же, как и у всей продолжительности X . Действительно, вероятность того, что оставшаяся длительность разговора будет не меньше t , равна

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

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

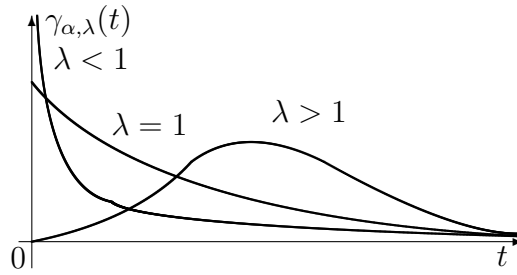
$$\gamma_{\alpha, \lambda}(t) = \begin{cases} \frac{\alpha^\lambda}{\Gamma(\lambda)} t^{\lambda-1} e^{-\alpha t}, & t > 0, \\ 0, & t \leq 0. \end{cases}$$

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

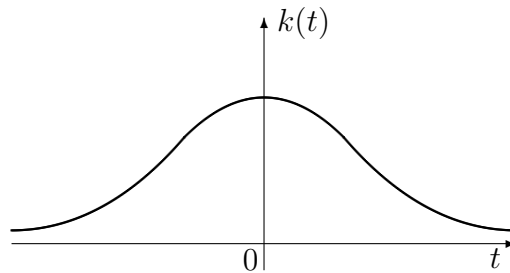
$$\Gamma_{\alpha, \lambda}(y) = \int_0^y \gamma_{\alpha, \lambda}(t) dt$$

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.

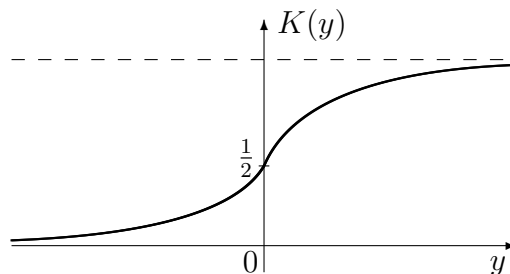
5. *Распределение Коши К.* Плотность задается формулой

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



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:

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



Есть еще так называемые *сингулярные* функции распределения. Они являются непрерывными, однако их точки роста образуют множество нулевой лебеговой меры (точка x называется точкой роста, если для любого $\varepsilon > 0$ $F(x + \varepsilon) - F(x - \varepsilon) > 0$).

Таким образом, $\frac{dF_X(t)}{dt} = 0$ почти всюду, однако $F(\infty) - F(-\infty) = 1$. Это трудно представить, но такие функции распределения есть. Примером может служить известная кривая Кантора, все изменение которой сосредоточено на отрезке $[0, 1]$: $F(x) = 0$ при $x \leq 0$ и $F(x) = 1$ при $x \geq 1$. Строится она следующим образом. Отрезок $[0, 1]$ разбивается на три равные части $[0, 1/3]$, $[1/3, 2/3]$, $[2/3, 1]$. На внутреннем сегменте полагаем $F(x) = 1/2$. Оставшиеся два сегмента снова разбиваются на три равные части каждый, и на внутренних сегментах $F(x)$ полагается равной соответственно $1/4$ и $3/4$. Каждый из оставшихся сегментов снова делится на три части, и на внутренних $F(x)$ определяется как постоянная, равная среднему арифметическому между соседними, уже определенными значениями $F(x)$, и т. д. Нетрудно видеть, что построенная функция является непрерывной и суммарная длина отрезков постоянства для $F(x)$ равна

$$\frac{1}{3} + \frac{2}{9} + \frac{4}{27} + \dots = \frac{1}{3} \sum_{k=0}^{\infty} \left(\frac{2}{3}\right)^k = \frac{1}{3} \frac{1}{1 - 2/3} = 1,$$

так что функция $F(x)$ растет на множестве меры нуль, но без скачков.

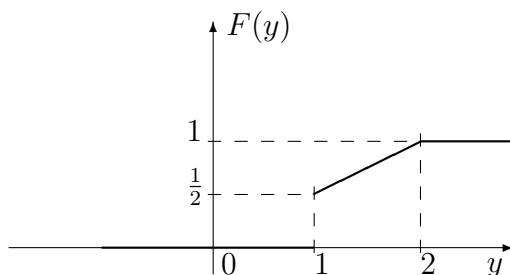
Если F_1 , F_2 и F_3 — функции распределения соответственно абсолютно непрерывного, дискретного и сингулярного типов, то любая линейная комбинация вида

$$F(y) = \alpha F_1(y) + \beta F_2(y) + \gamma F_3(y), \quad \alpha \geq 0, \quad \beta \geq 0, \quad \gamma \geq 0, \quad \alpha + \beta + \gamma = 1,$$

также будет функцией распределения. Отнесем ее к смешанному типу. Известная теорема Лебега утверждает, что любая функция распределения единственным образом представляется в таком виде как смесь абсолютно непрерывной, дискретной и сингулярной компонент.

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

Пример функции распределения смешанного типа. На рисунке изображен график некоторой функции распределения. Очевидно, 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 единицу), $\beta = 1/2$. Тогда ясно, что $F_1(y) = U_{1,2}(y)$, $\alpha = 1/2$, $\gamma = 0$.

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

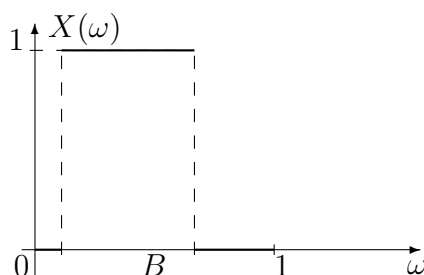
распределение, а вот по распределению восстановить случайную величину невозможно.

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 $\Omega = [0, 1]$, событиями будем считать все борелевские подмножества $[0, 1]$. 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. Независимость случайных величин и классов событий

Определение. Случайные величины X_1, X_2, \dots, X_n называются *независимыми*, если для любых борелевских множеств $B_1 \in \mathcal{B}(\mathbb{R}), \dots, B_n \in \mathcal{B}(\mathbb{R})$ выполняется соотношение

$$\mathbf{P}(X_1 \in B_1, X_2 \in B_2, \dots, X_n \in B_n) = \mathbf{P}(X_1 \in B_1)\mathbf{P}(X_2 \in B_2)\dots\mathbf{P}(X_n \in B_n) \quad (1)$$

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

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

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

Случайные величины $\{X_n\}_{n=1}^{\infty}$, заданные на одном вероятностном пространстве, будут называться независимыми, если при любом n случайные величины X_1, \dots, X_n независимы в смысле предыдущего определения.

Для произвольной случайной величины X рассмотрим совокупность $\sigma(X)$ событий вида $X^{-1}(B)$, $B \in \mathcal{B}(\mathbb{R})$. Легко проверить, что $\sigma(X)$ является σ -алгеброй. По определению, она называется σ -алгеброй, порожденной случайной величиной X .

Примеры Пусть $\Omega = [0, 1]$, \mathcal{S} — совокупность борелевских подмножеств $[0, 1]$. Тогда если $X(\omega) = \omega$, то $\sigma(X) = \mathcal{S}$. Если же $X = 0$ при $\omega \leq 1/2$ и $X = 1$ при $\omega > 1/2$, то в этом случае $\sigma(X) = \{\emptyset, \Omega, [0, 1/2], (1/2, 1]\}$.

Пусть $\mathcal{A}_1, \dots, \mathcal{A}_n$ — некоторые классы событий. Будем говорить, что они независимы, если для любых $B_1 \in \mathcal{A}_1, \dots, B_n \in \mathcal{A}_n$ выполняется

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

Таким образом, независимость случайных величин X_1, X_2, \dots, X_n может пониматься как независимость σ -алгебр $\sigma(X_1), \dots, \sigma(X_n)$.

Вернемся к соотношению (1). Оказывается, если оно выполняется для всех множеств вида $B_i = (-\infty, y_i)$, $i = 1, \dots, n$, то оно будет иметь место и для любых борелевских множеств $B_i \in \mathcal{B}(\mathbb{R})$, $i = 1, \dots, n$. Приведем доказательство этого факта.

Если (1) справедливо для множеств вида $B_i = (-\infty, y_i)$, то оно справедливо и для множеств вида $B_i = [x_i, y_i)$ — это легко показывается. Значит, (1) будет выполняться также для любых множеств из алгебры \mathcal{A} , элементы которой представимы в виде конечного объединения полуинтервалов такого вида. По-другому можно сказать, что независимыми будут алгебры \mathcal{A}_i , состоящие из событий вида $X_i^{-1}(B)$, $B \in \mathcal{A}$. Нетрудно видеть, что σ -алгебры, порожденные алгебрами \mathcal{A}_i , совпадают с σ -алгебрами $\sigma(X_i)$. Завершает наши рассуждения следующая теорема.

Теорема. Пусть \mathcal{A}_1 и \mathcal{A}_2 — две произвольные независимые алгебры событий. Тогда независимыми будут порожденные ими σ -алгебры $\sigma(\mathcal{A}_1)$ и $\sigma(\mathcal{A}_2)$.

Для доказательства теоремы нам потребуется аппроксимировать события из σ -алгебры событиями из порождающей ее алгебры. Чтобы осуществить это, для всякой пары событий A и B введем функционал

$$d(A, B) = \mathbf{P}(A\bar{B} \cup \bar{A}B) = \mathbf{P}(A\bar{B}) + \mathbf{P}(\bar{A}B) = \mathbf{P}(A \setminus B) + \mathbf{P}(B \setminus A).$$

Это — почти расстояние (равенство $d(A, B) = 0$ не означает, что $A = B$). Введенный функционал обладает следующими полезными свойствами.

- 1) $d(A, B) = d(B, A) = d(\bar{A}, \bar{B})$;
- 2) $d(A, C) \leq d(A, B) + d(B, C)$;
- 3) $d(AB, CD) \leq d(A, C) + d(B, D)$;
- 4) $|\mathbf{P}(A) - \mathbf{P}(B)| \leq d(A, B)$.

Первое из них очевидно. Второе вытекает из соотношений

$$\begin{aligned} d(A, C) &= \mathbf{P}(A\bar{C}) + \mathbf{P}(\bar{A}C) = \mathbf{P}(A\bar{C}\bar{B}) + \mathbf{P}(A\bar{C}B) + \mathbf{P}(\bar{A}C\bar{B}) + \mathbf{P}(\bar{A}CB) \leq \\ &\leq \mathbf{P}(\bar{C}B) + \mathbf{P}(A\bar{B}) + \mathbf{P}(\bar{A}B) + \mathbf{P}(C\bar{B}) = d(A, B) + d(B, C). \end{aligned}$$

Докажем третье. С помощью формулы двойственности получаем

$$\begin{aligned} \mathbf{P}(AB\bar{C}\bar{D} \cup \bar{A}\bar{B}CD) &= \mathbf{P}(AB(\bar{C} \cup \bar{D})) + \mathbf{P}((\bar{A} \cup \bar{B})CD) \leq \\ &\leq \mathbf{P}(AB\bar{C}) + \mathbf{P}(AB\bar{D}) + \mathbf{P}(\bar{A}CD) + \mathbf{P}(\bar{B}CD) \leq \\ &\leq \mathbf{P}(A\bar{C}) + \mathbf{P}(B\bar{D}) + \mathbf{P}(\bar{A}C) + \mathbf{P}(\bar{B}D). \end{aligned}$$

Четвертое свойство следует из представлений

$$\mathbf{P}(A) = \mathbf{P}(A \setminus B) + \mathbf{P}(AB), \quad \mathbf{P}(B) = \mathbf{P}(B \setminus A) + \mathbf{P}(AB).$$

Лемма об аппроксимации. Пусть \mathcal{A} — некоторая алгебра событий. Тогда для любого события $A \in \sigma(\mathcal{A})$ существует последовательность событий $\{A_n\}$, $A_n \in \mathcal{A}$, такая, что

$$\lim_{n \rightarrow \infty} d(A_n, A) = 0$$

(а это значит, что $\mathbf{P}(A) = \lim_{n \rightarrow \infty} \mathbf{P}(A_n)$).

Доказательство. Обозначим \mathcal{A}^* множество событий, аппроксимируемых в этом смысле. Ясно, что $\mathcal{A} \subset \mathcal{A}^*$, т. к. всегда $d(A, A) = 0$. Далее покажем, что \mathcal{A}^* — σ -алгебра. Отсюда будет следовать $\sigma(\mathcal{A}) \subset \mathcal{A}^*$, поскольку $\sigma(\mathcal{A})$ — наименьшая из всех σ -алгебр, содержащих \mathcal{A} .

То, что \mathcal{A}^* является алгеброй, легко проверяется, поскольку соотношения $A \in \mathcal{A}^*$, $B \in \mathcal{A}^*$ влекут за собой принадлежность \mathcal{A}^* множеств \bar{A} , $A \cup B$, AB . Например, если $d(A, A_n) \rightarrow 0$, $d(B, B_n) \rightarrow 0$, то в силу третьего свойства $d(AB, A_n B_n) \leq d(A, A_n) + d(B, B_n) \rightarrow 0$, т. е. $AB \in \mathcal{A}^*$. Кроме того $d(\bar{A}, \bar{A}_n) \rightarrow 0$, т. е. $\bar{A} \in \mathcal{A}^*$, $A \cup B = \overline{\bar{A} \bar{B}}$.

Пусть теперь

$$C = \bigcup_{k=1}^{\infty} C_k, \quad C_k \in \mathcal{A}^*.$$

Так как \mathcal{A}^* — алгебра, то $D_n = \bigcup_{k=1}^n C_k \in \mathcal{A}^*$, при этом

$$d(D_n, C) = \mathbf{P}(C \setminus D_n) = \mathbf{P}(C) - \mathbf{P}(D_n) \rightarrow 0$$

в силу свойства непрерывности вероятности. Поэтому можно выбрать $A_n \in \mathcal{A}$ так, чтобы $d(D_n, A_n) \leq 1/n$, откуда сразу получаем

$$d(C, A_n) \leq d(C, D_n) + d(D_n, A_n) \rightarrow 0.$$

Тем самым мы установили, что $C \in \mathcal{A}^*$, т. е. \mathcal{A}^* образует σ -алгебру. Лемма доказана.

Вернемся к доказательству теоремы. Для каждого из двух событий $A_1 \in \sigma(\mathcal{A}_1)$ и $A_2 \in \sigma(\mathcal{A}_2)$ выберем последовательности событий $\{A_{1n}\}$ из \mathcal{A}_1 и $\{A_{2n}\}$ из \mathcal{A}_2 так, что $d(A_i, A_{in}) \rightarrow 0$. Тогда

$$d(A_1 A_2, A_{1n} A_{2n}) \leq d(A_1, A_{1n}) + d(A_2, A_{2n}) \rightarrow 0,$$

$$\mathbf{P}(A_1 A_2) = \lim_{n \rightarrow \infty} \mathbf{P}(A_{1n} A_{2n}) = \lim_{n \rightarrow \infty} \mathbf{P}(A_{1n}) \mathbf{P}(A_{2n}) = \mathbf{P}(A_1) \mathbf{P}(A_2).$$

Теорема доказана.

Если случайные величины X_1, X_2, \dots, X_n дискретны, то определение их независимости удобно давать в следующей эквивалентной форме.

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

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

2.4. 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

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

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).

$$3. \text{ а) } \lim_{y_n \rightarrow -\infty} F_{X_1, \dots, X_n}(y_1, \dots, y_n) = 0,$$

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

$$\text{В частности, } F_{X_1}(y_1) = \lim_{y_2 \rightarrow \infty, \dots, y_n \rightarrow \infty} F_{X_1, \dots, X_n}(y_1, y_2, \dots, y_n).$$

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 $\{X_1 < y_1, X_2 < y_2, \dots, X_{n-1} < y_{n-1}\}$.

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

$$F_{X_1, \dots, X_n}(y_1, \dots, y_n) = F_{X_1}(y_1) \dots F_{X_n}(y_n). \quad (2)$$

Мы установили в предыдущем разделе, что это соотношение можно использовать в качестве определения независимости случайных величин.

Как видим, 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.

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

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

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 (2) поочередным дифференцированием по каждой переменной, а само оно превращается в соотношение (2) 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.

2.5. 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 — борелевские функции из \mathbb{R} в \mathbb{R} . Тогда случайные величины $g(X)$ и $h(Y)$ также независимы.

Напомним, что функция g называется борелевской, если $g^{-1}(B) \in \mathcal{B}(\mathbb{R})$ для любого $B \in \mathcal{B}(\mathbb{R})$.

Доказательство. Для любых $B_1 \in \mathcal{B}(\mathbb{R})$, $B_2 \in \mathcal{B}(\mathbb{R})$

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

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 $f_X(t)$. Образум новую случайную величину $Y = g(X)$, где g — некоторая неслучайная борелевская функция из \mathbb{R} в \mathbb{R} . 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 $a \neq 0$.

1. 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

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

2. 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, доказано следующее утверждение.

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

$$f_{aX+b}(t) = \frac{1}{|a|} f_X\left(\frac{t-b}{a}\right). \quad (3)$$

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 (3). 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.

Далее поговорим о *квантильных* преобразованиях. Вообще, для монотонной непрерывной функции распределения *квантилью* q_y называется $q_y = F^{-1}(y)$. Если же F — произвольная функция распределения, то квантиль можно определить как

$$q_y = \sup\{t : F(t) < y\}.$$

Теорема 3. Пусть $X \in F$ и F непрерывна. Тогда $Y = F(X) \in U_{0,1}$.

Доказательство будет простым, если дополнительно предположить, что F строго монотонна. Тогда определена обратная функция F^{-1} и для $y \in (0, 1)$

$$\mathbf{P}(F(X) < y) = \mathbf{P}(X < F^{-1}(y)) = F(F^{-1}(y)) = y. \quad (4)$$

Если же F не является строго монотонной, то можно положить $F^{-1}(y) = q_y$. Нетрудно видеть, что при этом соотношение (4) по-прежнему сохраняется в силе.

Из аналогичных рассуждений следует, что если $Y \in U_{0,1}$ и F — некоторая непрерывная функция распределения, то мы можем построить новую случайную величину $X \in F$, взяв $X = F^{-1}(Y)$. Этот факт часто используется при моделировании случайных величин с заданной функцией распределения.

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.

Итак, пусть 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, очевидно, $X + Y$ также будет принимать возможные значения $k = 0, 1, 2, \dots$,

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

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 $\{q_k, \quad k = 0, 1, 2, \dots\}$.

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 .

Заметим, что мы попутно установили формулу свертки и для абсолютно непрерывных функций распределения:

$$F_{X+Y}(y) = \int_{-\infty}^{\infty} f_X(u) \int_{-\infty}^{y-u} f_Y(v) dv du = \int_{-\infty}^{\infty} F_Y(y-u) dF_X(u).$$

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

$$Y_1 = \frac{X_1 - \alpha_1}{\sigma_1}, \quad Y_2 = \frac{X_2 - \alpha_2}{\sigma_2}.$$

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

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

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

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. Интеграл по вероятностной мере. Математическое ожидание

Пусть имеется некоторое вероятностное пространство $\langle \Omega, \mathcal{S}, \mathbf{P} \rangle$ и случайная величина X , заданная на нем. Наша ближайшая цель — построить конструкцию интеграла от этой случайной величины по вероятностной мере, который будет обозначаться

$$\mathbf{E} X = \int_{\Omega} X(\omega) \mathbf{P}(d\omega).$$

и называться *математическим ожиданием* случайной величины X .

Сделаем это в несколько этапов.

I. Интеграл от простой случайной величины. По определению, случайная величина называется простой, если она дискретна и множество ее значений конечно. Индикатор события

$$I_A(\omega) = \begin{cases} 1, & \omega \in A, \\ 0, & \text{иначе} \end{cases}$$

— простая случайная величина. Произвольную простую случайную величину можно записать в виде

$$X(\omega) = \sum_{k=1}^n y_k I_{A_k}(\omega), \quad (5)$$

где $A_k = \{\omega : X(\omega) = y_k\}$. Множества A_1, \dots, A_n не пересекаются и $\bigcup_{k=1}^n A_k = \Omega$.

Определение. Интегралом от простой случайной величины (5) называется

$$\mathbf{E} X = \int_{\Omega} X(\omega) \mathbf{P}(d\omega) = \sum_{k=1}^n y_k \mathbf{P}(A_k)$$

(иногда также используются обозначения $\mathbf{E} X = \int X d\mathbf{P} = \int X(\omega) d\mathbf{P}(\omega)$).

Легко видеть, что математическое ожидание есть среднее взвешенное значений случайной величины; вероятности значений выступают в роли весовых коэффициентов и поэтому более вероятные значения вносят бóльший вклад в сумму.

Следующие свойства математических ожиданий простых случайных величин легко следуют из определения.

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

3. Если $X \leq Y$, то $\mathbf{E} X \leq \mathbf{E} Y$.

4. $|\mathbf{E} X| \leq \mathbf{E} |X|$.

5. $\mathbf{E}(X + Y) = \mathbf{E} X + \mathbf{E} Y$.

Докажем последнее соотношение. Если X принимает возможные значения x_1, x_2, \dots, x_m , а Y — возможные значения y_1, y_2, \dots, y_n , то $X + Y$ будет принимать возможные значения вида $x_i + y_j$, на множестве $\{X = x_i, Y = y_j\}$, $i = 1, 2, \dots, m$, $j = 1, 2, \dots, n$, и

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

Определение. Интегралом по множеству $A \in \mathcal{S}$ от простой случайной величины X называется

$$\mathbf{E}(X; A) = \int_A X(\omega) \mathbf{P}(d\omega) = \int_{\Omega} X(\omega) I_A(\omega) \mathbf{P}(d\omega) = \sum_{k=1}^n y_k \mathbf{P}(X = y_k, A).$$

Напомним, что перечисление событий через запятую соответствует их пересечению.

II. Следующим шагом мы будем строить математическое ожидание для произвольной неотрицательной случайной величины $X \geq 0$.

Лемма 1. Пусть $X \geq 0$. Тогда существует последовательность $\{X_n\}$ неотрицательных простых случайных величин такая, что $X_n(\omega) \nearrow X(\omega)$ для любого ω при $n \rightarrow \infty$.

Доказательство. Разобьем отрезок $[0, n]$ на $n2^n$ равных отрезков. Пусть $y_0 = 0, \dots, y_{n2^n} = n$ — точки деления, $y_{i+1} - y_i = 1/2^n$. Положим

$$A_i = \{\omega : y_i \leq X(\omega) < y_{i+1}\}, \quad i = 1, \dots, n2^n - 1,$$

$$A_0 = \{\omega : 0 \leq X(\omega) < y_1\} \cup \{\omega : X(\omega) \geq n\}.$$

Тогда

$$X_n(\omega) \equiv \sum_{i=0}^{n2^n-1} y_i I_{A_i}(\omega) \leq X(\omega).$$

Очевидно, X_n — измеримые функции, т. е. являются случайными величинами, $X_n(\omega)$ не убывает по n и для любого ω при $n > X(\omega)$ имеет место

$$0 \leq X(\omega) - X_n(\omega) \leq \frac{1}{2^n}.$$

Лемма доказана.

Лемма 2. Пусть $X \geq 0$, $X_n \nearrow X$ и $Y_n \nearrow X$ — последовательности неотрицательных простых случайных величин. Тогда

$$\lim_{n \rightarrow \infty} \mathbf{E}X_n = \lim_{n \rightarrow \infty} \mathbf{E}Y_n.$$

Заметим, что в силу того, что последовательности X_n, Y_n не убывают по n , то же самое верно и для последовательностей $\mathbf{E}X_n, \mathbf{E}Y_n$, значит оба предела в формулировке теоремы существуют, хотя и могут принимать значения $+\infty$.

Доказательство. Убедимся, что для любого m

$$\mathbf{E}X_m \leq \lim_{n \rightarrow \infty} \mathbf{E}Y_n.$$

Случайная величина X_m простая, поэтому с вероятностью 1 она ограничена: $X_m \leq c_m$. Тогда для любого $\varepsilon > 0$ и для любого n

$$X_m - Y_n = (X_m - Y_n)I_{\{X_m - Y_n < \varepsilon\}} + (X_m - Y_n)I_{\{X_m - Y_n \geq \varepsilon\}} \leq \varepsilon + c_m I_{\{X_m - Y_n \geq \varepsilon\}}.$$

Отсюда

$$\mathbf{E}X_m \leq \varepsilon + c_m \mathbf{P}(X_m - Y_n \geq \varepsilon) + \mathbf{E}Y_n,$$

$$\mathbf{P}(X_m - Y_n \geq \varepsilon) \leq \mathbf{P}(X - Y_n \geq \varepsilon) \rightarrow 0,$$

поскольку, обозначив $B_n = \{X - Y_n \geq \varepsilon\}$, будем иметь $B_1 \supset B_2 \supset \dots$ и по свойству непрерывности вероятности

$$\lim \mathbf{P}(B_n) = \mathbf{P}(\cap B_n) = \mathbf{P}(\emptyset) = 0.$$

Множество $\cap B_n$ пусто, так как если $\omega \in \cap B_n$, то $X(\omega) - Y_n(\omega) \geq \varepsilon$ для любого n , что невозможно.

Устремляя теперь $n \rightarrow \infty$, получим

$$\mathbf{E}X_m \leq \varepsilon + \lim_{n \rightarrow \infty} \mathbf{E}Y_n.$$

Так как ε произвольно, то

$$\mathbf{E}X_m \leq \lim_{n \rightarrow \infty} \mathbf{E}Y_n, \quad \text{и} \quad \lim_{m \rightarrow \infty} \mathbf{E}X_m \leq \lim_{n \rightarrow \infty} \mathbf{E}Y_n.$$

Меняя местами X_m и Y_n , получим противоположное неравенство. Лемма доказана.

Таким образом, для $X \geq 0$ можно положить по определению

$$\mathbf{E} X = \int_{\Omega} X d\mathbf{P} = \lim_{n \rightarrow \infty} \mathbf{E} X_n,$$

причем будем говорить, что интеграл существует, если $\mathbf{E} X < \infty$.

III. Если X — произвольная случайная величина, то $X = X^+ - X^-$, где

$$X^+ = \max(0, X), \quad X^- = \max(0, -X),$$

и по определению положим $\mathbf{E} X = \mathbf{E} X^+ - \mathbf{E} X^-$. Будем считать, что $\mathbf{E} X$ существует, если $\mathbf{E} |X| = \mathbf{E} X^+ + \mathbf{E} X^- < \infty$.

Если существует $\mathbf{E} X$, то для любого $A \in \mathcal{S}$ существует также и

$$\mathbf{E}(X; A) = \int_A X(\omega) \mathbf{P}(d\omega) = \int_{\Omega} X(\omega) I_A(\omega) \mathbf{P}(d\omega).$$

Закончив с определениями, мы должны подробно рассмотреть основные свойства интеграла. Сразу отметим, что приведенная конструкция отличается от хорошо известного интеграла Римана тем, что интегральные суммы формируются разбиением на мелкие кусочки не области интегрирования, а области значений интегрируемой функции.

Основные свойства математического ожидания

1. $\mathbf{E}(\alpha X) = \alpha \mathbf{E} X$.
2. Если $X \leq Y$, то $\mathbf{E} X \leq \mathbf{E} Y$.
3. $|\mathbf{E} X| \leq \mathbf{E} |X|$.
4. $\mathbf{E}(X + Y) = \mathbf{E} X + \mathbf{E} Y$.

Перечисленные свойства выполняются для простых случайных величин, после чего предельным переходом они устанавливаются для произвольных случайных величин. Разумеется, везде речь идет о случайных величинах, для которых математическое ожидание существует.

5. Если $X \geq 0$, то для любого $\varepsilon > 0$

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

Этот факт известен как *неравенство Чебышева*. Назовем его первым неравенством Чебышева, поскольку далее в курсе будет получено второе.

Доказательство. Имеем

$$\mathbf{E} X = \int_{\Omega} X d\mathbf{P} \geq \int_{\omega: X(\omega) \geq \varepsilon} X d\mathbf{P} \geq \varepsilon \int_{\omega: X(\omega) \geq \varepsilon} d\mathbf{P} = \varepsilon \mathbf{P}(X \geq \varepsilon).$$

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

Это сразу же следует из установленного выше неравенства Чебышева, поскольку для любого $\varepsilon > 0$ имеем $\mathbf{P}(X \geq \varepsilon) = 0$, что возможно только при $\mathbf{P}(X = 0) = 1$.

7. Неравенство Коши — Буняковского:

$$\mathbf{E} |XY| \leq \sqrt{\mathbf{E} X^2 \mathbf{E} Y^2}.$$

Для доказательства следует воспользоваться неравенством $2|ab| \leq a^2 + b^2$, положив в нем

$$a^2 = \frac{X^2}{\mathbf{E}X^2}, \quad b^2 = \frac{Y^2}{\mathbf{E}Y^2},$$

после чего взять математические ожидания от обеих частей неравенства.

8. Пусть $\mathbf{E}|X| < \infty$, и последовательность событий $\{A_n\}$ такова, что $A_i A_j = \emptyset$ при $i \neq j$ и $A = \bigcup_{n=1}^{\infty} A_n$. Тогда

$$\int_A X d\mathbf{P} = \sum_{n=1}^{\infty} \int_{A_n} X d\mathbf{P}, \quad \text{или по-другому} \quad \mathbf{E}(X; A) = \sum_n \mathbf{E}(X; A_n).$$

Это значит, что интеграл как функция множества обладает свойством счетной аддитивности. Для доказательства нам потребуется следующее вспомогательное утверждение.

Лемма. Пусть $X \geq 0$, $\mathbf{E}X < \infty$ и последовательность событий $\{A_n\}$ такова, что $\mathbf{P}(A_n) \rightarrow 0$ при $n \rightarrow \infty$. Тогда $\mathbf{E}(X; A_n) \rightarrow 0$.

Доказательство. Пусть X_m — последовательность неотрицательных простых функций, $X_m \nearrow X$. Обозначим $B_m = \{X \leq m\}$. Тогда

$$X \geq XI_{B_m} \geq X_m I_{B_m},$$

$$\mathbf{E}X \geq \lim_{m \rightarrow \infty} \mathbf{E}X I_{B_m} \geq \lim_{m \rightarrow \infty} \mathbf{E}X_m I_{B_m} = \mathbf{E}X,$$

поскольку $X_m I_{B_m} \nearrow X$. Отсюда получаем

$$\mathbf{E}X = \lim_{m \rightarrow \infty} \mathbf{E}X I_{B_m} = \lim_{m \rightarrow \infty} \mathbf{E}(X; X \leq m),$$

и значит для любого $\varepsilon > 0$ найдется число $m(\varepsilon)$ такое, что при $m \geq m(\varepsilon)$

$$\mathbf{E}X - \mathbf{E}(X; X \leq m) = \mathbf{E}(X; X > m) \leq \varepsilon.$$

Поэтому для таких m

$$\mathbf{E}(X; A_n) = \mathbf{E}(X; \{X \leq m\} A_n) + \mathbf{E}(X; \{X > m\} A_n) \leq m\mathbf{P}(A_n) + \varepsilon$$

и, следовательно, $\overline{\lim}_{n \rightarrow \infty} \mathbf{E}(X; A_n) \leq \varepsilon$. Лемма доказана.

Перейдем к доказательству свойства 8. Достаточно доказать это для $X \geq 0$. Для простых случайных величин утверждение очевидно, так как

$$\{X = y_k\} \cap A = \bigcup_n \{X = y_k\} A_n,$$

$$\int_A X d\mathbf{P} = \sum_k y_k \mathbf{P}(X = y_k, A) = \sum_n \sum_k y_k \mathbf{P}(X = y_k, A_n) = \sum_n \int XI_{A_n} d\mathbf{P}.$$

В общем случае имеем

$$\int_A X d\mathbf{P} = \lim_{m \rightarrow \infty} \int_A X_m d\mathbf{P} = \lim_{m \rightarrow \infty} \sum_n \int_{A_n} X_m d\mathbf{P} = \sum_n \lim_{m \rightarrow \infty} \int_{A_n} X_m d\mathbf{P} = \sum_n \int_{A_n} X d\mathbf{P}.$$

Обоснуем смену порядка действий. При $N \rightarrow \infty$

$$\mathbf{P}\left(\bigcup_{n=N}^{\infty} A_n\right) = \sum_{n=N}^{\infty} \mathbf{P}(A_n) \rightarrow 0$$

в силу сходимости ряда $\sum_n \mathbf{P}(A_n) = \mathbf{P}(A)$. Отсюда с помощью леммы получаем

$$\sum_{n=N}^{\infty} \int_{A_n} X_m d\mathbf{P} = \mathbf{E}\left(X_m; \bigcup_{n=N}^{\infty} A_n\right) \leq \mathbf{E}\left(X; \bigcup_{n=N}^{\infty} A_n\right) \rightarrow 0$$

при $N \rightarrow \infty$ равномерно по m .

9. Пусть $\mathbf{P}(B) > 0$ для некоторого события B . Введем вероятностную меру $\mathbf{P}_B(A) = \mathbf{P}(A/B)$ и новое вероятностное пространство $\langle \Omega, \mathcal{S}, \mathbf{P}_B \rangle$. На этом вероятностном пространстве построим математическое ожидание случайной величины X :

$$\mathbf{E}(X/B) = \int_{\Omega} X(\omega) \mathbf{P}_B(d\omega).$$

Но по определению

$$\mathbf{P}_B(d\omega) = \frac{\mathbf{P}(d\omega \cap B)}{\mathbf{P}(B)}, \quad \text{то есть} \quad \mathbf{E}(X/B) = \frac{1}{\mathbf{P}(B)} \int_{\Omega} X(\omega) \mathbf{P}(d\omega \cap B) = \frac{1}{\mathbf{P}(B)} \mathbf{E}(X; B).$$

Если в условиях свойства 8 $\mathbf{P}(A_j) > 0$ при всех j , то

$$\int_A X d\mathbf{P} = \sum_j \int_{A_j} X d\mathbf{P} = \sum_j \mathbf{E}(X; A_j) = \sum_j \mathbf{P}(A_j) \mathbf{E}(X/A_j).$$

Мы получили тем самым аналог формулы полной вероятности для математических ожиданий.

Далее рассмотрим вопрос о способах вычисления математических ожиданий.

Уже отмечалось ранее, что каждая случайная величина X порождает вероятностную меру на прямой: для любого $B \in \mathcal{B}(\mathbb{R})$

$$\mathbf{P}_X(B) = \mathbf{P}(\omega : X(\omega) \in B).$$

Сделав замену переменных $t = X(\omega)$, получим

$$\mathbf{E}X = \int_{\Omega} X(\omega) \mathbf{P}(d\omega) = \int_{-\infty}^{\infty} t \mathbf{P}_X(dt) = \int_{-\infty}^{\infty} t \mathbf{P}(X \in dt).$$

Достаточно аккуратно выписать определения этих двух интегралов, чтобы убедиться в их равенстве. Для простой случайной величины мы это уже наблюдали:

$$\mathbf{E}X = \sum_k y_k \mathbf{P}(X = y_k) = \sum_k y_k \mathbf{P}_X(y_k).$$

И для произвольной дискретной случайной величины, т. е. когда $\sum_k \mathbf{P}(X = y_k) = 1$, имеем

$$\mathbf{E}X = \sum_k y_k \mathbf{P}(X = y_k), \quad \text{если} \quad \sum_k |y_k| \mathbf{P}(X = y_k) < \infty.$$

Если распределение X абсолютно непрерывно, то $\mathbf{P}(X \in dt) = f_X(t)dt$, и

$$\mathbf{E}X = \int_{-\infty}^{\infty} t f_X(t) dt, \quad \text{если} \quad \int_{-\infty}^{\infty} |t| f_X(t) dt < \infty.$$

Здесь мы имеем непрерывный аналог среднего взвешенного. Переменная t под знаком интеграла есть текущее значение случайной величины, а плотность $f_X(t)$ играет роль весовой функции.

Если g — борелевская функция из \mathbb{R} в \mathbb{R} , то

$$\mathbf{E}g(X) = \int_{\Omega} g(X(\omega))\mathbf{P}(d\omega) = \int_{-\infty}^{\infty} g(t)\mathbf{P}_X(dt).$$

Эта формула превращается в

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

если распределение X абсолютно непрерывно, при этом случайная величина $g(X)$ может и не иметь плотности распределения.

Часто используется запись вида

$$\mathbf{E}X = \int_{-\infty}^{\infty} t dF_X(t).$$

Здесь используется то обстоятельство, что распределение индуцируется приращениями функции распределения:

$$\mathbf{P}_X([a, b)) = F_X(b) - F_X(a).$$

В таком виде интеграл называется интегралом Лебега-Стилтьеса.

Если F_X распределение смешанного типа, то для вычисления математического ожидания полезно сначала выделить компоненты распределения. Пусть, к примеру,

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

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

$$\mathbf{E}X = \alpha \int_{-\infty}^{\infty} t dF_1(t) + \beta \int_{-\infty}^{\infty} t dF_2(t) = \alpha \int_{-\infty}^{\infty} t f(t) dt + \beta \sum_{k=1}^{\infty} y_k p_k,$$

если только абсолютно сходятся участвующие здесь интеграл и сумма ряда.

Подводя итог нашим построениям отметим, что мы определили математическое ожидание в виде интеграла

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

который может превращаться в сумму (для дискретных распределений) или в обычный интеграл (для распределений абсолютно непрерывного типа). Для вычисления математического ожидания достаточно знать только лишь распределение случайной

величины, т. е. математическое ожидание — это числовая характеристика распределения.

Приведем некоторые примеры вычисления математических ожиданий.

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

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

Найдем $\mathbf{E}X$ другим способом. Случайная величина X распределена так же, как и число успехов S_n в n испытаниях схемы Бернулли, поэтому $\mathbf{E}X = \mathbf{E}S_n$. 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 $S_n = X_1 + \dots + X_n$ и $\mathbf{E}X = \mathbf{E}S_n = \mathbf{E}X_1 + \dots + \mathbf{E}X_n = np$.

3. 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.

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

Если $X = (X_1, \dots, X_n)$ — случайный вектор, то по аналогии с одномерным случаем введем распределение

$$\mathbf{P}_X(B) = \mathbf{P}((X_1, \dots, X_n) \in B), \quad B \in \mathcal{B}(\mathbb{R}^n).$$

Тогда для борелевской функции $g: \mathbb{R}^n \rightarrow \mathbb{R}$

$$\mathbf{E}g(X_1, \dots, X_n) = \int_{\Omega} g(X_1, \dots, X_n) d\mathbf{P} = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} g(t_1, \dots, t_n) \mathbf{P}_X(dt_1, \dots, dt_n).$$

Если распределение случайного вектора X абсолютно непрерывно, то

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

где $f_X(t_1, \dots, t_n)$ — плотность совместного распределения случайного вектора $X = (X_1, \dots, X_n)$.

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

Пусть теперь $n = 2$ и случайные величины X_1, X_2 независимы. Тогда для любых борелевских множеств $B_1 \in \mathcal{B}(\mathbb{R}), B_2 \in \mathcal{B}(\mathbb{R})$

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

Это соотношение задает вероятностную меру на «прямоугольниках», что дает возможность однозначно продолжить ее на любые борелевские подмножества \mathbb{R}^2 . В этом случае говорят, что полученная мера $\mathbf{P}_X(dt)$ является прямым произведением мер $\mathbf{P}_{X_1}(dt_1)$ и $\mathbf{P}_{X_2}(dt_2)$ (здесь $X = (X_1, X_2)$, $dt = (dt_1, dt_2)$). Имеет место следующее свойство (теорема Фубини):

$$\begin{aligned} \mathbf{E}g(X_1, X_2) &= \int_{\Omega} g(X_1, X_2)d\mathbf{P} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(t_1, t_2)\mathbf{P}_X(dt) = \\ &= \int_{-\infty}^{\infty} \left\{ \int_{-\infty}^{\infty} g(t_1, t_2)\mathbf{P}_{X_2}(dt_2) \right\} \mathbf{P}_{X_1}(dt_1), \end{aligned}$$

если все участвующие здесь математические ожидания существуют. В частности, при $g(t_1, t_2) = t_1 t_2$ получаем следующее утверждение.

Теорема. Если X_1, X_2 независимы, то $\mathbf{E}(X_1 X_2) = \mathbf{E}X_1 \mathbf{E}X_2$.

Обратное неверно. Для примера можно взять несомненно зависимые случайные величины $X_1 \in U_{-1,1}$ и $X_2 = X_1^2$. Ясно, что

$$\mathbf{E}(X_1 X_2) = \mathbf{E}X_1^3 = \frac{1}{2} \int_{-1}^1 t^3 dy = 0, \quad \mathbf{E}X_1 = 0, \quad \mathbf{E}X_1^2 = \frac{1}{3},$$

то есть $\mathbf{E}(X_1 X_2) \neq \mathbf{E}X_1 \mathbf{E}X_2$.

Вернемся к распределению суммы независимых случайных величин.

$$\begin{aligned} F_{X+Y}(t) &= \mathbf{E}I_{\{X+Y < t\}} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} I_{\{u+v < t\}} \mathbf{P}_X(du) \mathbf{P}_Y(dv) = \\ &= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} I_{\{u+v < t\}} \mathbf{P}_Y(dv) \right) \mathbf{P}_X(du) = \int_{-\infty}^{\infty} \mathbf{E}I_{\{u+Y < t\}} \mathbf{P}_X(du) = \int_{-\infty}^{\infty} \mathbf{P}(u+Y < t) \mathbf{P}_X(du) = \\ &= \int_{-\infty}^{\infty} F_Y(t-u) dF_X(u) = \int_{-\infty}^{\infty} F_X(t-v) dF_Y(v). \end{aligned}$$

Это выражение называется сверткой функций распределения, как и в случае плотностей. Если хотя бы одна из случайных величин (например, Y) обладает плотностью, то распределение $X + Y$ будет абсолютно непрерывным,

$$F_{X+Y}(y) = \int_{-\infty}^{\infty} \int_{-\infty}^{y-u} f_Y(v) dv dF_X(u) =$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^y f_Y(t-u) dt dF_X(u) = \int_{-\infty}^y \int_{-\infty}^{\infty} f_Y(t-u) dF_X(u) dt,$$

т. е.

$$f_{X+Y}(t) = \int_{-\infty}^{\infty} f_Y(t-u) dF_X(u).$$

3.2. Моменты

Определение. Моментом k -го порядка случайной величины X называется

$$\mathbf{E}X^k = \int_{-\infty}^{\infty} t^k dF_X(t), \quad k > 0.$$

Как и всякое математическое ожидание, момент k -го порядка существует тогда и только тогда, когда

$$\mathbf{E}|X|^k = \int_{-\infty}^{\infty} |t|^k dF_X(t) < \infty.$$

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

$$\mathbf{D}X = \mathbf{E}(X - \mathbf{E}X)^2.$$

Другими словами, дисперсия — это второй центральный момент случайной величины. Она действительно показывает, насколько велик разброс значений случайной величины. Вычитая $\mathbf{E}X$ 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:

$$\begin{aligned} \mathbf{D}X &= \mathbf{E}(X - \mathbf{E}X)^2 = \mathbf{E}(X^2 - 2X\mathbf{E}X + (\mathbf{E}X)^2) = \\ &= \mathbf{E}X^2 - 2\mathbf{E}X\mathbf{E}X + (\mathbf{E}X)^2 = \mathbf{E}X^2 - (\mathbf{E}X)^2. \end{aligned}$$

Для дискретных распределений дисперсия вычисляется по формулам

$$\mathbf{D}X = \sum_{k=1}^{\infty} (y_k - \mathbf{E}X)^2 \mathbf{P}(X = y_k) = \sum_{k=1}^{\infty} y_k^2 \mathbf{P}(X = y_k) - (\mathbf{E}X)^2,$$

для распределений абсолютно непрерывного типа имеем

$$\mathbf{D}X = \int_{-\infty}^{\infty} (t - \mathbf{E}X)^2 f_X(t) dt = \int_{-\infty}^{\infty} t^2 f_X(t) dt - (\mathbf{E}X)^2.$$

$\sqrt{\mathbf{D}X}$ называется *стандартным отклонением*.

Свойства дисперсии

Здесь и всюду в дальнейшем буквой C будут обозначаться константы.

1. $\mathbf{D}X \geq 0$. Свойство очевидно.
2. $\mathbf{D}C = 0$. Следует из определения, поскольку $C = \mathbf{E}C$.

3. Если $\mathbf{D}X = 0$, то $\mathbf{P}(X = C) = 1$ для некоторой постоянной C . Действительно, из свойства 6 математических ожиданий вытекает: соотношения $(X - \mathbf{E}X)^2 \geq 0$ и $\mathbf{E}(X - \mathbf{E}X)^2 = 0$ влекут $\mathbf{P}(X - \mathbf{E}X = 0) = 1$.

4. $\mathbf{D}(CX) = C^2\mathbf{D}X$; в частности, $\mathbf{D}(-X) = \mathbf{D}X$. Это вновь следует из определения: $\mathbf{D}(CX) = \mathbf{E}(CX - \mathbf{E}CX)^2 = C^2\mathbf{E}(X - \mathbf{E}X)^2 = C^2\mathbf{D}X$.

5. $\mathbf{D}(X + C) = \mathbf{D}X$. Следует из определения: $\mathbf{E}(X + C - \mathbf{E}(X + C))^2 = \mathbf{E}(X + C - \mathbf{E}X - C)^2 = \mathbf{D}X$.

6. Если X и Y независимы, то $\mathbf{D}(X \pm Y) = \mathbf{D}X + \mathbf{D}Y$.

Доказательство.

$$\begin{aligned}\mathbf{D}(X \pm Y) &= \mathbf{E}(X \pm Y - \mathbf{E}(X \pm Y))^2 = \mathbf{E}((X - \mathbf{E}X) \pm (Y - \mathbf{E}Y))^2 \\ &= \mathbf{E}(X - \mathbf{E}X)^2 + \mathbf{E}(Y - \mathbf{E}Y)^2 \pm 2\mathbf{E}((X - \mathbf{E}X)(Y - \mathbf{E}Y)) \\ &= \mathbf{D}X + \mathbf{D}Y \pm 2\mathit{Cov}(X, Y),\end{aligned}$$

где обозначено

$$\mathit{Cov}(X, Y) = \mathbf{E}((X - \mathbf{E}X)(Y - \mathbf{E}Y)).$$

Эта величина называется *ковариацией* между случайными величинами X и Y . Если X и Y независимы, то $X - \mathbf{E}X$ и $Y - \mathbf{E}Y$ также независимы и поэтому

$$\mathit{Cov}(X, Y) = \mathbf{E}((X - \mathbf{E}X)(Y - \mathbf{E}Y)) = \mathbf{E}(X - \mathbf{E}X)\mathbf{E}(Y - \mathbf{E}Y) = 0.$$

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 матожидания 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. $|\rho(X, Y)| \leq 1$.

Это свойство напрямую получается из неравенства Коши-Буняковского.

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 .

Для доказательства обратного утверждения 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,

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

Таким образом,

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

Пусть теперь, к примеру, $\rho(X, Y) = 1$. Воспользуемся полученным 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 $t = (t_1, \dots, t_n)^T$, $Q = (\mathbf{C}(X))^{-1} = (AC(Y)A^T)^{-1} = (AA^T)^{-1}$.

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

Доказательство. 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 . Теорема доказана.

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 $\mathbf{C}(Y) = A\mathbf{C}(X)A^T = AA^T = E$ 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.

3.7. Математическое ожидание суммы случайного числа слагаемых

Ранее было введено понятие σ -алгебры $\sigma(X) = \{X^{-1}(B), B \in \mathcal{B}(\mathbb{R})\}$, порожденной случайной величиной X . Она состоит из событий, о реализации которых можно судить по значениям X . Пусть имеется теперь последовательность случайных величин $\{X_n\}$. Для любых $k \leq m \leq \infty$ введем σ -алгебру $\sigma(X_k, \dots, X_m)$, которая порождена событиями вида $\bigcap_{i=k}^m A_i$, где $A_i \in \sigma(X_i)$. Это более богатая σ -алгебра,

чем, к примеру, $\sigma(X_k)$ или $\sigma(X_k, \dots, X_{m-1})$, так как некоторые из множеств A_i могут совпадать с Ω .

Теорема. Если случайные величины X_1, X_2, \dots независимы, то для любых $k \geq 1$, $n \geq 1$ σ -алгебры $\sigma(X_1, \dots, X_n)$ и $\sigma(X_{n+k}, X_{n+k+1}, \dots)$ независимы.

Схема доказательства состоит в следующем. Сначала проверяется независимость произвольного события вида $\bigcap_{i=1}^n A_i$ и события вида $\bigcap_{i=n+k}^m A_i$, где $A_i \in \sigma(X_i)$, $n+k \leq m < \infty$. Затем устанавливается независимость алгебр, порожденных событиями такого вида. Как уже доказано ранее, независимость алгебр обеспечивает независимость порожденных ими σ -алгебр. Переход к пределу при $m \rightarrow \infty$ осуществляется с помощью свойства непрерывности вероятности.

Пусть на вероятностном пространстве $\langle \Omega, \mathcal{S}, \mathbf{P} \rangle$ задана последовательность $\{X_n\}$ независимых случайных величин и некоторая целочисленная случайная величина $\nu \geq 1$. Значения n можно интерпретировать как дискретные моменты времени.

Определение. Случайная величина ν называется *не зависящей от будущего*, если для любого n событие $\{\nu \leq n\}$ не зависит от σ -алгебры $\sigma(X_{n+1}, X_{n+2}, \dots)$.

Определение. Если $\{\nu \leq n\} \in \sigma(X_1, \dots, X_n)$ при всех $n \geq 1$, то такая случайная величина ν называется *марковской* или *марковским моментом*.

Ясно, что марковский момент не зависит от будущего, так как $\sigma(X_1, \dots, X_n)$ и $\sigma(X_{n+1}, X_{n+2}, \dots)$ независимы.

Примерами марковских моментов могут служить случайные величины

$$\nu_1 = \inf\{k \geq 1 : X_k \geq N\},$$

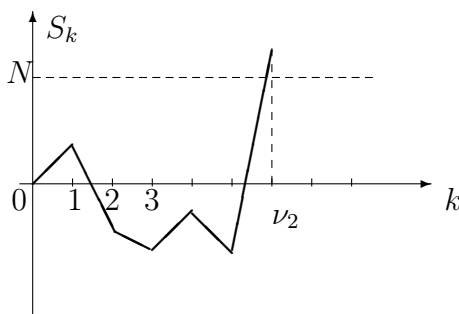
поскольку

$$\{\nu_1 \leq n\} = \bigcup_{k=1}^n \{X_k \geq N\} \in \sigma(X_1, \dots, X_n),$$

а также

$$\nu_2 = \inf\{k \geq 1 : S_k \geq N\},$$

где $S_k = X_1 + \dots + X_k$, $N > 0$ — произвольное положительное число.



Теорема (Колмогоров - Прохоров). Пусть случайная величина ν не зависит от будущего. Тогда если

$$\sum_{k=1}^{\infty} \mathbf{P}(\nu \geq k) \mathbf{E}|X_k| < \infty,$$

то

$$\mathbf{E}S_\nu = \sum_{k=1}^{\infty} \mathbf{P}(\nu \geq k) \mathbf{E}X_k.$$

Доказательство. Предположим сначала, что все случайные величины X_k неотрицательны. Тогда вне зависимости от того, конечен интеграл $\mathbf{E}S_\nu$ или нет, мы можем написать

$$\begin{aligned}\mathbf{E}S_\nu &= \sum_{n=1}^{\infty} \mathbf{E}(S_\nu; \nu = n) = \sum_{n=1}^{\infty} \mathbf{E}(S_n; \nu = n) = \\ &= \sum_{n=1}^{\infty} \sum_{k=1}^n \mathbf{E}(X_k; \nu = n) = \sum_{k=1}^{\infty} \sum_{n=k}^{\infty} \mathbf{E}(X_k; \nu = n) = \sum_{k=1}^{\infty} \mathbf{E}(X_k; \nu \geq k).\end{aligned}$$

Событие $\{\nu \geq k\} = \{\nu > k - 1\} = \Omega \setminus \{\nu \leq k - 1\}$ не зависит от σ -алгебры $\sigma(X_k, X_{k+1}, \dots)$, поэтому оно не зависит от $\sigma(X_k)$. Значит,

$$\mathbf{E}(X_k; \nu \geq k) = \mathbf{E}X_k I_{\{\nu \geq k\}} = \mathbf{E}X_k \mathbf{E}I_{\{\nu \geq k\}} = \mathbf{E}X_k \mathbf{P}(\nu \geq k)$$

и $\mathbf{E}S_\nu = \sum_{k=1}^{\infty} \mathbf{P}(\nu \geq k) \mathbf{E}X_k < \infty$.

Если X_k могут принимать значения разных знаков, то, очевидно,

$$\mathbf{E}|S_\nu| \leq \mathbf{E} \sum_{k=1}^{\nu} |X_k| = \sum_{k=1}^{\infty} \mathbf{P}(\nu \geq k) \mathbf{E}|X_k| < \infty,$$

и мы вновь можем повторить наши рассуждения, обосновав перемену порядка суммирования абсолютной сходимостью ряда

$$\begin{aligned}\sum_{k=1}^{\infty} \sum_{n=k}^{\infty} |\mathbf{E}(X_k; \nu = n)| &\leq \sum_{k=1}^{\infty} \sum_{n=k}^{\infty} \mathbf{E}(|X_k|; \nu = n) = \\ &= \sum_{k=1}^{\infty} \mathbf{E}(|X_k|; \nu \geq k) = \sum_{k=1}^{\infty} \mathbf{E}|X_k| \mathbf{P}(\nu \geq k) < \infty.\end{aligned}$$

Следствие (тождество Вальда). Если X_1, X_2, \dots — независимые одинаково распределенные случайные величины, $\mathbf{E}|X_1| < \infty$, ν не зависит от будущего и $\mathbf{E}\nu < \infty$, то $\mathbf{E}S_\nu = \mathbf{E}X_1 \mathbf{E}\nu$.

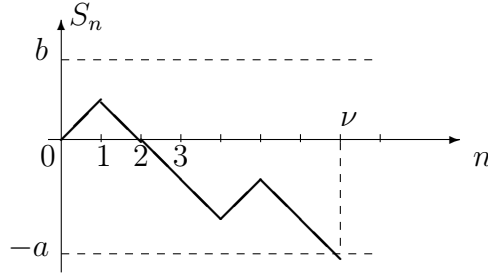
Доказательство. Достаточно заметить, что

$$\sum_{k=1}^{\infty} \mathbf{P}(\nu \geq k) = \sum_{k=1}^{\infty} \sum_{i=k}^{\infty} \mathbf{P}(\nu = i) = \sum_{i=1}^{\infty} i \mathbf{P}(\nu = i) = \mathbf{E}\nu.$$

В качестве примера использования тождества Вальда рассмотрим задачу о разорении.

Два игрока, А и В, подбрасывают монету, и в случае выпадения герба А выигрывает единицу, если решка, то единицу выигрывает В. Это так называемая безобидная игра, игроки каждый раз выигрывают единицу с равной вероятностью. Пусть a и b — начальные капиталы игроков. Игра заканчивается, когда один из игроков проиграет весь свой капитал. Обозначим P_A и P_B вероятности разорения для А и В соответственно. Наша задача — найти эти вероятности.

Если обозначить X_n выигрыш А в n -м туре (он может равняться 1 или -1 с равными вероятностями), то $S_n = X_1 + \dots + X_n$ есть суммарный выигрыш А после n бросаний монеты. Удобно изображать развитие игры на графике.



Покажем, что $P_A + P_B = 1$, т. е. с вероятностью единица траектория обязательно выйдет из полосы. Действительно, для любого натурального N

$$\begin{aligned} 1 - P_A - P_B &= \mathbf{P}(S_1 \in (-a, b), S_2 \in (-a, b), \dots) \leq \\ &\leq \mathbf{P}(|X_1 + \dots + X_N| < a + b, |X_{N+1} + \dots + X_{2N}| < a + b, \dots) = \\ &= \mathbf{P}(|X_1 + \dots + X_N| < a + b) \mathbf{P}(|X_{N+1} + \dots + X_{2N}| < a + b) \dots = 0, \end{aligned}$$

если N таково, что $\mathbf{P}(|X_1 + \dots + X_N| < a + b) < 1$. Но для этого достаточно взять $N = a + b$, так как при этом

$$\mathbf{P}(|X_1 + \dots + X_N| = a + b) = \frac{2}{2^{a+b}} > 0$$

и

$$\mathbf{P}(|X_1 + \dots + X_N| < a + b) = 1 - \frac{2}{2^{a+b}} < 1.$$

Обозначим через ν момент прекращения игры, т. е. разорения одного из игроков. Ясно, что ν — марковский момент; $S_\nu = -a$ с вероятностью P_A и $S_\nu = b$ с вероятностью P_B . Поэтому в силу тождества Вальда

$$\mathbf{E}S_\nu = -aP_A + b(1 - P_A) = \mathbf{E}\nu \mathbf{E}X_1 = 0,$$

если $\mathbf{E}\nu < \infty$. Отсюда сразу следует $P_A = \frac{b}{a+b}$. Отметим, что $P_A = 1$, если $a < \infty$, $b = \infty$, то есть внешне безобидная игра с бесконечно богатым соперником всегда заканчивается разорением.

Покажем, что $\mathbf{E}\nu < \infty$. Вычисление вероятностей $\mathbf{P}(\nu = k)$ является весьма сложной задачей, поэтому мы подходящим образом оценим сверху $\mathbf{P}(\nu \geq k)$ и воспользуемся формулой $\mathbf{E}\nu = \sum_{k=1}^{\infty} \mathbf{P}(\nu \geq k)$. Имеем для уже выбранного числа N

$$\begin{aligned} \mathbf{P}(\nu > mN) &\leq \mathbf{P}(|X_1 + \dots + X_N| < a + b, \dots, |X_{(m-1)N+1} + \dots + X_{mN}| < a + b) = \\ &= \left(1 - \frac{2}{2^{a+b}}\right)^m, \end{aligned}$$

$$\begin{aligned} \mathbf{E}\nu &= \sum_{k=1}^{\infty} \mathbf{P}(\nu \geq k) = \sum_{m=0}^{\infty} \sum_{mN < k \leq (m+1)N} \mathbf{P}(\nu \geq k) \leq \\ &\leq \sum_{m=0}^{\infty} N \mathbf{P}(\nu > mN) \leq N \sum_{m=0}^{\infty} \left(1 - \frac{2}{2^{a+b}}\right)^m < \infty. \end{aligned}$$

3.8. Условное математическое ожидание

Сначала введем понятие условного распределения. Пусть X и Y — две дискретные случайные величины, заданные на одном вероятностном пространстве. Обозначим

$$p_X(u/v) = \frac{\mathbf{P}(X = u, Y = v)}{\mathbf{P}(Y = v)}$$

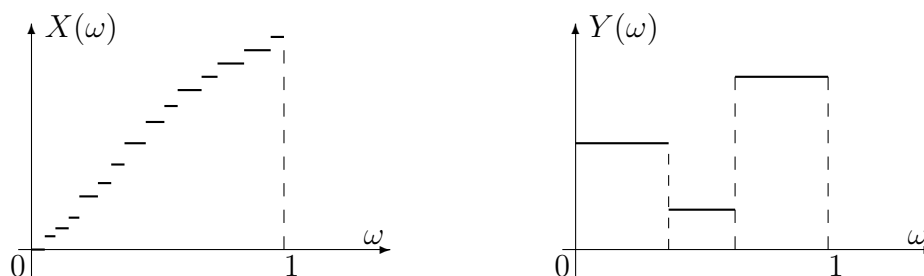
и назовем эту функцию условным распределением случайной величины X при условии $Y = v$. Здесь переменная v принимает только те значения, для которых $\mathbf{P}(Y = v) > 0$.

Далее рассмотрим математическое ожидание, соответствующее этому условному распределению (если оно существует):

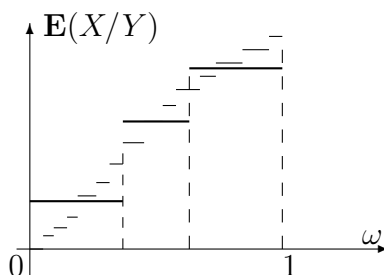
$$\mathbf{E}(X/v) = \sum_u u p_X(u/v).$$

Мы получили тем самым функцию от текущего значения v случайной величины Y . Рассматривая теперь $\mathbf{E}(X/Y)$ как случайную величину, которая каждому ω ставит в соответствие $\mathbf{E}(X/Y(\omega))$, мы получаем *условное математическое ожидание* X относительно Y .

Проиллюстрируем введенное понятие на рисунке. Пусть, как и ранее, $\Omega = [0, 1]$, и графики случайных величин выглядят следующим образом.



А вот как выглядит график случайной величины $\mathbf{E}(X/Y)$ (он показан жирной линией):



Случайная величина $\mathbf{E}(X/Y)$ получается из X следующим образом: на каждом из трех отрезков, где Y постоянна, X заменяется на среднее значение, вычисленное по этому отрезку.

Из приведенного выше определения непосредственно вытекают следующие свойства.

1. Если X и Y независимы, то $\mathbf{E}(X/Y) = \mathbf{E}X$. Это очевидно.
2. Для любой борелевской функции g имеет место

$$\mathbf{E}(g(Y)X/Y) = g(Y)\mathbf{E}(X/Y),$$

т. е. $g(Y)$ выносится из-под знака математического ожидания подобно константе. В частности, при $X = 1$ получаем $\mathbf{E}(g(Y)/Y) = g(Y)$.

Доказательство.

$$p_{g(Y)X}(u/v) = \frac{\mathbf{P}(g(Y)X = u, Y = v)}{\mathbf{P}(Y = v)} = \frac{\mathbf{P}(g(v)X = u, Y = v)}{\mathbf{P}(Y = v)},$$

поэтому

$$\mathbf{E}(g(Y)X/v) = \sum_u u \frac{\mathbf{P}(g(v)X = u, Y = v)}{\mathbf{P}(Y = v)} = \sum_t g(v)t \frac{\mathbf{P}(X = t, Y = v)}{\mathbf{P}(Y = v)} = g(v)\mathbf{E}(X/v).$$

3. $\mathbf{E}(X_1 + X_2/Y) = \mathbf{E}(X_1/Y) + \mathbf{E}(X_2/Y)$.

Доказательство.

$$\begin{aligned} \mathbf{E}(X_1 + X_2/v) &= \sum_u u \frac{\mathbf{P}(X_1 + X_2 = u, Y = v)}{\mathbf{P}(Y = v)} = \sum_{t_1, t_2} (t_1 + t_2) \frac{\mathbf{P}(X_1 = t_1, X_2 = t_2, Y = v)}{\mathbf{P}(Y = v)} = \\ &= \sum_{t_1} t_1 \sum_{t_2} \frac{\mathbf{P}(X_1 = t_1, X_2 = t_2, Y = v)}{\mathbf{P}(Y = v)} + \sum_{t_2} t_2 \sum_{t_1} \frac{\mathbf{P}(X_1 = t_1, X_2 = t_2, Y = v)}{\mathbf{P}(Y = v)} = \\ &= \mathbf{E}(X_1/v) + \mathbf{E}(X_2/v). \end{aligned}$$

4. $\mathbf{E}X = \mathbf{E}[\mathbf{E}(X/Y)]$.

Доказательство.

$$\begin{aligned} \mathbf{E}X &= \sum_u u \mathbf{P}(X = u) = \sum_u u \sum_v \frac{\mathbf{P}(X = u, Y = v)}{\mathbf{P}(Y = v)} \mathbf{P}(Y = v) = \\ &= \sum_v \sum_u u p_X(u/v) \mathbf{P}(Y = v) = \sum_v \mathbf{E}(X/v) \mathbf{P}(Y = v). \end{aligned}$$

Далее введем условное математическое ожидание для случайных величин, имеющих абсолютно непрерывное совместное распределение.

Пусть $f_{X,Y}(u, v)$ — плотность совместного распределения пары случайных величин (X, Y) , $f_X(u)$ и $f_Y(v)$ — соответственно плотности случайных величин X и Y . Для тех значений v , для которых $f_Y(v) > 0$, введем условную плотность

$$f_X(u/v) = \frac{f_{X,Y}(u, v)}{f_Y(v)}$$

и математическое ожидание (если оно существует)

$$\mathbf{E}(X/v) = \int_{-\infty}^{\infty} u f_X(u/v) du.$$

Подставляя в это выражение Y вместо v , получим условное математическое ожидание $\mathbf{E}(X/Y)$.

Перечисленные выше свойства 1 – 4 сохраняются в силе и в этом случае, однако доказательства будут выглядеть сложнее и мы их не будем приводить. Отметим только, что во всех этих конструкциях X и Y могут быть векторными — это не вызовет принципиальных изменений.

3.9. Задача о наилучшем приближении

В этом разделе мы будем рассматривать случайные величины, заданные на одном вероятностном пространстве и обладающие конечным вторым моментом. Для любых двух таких величин X, Y введем скалярное произведение $(X, Y) = \mathbf{E}(XY)$. Оно конечно в силу неравенства Коши–Буняковского. Введем также норму

$$\|X\|^2 = (X, X) = \mathbf{E}X^2.$$

Сходимость последовательности случайных величин $\{X_n\}$ по этой норме называется *сходимостью в среднем квадратическом*, $\|X_n - X\| \rightarrow 0 \iff \mathbf{E}(X_n - X)^2 \rightarrow 0$. Можно показать, что $\|X_n - X\| \rightarrow 0$ при $n \rightarrow \infty \iff \|X_n - X_m\| \rightarrow 0$ при $n, m \rightarrow \infty$.

Пусть H — линейное пространство случайных величин с конечным вторым моментом, замкнутое относительно сходимости в среднем квадратическом. Пусть также L — некоторое замкнутое линейное подпространство в H . Задача состоит в том, чтобы приблизить произвольную случайную величину $X \in H$ с помощью случайной величины $Y \in L$.

Назовем $\hat{X} \in L$ *наилучшим приближением* для $X \in H$, если

$$\|X - \hat{X}\| = \min_{Y \in L} \|X - Y\|.$$

Теорема. $\hat{X} \in L$ является наилучшим приближением для $X \in H$ тогда и только тогда, когда

$$(X - \hat{X}, Y) = 0 \quad \text{для любого } Y \in L,$$

т. е. $X - \hat{X}$ ортогонально L и в этом смысле \hat{X} есть «проекция» X на L .

Доказательство. Пусть \hat{X} — наилучшее приближение, тогда

$$\|X - \hat{X}\| = \min_{Y \in L} \|X - Y\| = \min_{\Delta \in L} \|X - \hat{X} - \Delta\|,$$

где $\Delta = Y - \hat{X}$. Зафиксируем некоторое $\Delta_0 \neq 0$ из L и рассмотрим минимум по выделенному направлению:

$$\min_{\lambda} \|X - \hat{X} - \lambda\Delta_0\|^2 = \min_{\lambda} (X - \hat{X} - \lambda\Delta_0, X - \hat{X} - \lambda\Delta_0) = \min_{\lambda} (A^2 - 2\lambda B + \lambda^2 C^2),$$

где обозначено

$$A^2 = \|X - \hat{X}\|^2, \quad B = (X - \hat{X}, \Delta_0), \quad C^2 = \|\Delta_0\|^2.$$

Очевидно, что при $\lambda = 0$ достигается минимум левой части, причем минимум по направлению совпадает с минимумом по всему пространству L . Следовательно, минимум правой части, где стоит квадратный трехчлен, также должен достигаться при $\lambda = 0$, т. е. $B = 0$. Это значит, что

$$(X - \hat{X}, \Delta) = 0 \quad \text{для любого } \Delta \in L.$$

Обратно, пусть для некоторого $\hat{X} \in L$ имеет место

$$(X - \hat{X}, Y) = 0 \quad \text{для любого } Y \in L.$$

Тогда

$$\|X - Y\|^2 = \|(X - \hat{X}) - (Y - \hat{X})\|^2 =$$

$$= \|X - \widehat{X}\|^2 - 2(X - \widehat{X}, Y - \widehat{X}) + \|Y - \widehat{X}\|^2 = \|X - \widehat{X}\|^2 + \|Y - \widehat{X}\|^2 \geq \|X - \widehat{X}\|^2,$$

и минимум достигается при $Y = \widehat{X}$. Теорема доказана.

Отметим, что наилучшее приближение единственно. Действительно, если $\widehat{X}_1 \in L$ и $\widehat{X}_2 \in L$ таковы, что

$$(X - \widehat{X}_1, Y) = 0, \quad (X - \widehat{X}_2, Y) = 0 \quad \text{для любого } Y \in L,$$

то, в частности, $(X - \widehat{X}_1, \widehat{X}_2 - \widehat{X}_1) = 0$ и $(X - \widehat{X}_2, \widehat{X}_2 - \widehat{X}_1) = 0$. Вычитая одно равенство из другого, получим $\|\widehat{X}_2 - \widehat{X}_1\| = 0$, т. е. $\widehat{X}_2 = \widehat{X}_1$ почти наверное.

Рассмотрим далее *задачу прогноза*.

Предположим, что n раз проводились эксперименты, в результате которых получены случайные величины Y_1, Y_2, \dots, Y_n . Нам предстоит провести следующий по счету эксперимент и получить в результате него случайную величину X . Можем ли мы с некоторой точностью спрогнозировать значение X , если совместное распределение вектора $(X, Y_1, Y_2, \dots, Y_n)$ нам известно?

Будем предполагать, что распределение этого вектора дискретно или абсолютно непрерывно и все его компоненты обладают конечными вторыми моментами, т. е. принадлежат H . Обозначим $Y = (Y_1, Y_2, \dots, Y_n)$. Рассмотрим подпространство $L \subset H$ всех случайных величин вида $g(Y)$, где g может быть произвольной борелевской функцией такой, что $\mathbf{E}g^2(Y) < \infty$. Мы будем приближать X функциями от уже имеющихся случайных величин, т. е. элементами из L .

Теорема. *Наилучшим приближением для X является $\widehat{X} = \mathbf{E}(X/Y)$.*

Доказательство. То, что \widehat{X} есть функция от Y , следует из определения условного математического ожидания. Покажем, что $\mathbf{E}\widehat{X}^2 < \infty$. Имеем $[\mathbf{E}(X/v)]^2 \leq \mathbf{E}(X^2/v)$, так как дисперсия условного распределения неотрицательна. Отсюда следует неравенство для случайных величин $\widehat{X}^2 \leq \mathbf{E}(X^2/Y)$ и для математических ожиданий этих величин:

$$\mathbf{E}\widehat{X}^2 \leq \mathbf{E}[\mathbf{E}(X^2/Y)] = \mathbf{E}X^2 < \infty.$$

Проверим далее условие ортогональности $\mathbf{E}[(X - \widehat{X})g(Y)] = 0$ для любой случайной величины $g(Y) \in L$. Имеем

$$\mathbf{E}[Xg(Y)] = \mathbf{E}[\mathbf{E}(Xg(Y)/Y)] = \mathbf{E}[g(Y)\mathbf{E}(X/Y)] = \mathbf{E}(\widehat{X}g(Y)).$$

Теорема доказана.

4. Сходимость случайных величин и распределений. Предельные теоремы

4.1. Сходимость последовательностей случайных величин

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 рассмотрим некоторые типы сходимости.

Пусть случайные величины X, X_1, X_2, \dots заданы на одном и том же вероятностном пространстве. Ранее уже вводился один тип сходимости — сходимость в

среднем квадратическом. Напомним, что $X_n \rightarrow X$ в среднем квадратическом, если $\mathbf{E}(X_n - X)^2 \rightarrow 0$ при $n \rightarrow \infty$. Далее введем понятие *сходимости по вероятности*.

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 $X_n \xrightarrow{P} X$.

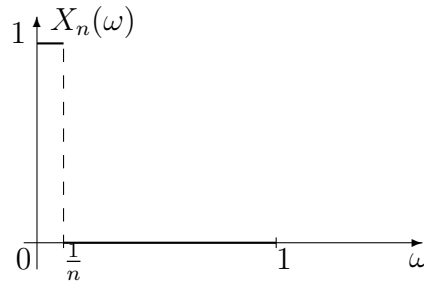
Эквивалентное определение: для любого $\varepsilon > 0$

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

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, $\Omega = [0, 1]$, $\mathcal{S} = \mathcal{B}(\Omega)$. 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 $n \rightarrow \infty$.

Сходимость по вероятности обладает рядом естественных свойств. Например, таким: если $X_n \xrightarrow{P} X$, $Y_n \xrightarrow{P} Y$, то $X_n + Y_n \xrightarrow{P} X + Y$.

Приведем доказательство этого факта. Для любого $\varepsilon > 0$

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

Введем еще один тип сходимости.

Определение. Последовательность $\{X_n\}$ сходится к случайной величине X *почти наверное (п. н.)* (или *с вероятностью единица*), если $X_n(\omega) \rightarrow X(\omega)$ при $n \rightarrow \infty$ для всех $\omega \in \Omega \setminus N$, где $\mathbf{P}(N) = 0$.

Можно также записать: $\mathbf{P}\{\omega : X_n(\omega) \rightarrow X(\omega)\} = 1$.

Разберемся, что представляет из себя событие $\{\omega : X_n(\omega) \rightarrow X(\omega)\}$. Нетрудно видеть, что его можно представить в виде

$$\{X_n \rightarrow X\} = \bigcap_{k=1}^{\infty} \bigcup_{n=1}^{\infty} \bigcap_{m \geq n} \left\{ |X_m - X| \leq \frac{1}{k} \right\}.$$

Это означает, что для любого $k \geq 1$ существует $n \geq 1$ такое, что при всех $m \geq n$ выполняется $|X_m - X| \leq \frac{1}{k}$.

Запишем также противоположное событие:

$$\{X_n \not\rightarrow X\} = \bigcup_{k=1}^{\infty} \bigcap_{n=1}^{\infty} \bigcup_{m \geq n} \left\{ |X_m - X| > \frac{1}{k} \right\}.$$

Теорема 1. *Сходимость $X_n \rightarrow X$ п. н. эквивалентна условию: для любого $\varepsilon > 0$ $\mathbf{P}\left(\sup_{m \geq n} |X_m - X| \geq \varepsilon\right) \rightarrow 0$ при $n \rightarrow \infty$.*

Доказательство. Для начала заметим, что условие « $\mathbf{P}\left(\sup_{m \geq n} |X_m - X| \geq \varepsilon\right) \rightarrow 0$ для любого $\varepsilon > 0$ » эквивалентно условию « $\mathbf{P}\left(\sup_{m \geq n} |X_m - X| > \frac{1}{k}\right) \rightarrow 0$ для любого $k \geq 1$ ». Это следует из того, что для любого $\varepsilon > 0$ найдется число $k \geq 1$ такое, что $1/k < \varepsilon$, и наоборот: для любого $k \geq 1$ найдется $\varepsilon > 0$ такое, что $\varepsilon < 1/k$. Имеем далее

$$\begin{aligned} \mathbf{P}(X_n \rightarrow X) = 1 &\iff \mathbf{P}(X_n \not\rightarrow X) = 0 \iff \\ &\iff \mathbf{P}\left(\bigcap_{n=1}^{\infty} \bigcup_{m \geq n} \left\{ |X_m - X| > \frac{1}{k} \right\}\right) = 0 \quad \text{для любого } k \geq 1. \end{aligned}$$

Но события $\bigcup_{m \geq n} \left\{ |X_m - X| > \frac{1}{k} \right\}$ сужаются с ростом n , поэтому в силу свойства непрерывности вероятности

$$\begin{aligned} \mathbf{P}\left(\bigcap_{n=1}^{\infty} \bigcup_{m \geq n} \left\{ |X_m - X| > \frac{1}{k} \right\}\right) &= \lim_{n \rightarrow \infty} \mathbf{P}\left(\bigcup_{m \geq n} \left\{ |X_m - X| > \frac{1}{k} \right\}\right) = \\ &= \lim_{n \rightarrow \infty} \mathbf{P}\left(\sup_{m \geq n} |X_m - X| > \frac{1}{k}\right). \end{aligned}$$

Теорема доказана.

Следствие. *Если $X_n \rightarrow X$ п. н., то $X_n \xrightarrow{P} X$.*

Утверждение следует из соотношения

$$\left\{ \sup_{m \geq n} |X_m - X| \geq \varepsilon \right\} \supset \left\{ |X_n - X| \geq \varepsilon \right\}.$$

Обратное утверждение неверно. Для того, чтобы это показать, построим пример последовательности случайных величин, которая сходится по вероятности, но не сходится почти наверное. Мы модифицируем приведенный выше (см. рисунок) пример сходящейся по вероятности последовательности. По-прежнему X_n будет принимать значения 1 с вероятностью $1/n$ и 0 с вероятностью $1 - 1/n$, однако множества A_n , на которых $X_n = 1$, с ростом n будут двигаться по отрезку $[0, 1]$ так, что любая точка отрезка бесконечное число раз будет попадать в такие множества. Другими

словами, мы организуем «бегающую» ступеньку. Этого можно достичь, к примеру, если положить $A_1 = [0, 1]$, $A_2 = [0, 1/2]$, $A_3 = [1/2, 1/2 + 1/3]$, $A_4 = [5/6, 1] \cup [0, 1/12]$, $A_5 = [1/12, 1/12 + 1/5]$ и т. д. Для случайной величины X_n ступенька начинается с той точки, где закончилась ступенька предыдущей случайной величины X_{n-1} ; если же вся ступенька не уместится на интервале $[0, 1]$, то та ее часть, которая оказалась правее единицы, переносится в начало интервала.

Следующая теорема дает весьма полезное достаточное условие сходимости почти наверное.

Теорема 2. Если ряд

$$\sum_{n=1}^{\infty} \mathbf{P}(|X_n - X| \geq \varepsilon) < \infty$$

сходится при любом $\varepsilon > 0$, то $X_n \rightarrow X$ п. н.

Доказательство. Утверждение следует из соотношения

$$\mathbf{P}\left(\sup_{m \geq n} |X_m - X| \geq \varepsilon\right) \leq \mathbf{P}\left(\bigcup_{m \geq n} \left\{|X_m - X| \geq \frac{\varepsilon}{2}\right\}\right) \leq \sum_{m=n}^{\infty} \mathbf{P}\left(|X_m - X| \geq \frac{\varepsilon}{2}\right) \rightarrow 0.$$

Следствие. Если $X_n \xrightarrow{P} X$, то существует подпоследовательность n_k такая, что $X_{n_k} \rightarrow X$ п. н. при $k \rightarrow \infty$.

Доказательство. Для каждого $k \geq 1$ выберем n_k так, чтобы

$$\mathbf{P}\left(|X_{n_k} - X| \geq \frac{1}{k}\right) \leq \frac{1}{k^2},$$

и воспользуемся предыдущим утверждением. Пусть число $\varepsilon > 0$ произвольно. Начиная с некоторого номера k_0 при $k \geq k_0$ будет иметь место $1/k \leq \varepsilon$, поэтому

$$\sum_{k=k_0}^{\infty} \mathbf{P}(|X_{n_k} - X| \geq \varepsilon) \leq \sum_{k=k_0}^{\infty} \mathbf{P}\left(|X_{n_k} - X| \geq \frac{1}{k}\right) \leq \sum_{k=k_0}^{\infty} \frac{1}{k^2} < \infty.$$

Следующая теорема устанавливает связь между сходимостью в среднем квадратическом и по вероятности.

Теорема 3. Если $X_n \rightarrow X$ в среднем квадратическом, то $X_n \xrightarrow{P} X$.

Доказательство. В силу неравенства Чебышева для любого $\varepsilon > 0$

$$\mathbf{P}(|X_n - X| \geq \varepsilon) = \mathbf{P}((X_n - X)^2 \geq \varepsilon^2) \leq \frac{\mathbf{E}(X_n - X)^2}{\varepsilon^2} \rightarrow 0.$$

Обратное утверждение неверно. Чтобы подтвердить это, вернемся к уже рассмотренному выше примеру и модифицируем его. Пусть случайная величина X_n принимает теперь значение n на множестве $[0, \frac{1}{n}]$ и равна 0 в противном случае. Ясно, что как и ранее, $X_n \xrightarrow{P} 0$ и даже $X_n \rightarrow 0$ почти наверное, однако $\mathbf{E}X_n^2 = n \rightarrow 0$.

Одновременно заметим, что из сходимости в среднем квадратическом сходимости почти наверное не следует. Подтверждением служит пример, приведенный перед формулировкой теоремы 2, в котором $\mathbf{E}X_n^2 = 1^2 \cdot \frac{1}{n} + 0^2(1 - \frac{1}{n}) \rightarrow 0$, однако X_n не сходится к нулю почти наверное.

Теорема 4. Пусть $X_n \rightarrow X$ п. н. (или $X_n \xrightarrow{P} X$), g — функция, непрерывная на множестве значений X , тогда $g(X_n) \rightarrow g(X)$ п. н. ($g(X_n) \xrightarrow{P} g(X)$).

Доказательство. Для сходимости почти наверное утверждение очевидно. Предположим, что $X_n \xrightarrow{P} X$ и $g(X_n)$ не сходится по вероятности к $g(X)$. Тогда найдутся числа $\delta > 0$, $\varepsilon > 0$ и подпоследовательность индексов n' такие, что

$$\mathbf{P}(|g(X_{n'}) - g(X)| \geq \varepsilon) > \delta.$$

Но $X_{n'} \xrightarrow{P} X$, значит, найдется подпоследовательность n'' последовательности n' , для которой $X_{n''} \rightarrow X$ п.н. Тогда $g(X_{n''}) \rightarrow g(X)$ п.н. и неизбежно $g(X_{n''}) \xrightarrow{P} g(X)$, что противоречит сделанному допущению. Теорема доказана.

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. О сходимости математических ожиданий

Пусть теперь $X_n \xrightarrow{P} X$ и существуют математические ожидания $\mathbf{E}X_n$ и $\mathbf{E}X$. Можно ли утверждать в этом случае, что $\mathbf{E}X_n \rightarrow \mathbf{E}X$? Ответ отрицательный, подтверждением этому служит пример, приведенный после доказательства теоремы 3 предыдущего параграфа. Тем самым мы приходим к необходимости помимо сходимости по вероятности накладывать некоторые дополнительные условия, чтобы обеспечить сходимость математических ожиданий.

Определение. Последовательность $\{X_n\}$ случайных величин называется *равномерно интегрируемой* (РИ), если

$$\sup_n \mathbf{E}(|X_n|; |X_n| \geq N) \rightarrow 0$$

при $N \rightarrow \infty$.

Заметим, что из условия РИ следует, что $\sup_n \mathbf{E}|X_n| < \infty$. Действительно, выбирая число $N > 0$ таким, что $\sup_n \mathbf{E}(|X_n|; |X_n| \geq N) < 1$, будем иметь

$$\begin{aligned} \sup_n \mathbf{E}|X_n| &= \sup_n [\mathbf{E}(|X_n|; |X_n| \geq N) + \mathbf{E}(|X_n|; |X_n| < N)] \leq \\ &\leq \sup_n \mathbf{E}(|X_n|; |X_n| \geq N) + \sup_n \mathbf{E}(|X_n|; |X_n| < N) < 1 + N. \end{aligned}$$

Очевидно, что последовательность одинаково распределенных случайных величин с конечным средним равномерно интегрируема.

Если X_n равномерно интегрируемы, то CX_n , $X_n + C$, где C — константа, также равномерно интегрируемы. Это легко проверяется.

Более того, если последовательности $\{X_n\}$ и $\{Y_n\}$ равномерно интегрируемы, то последовательности $Z_n = \max(|X_n|, |Y_n|)$, $Z_n = X_n + Y_n$, также равномерно интегрируемы.

Действительно, для $Z_n = \max(|X_n|, |Y_n|)$ имеем

$$\begin{aligned} \mathbf{E}(Z_n; Z_n > N) &= \mathbf{E}(Z_n; Z_n > N, |X_n| > |Y_n|) + \mathbf{E}(Z_n; Z_n > N, |X_n| \leq |Y_n|) \leq \\ &\leq \mathbf{E}(|X_n|; |X_n| > N) + \mathbf{E}(|Y_n|; |Y_n| \geq N) \rightarrow 0 \end{aligned}$$

при $N \rightarrow \infty$ равномерно по n .

Так как

$$|X_n + Y_n| \leq |X_n| + |Y_n| \leq 2 \max(|X_n|, |Y_n|),$$

то из сказанного выше следует также равномерная интегрируемость суммы $Z_n = X_n + Y_n$.

Теорема 5. Пусть $X_n \xrightarrow{P} X$ и последовательность $\{X_n\}$ удовлетворяет условию РИ. Тогда существует $\mathbf{E}|X|$ и $\mathbf{E}|X_n - X| \rightarrow 0$ при $n \rightarrow \infty$.

Разумеется, отсюда следует $\mathbf{E}X_n \rightarrow \mathbf{E}X$, поскольку $|\mathbf{E}X_n - \mathbf{E}X| \leq \mathbf{E}|X_n - X|$.

Доказательство. Сначала покажем, что существует $\mathbf{E}|X|$. Для любых $N > 0$, $\varepsilon > 0$

$$\begin{aligned} \mathbf{E} \min(|X|, N) &= \lim_{n \rightarrow \infty} \mathbf{E}(\min(|X|, N); |X_n - X| < \varepsilon) \leq \\ &\leq \overline{\lim}_{n \rightarrow \infty} \mathbf{E} \min(|X_n| + \varepsilon, N) \leq \overline{\lim}_{n \rightarrow \infty} \mathbf{E}(|X_n| + \varepsilon) \leq C + \varepsilon. \end{aligned}$$

При каждом N случайную величину $\min(|X|, N)$ можно приблизить снизу простыми случайными величинами, математические ожидания которых также не будут превосходить $C + \varepsilon$. Будем теперь неограниченно увеличивать значения N . Тогда у аппроксимирующих простых случайных величин будут появляться все новые ступеньки, и эти простые случайные величины будут в итоге приближать случайную величину $|X|$. Поскольку математические ожидания всех рассматриваемых простых случайных величин не превосходят $C + \varepsilon$, то же самое будет верно и для $\mathbf{E}|X|$.

Положим далее $Y_n = |X_n - X|$. Тогда $Y_n \xrightarrow{P} 0$ и последовательность $\{Y_n\}$ также удовлетворяет условию РИ. Поэтому для любых $N > \varepsilon > 0$ имеем

$$\begin{aligned} \mathbf{E}Y_n &= \mathbf{E}(Y_n; Y_n < \varepsilon) + \mathbf{E}(Y_n; \varepsilon \leq Y_n < N) + \mathbf{E}(Y_n; Y_n \geq N) \leq \\ &\leq \varepsilon + N\mathbf{P}(Y_n \geq \varepsilon) + \mathbf{E}(Y_n; Y_n \geq N). \end{aligned}$$

В силу условия РИ число N можно выбрать таким, чтобы $\mathbf{E}(Y_n; Y_n \geq N) \leq \varepsilon$. Кроме того, $\mathbf{P}(Y_n \geq \varepsilon) \rightarrow 0$. Поэтому

$$\overline{\lim}_{n \rightarrow \infty} \mathbf{E}Y_n \leq 2\varepsilon,$$

откуда заключаем, что $\lim_{n \rightarrow \infty} \mathbf{E}Y_n = 0$, так как ε произвольно. Теорема доказана.

Следствие 1 (теорема о мажорируемой сходимости). Пусть $X_n \xrightarrow{P} X$ и $|X_n| \leq Y$, где $\mathbf{E}Y < \infty$. Тогда $\mathbf{E}|X_n - X| \rightarrow 0$.

Доказательство. Установим выполнение условия РИ. При $N \rightarrow \infty$

$$\mathbf{E}(|X_n|; |X_n| \geq N) \leq \mathbf{E}(Y; |X_n| \geq N) \leq \mathbf{E}(Y; Y \geq N) \rightarrow 0.$$

Следствие 2. Пусть $X_n \xrightarrow{P} X$ и $\mathbf{E}|X_n|^{1+\alpha} \leq C < \infty$ при некотором $\alpha > 0$. Тогда $\mathbf{E}|X_n - X| \rightarrow 0$.

Доказательство. Проверим также выполнение условия РИ. При $N \rightarrow \infty$

$$\mathbf{E}(|X_n|; |X_n| \geq N) \leq \mathbf{E}(|X_n|^{1+\alpha}/N^\alpha; |X_n| \geq N) \leq \frac{1}{N^\alpha} \mathbf{E}|X_n|^{1+\alpha} \rightarrow 0.$$

4.3. Законы 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 1 (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 $n \rightarrow \infty$. Теорема доказана.

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. Мы сделаем это позже, изучив другой подход к доказательству.

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.

Оказывается, при тех же условиях можно доказать и более сильный результат — но ценой гораздо больших усилий.

Теорема 2 (усиленный закон больших чисел). *В условиях предыдущей теоремы*

$$\frac{S_n}{n} \rightarrow a \text{ п. н.}$$

Доказательство. Не ограничивая общности можно считать, что $a = 0$, иначе можно перейти к случайным величинам $Y_n = X_n - a$.

Для любого $n \geq 1$ найдется целое число m такое, что $m^2 \leq n < (m+1)^2$. Ясно, что $n \rightarrow \infty$ и $m \rightarrow \infty$ одновременно. Поэтому

$$\left| \frac{S_n}{n} \right| \leq \left| \frac{S_{m^2}}{m^2} \right| + \frac{Y_m}{m^2},$$

где $Y_m = \max_{m^2+1 \leq n < (m+1)^2} |X_{m^2+1} + \dots + X_n|$. Покажем, что с вероятностью единица

$$\left| \frac{S_{m^2}}{m^2} \right| \rightarrow 0, \quad \frac{Y_m}{m^2} \rightarrow 0$$

при $m \rightarrow \infty$. Применив второе неравенство Чебышева, для произвольного $\varepsilon > 0$ получаем

$$\mathbf{P} \left(\left| \frac{S_{m^2}}{m^2} \right| \geq \varepsilon \right) \leq \frac{\mathbf{D}S_{m^2}}{m^4\varepsilon^2} = \frac{m^2\sigma^2}{m^4\varepsilon^2} = \frac{\sigma^2}{\varepsilon^2} \frac{1}{m^2},$$

поэтому

$$\sum_{m=1}^{\infty} \mathbf{P} \left(\left| \frac{S_{m^2}}{m^2} \right| \geq \varepsilon \right) \leq \frac{\sigma^2}{\varepsilon^2} \sum_{m=1}^{\infty} \frac{1}{m^2} < \infty,$$

т. е. $\frac{S_{m^2}}{m^2} \rightarrow 0$ п. н. при $m \rightarrow \infty$. Далее,

$$\begin{aligned} \mathbf{P} \left(\left| \frac{Y_m}{m^2} \right| \geq \varepsilon \right) &= \mathbf{P} \left\{ \bigcup_{n=m^2+1}^{(m+1)^2-1} \left(\left| \frac{X_{m^2+1} + \dots + X_n}{m^2} \right| \geq \varepsilon \right) \right\} \leq \\ &\leq \sum_{n=m^2+1}^{(m+1)^2-1} \mathbf{P} \left\{ \left(\left| \frac{X_{m^2+1} + \dots + X_n}{m^2} \right| \geq \varepsilon \right) \right\} \leq \\ &\leq \sum_{n=m^2+1}^{(m+1)^2-1} \frac{(n-m^2)\sigma^2}{m^4\varepsilon^2} \leq 2m \frac{2m\sigma^2}{m^4\varepsilon^2}, \\ \sum_{m=1}^{\infty} \mathbf{P} \left(\left| \frac{Y_m}{m^2} \right| \geq \varepsilon \right) &\leq \frac{4\sigma^2}{\varepsilon^2} \sum_{m=1}^{\infty} \frac{1}{m^2} < \infty. \end{aligned}$$

Следовательно, $\frac{Y_m}{m^2} \rightarrow 0$ п. н. Теорема доказана.

Замечание. Усиленный закон больших чисел также справедлив только лишь при наличии первого момента у X_1 , однако доказательство этого факта выходит за рамки нашего курса.

4.4. Слабая сходимость распределений

Пусть имеется некоторая последовательность функций распределения $\{F_n(y)\}$ и еще одна функция распределения $F(y)$.

Определение. F_n слабо сходится к F при $n \rightarrow \infty$ (обозначается $F_n \Rightarrow F$), если для любой непрерывной и ограниченной функции g

$$\int_{-\infty}^{\infty} g(y) dF_n(y) \rightarrow \int_{-\infty}^{\infty} g(y) dF(y).$$

Напомним, что это то же самое, что и

$$\int_{-\infty}^{\infty} g(y) P_n(dy) \rightarrow \int_{-\infty}^{\infty} g(y) P(dy).$$

Можно говорить просто о слабой сходимости распределений. Данное определение можно также записать в виде

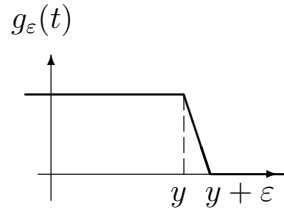
$$\mathbf{E}g(X_n) \rightarrow \mathbf{E}g(X),$$

где $X_n \in F_n$, $X \in F$.

Теорема (критерий слабой сходимости). $F_n \Rightarrow F$ тогда и только тогда, когда $F_n(y) \rightarrow F(y)$ для каждой точки y , в которой F непрерывна.

Заметим, что любая функция распределения F может иметь не более счетного множества разрывов. Действительно, можно перенумеровать все скачки функции распределения, выделив сначала скачки, размер которых превышает $1/2$, затем скачки, превышающие $1/3$, и т. д.

Доказательство. Пусть $F_n \Rightarrow F$ в смысле приведенного выше определения. Для фиксированного y и $\varepsilon > 0$ построим непрерывную ограниченную функцию $g_\varepsilon(t)$ следующим образом. Она равна единице при $t < y$, нулю при $t \geq y + \varepsilon$, и линейна на отрезке $[y, y + \varepsilon]$:



Так как

$$F_n(y) = \int_{-\infty}^y g_\varepsilon(t) dF_n(t) \leq \int_{-\infty}^{\infty} g_\varepsilon(t) dF_n(t),$$

то

$$\limsup_{n \rightarrow \infty} F_n(y) \leq \limsup_{n \rightarrow \infty} \int_{-\infty}^{\infty} g_\varepsilon(t) dF_n(t) = \int_{-\infty}^{\infty} g_\varepsilon(t) dF(t) \leq \int_{-\infty}^{y+\varepsilon} dF(t) = F(y + \varepsilon).$$

Устремим $\varepsilon \rightarrow 0$. Если y — точка непрерывности F , то

$$\limsup_{n \rightarrow \infty} F_n(y) \leq F(y).$$

Совершенно аналогично, взяв непрерывную функцию

$$h_\varepsilon(t) = g_\varepsilon(t + \varepsilon) = \begin{cases} 1, & t < y - \varepsilon, \\ 0, & t \geq y, \\ \text{линейна,} & t \in [y - \varepsilon, y], \end{cases}$$

получим неравенства

$$F_n(y) \geq \int_{-\infty}^{\infty} h_\varepsilon(t) dF_n(t),$$

$$\liminf_{n \rightarrow \infty} F_n(y) \geq F(y).$$

Утверждение в одну сторону доказано. Обратно, пусть g — непрерывная ограниченная функция, $|g(y)| \leq C$, и пусть $-M$ и N — точки непрерывности F такие, что

$$F(-M) < \frac{\varepsilon}{5C}, \quad 1 - F(N) < \frac{\varepsilon}{5C}.$$

Тогда

$$F_n(-M) < \frac{\varepsilon}{4C}, \quad 1 - F_n(N) < \frac{\varepsilon}{4C}$$

при всех достаточно больших n . Поэтому

$$\left| \int_{-\infty}^{\infty} g(y) dF_n(y) - \int_{-M}^N g(y) dF_n(y) \right| \leq C \left(\frac{\varepsilon}{4C} + \frac{\varepsilon}{4C} \right) = \frac{\varepsilon}{2},$$

$$\left| \int_{-\infty}^{\infty} g(y) dF(y) - \int_{-M}^N g(y) dF(y) \right| \leq C \left(\frac{\varepsilon}{5C} + \frac{\varepsilon}{5C} \right) = \frac{2\varepsilon}{5} < \frac{\varepsilon}{2}.$$

Далее будем оперировать с интегралами по множеству $[-M, N)$. Построим на $[-M, N)$ ступенчатую функцию g_ε , которая отличалась бы от g менее чем на $\varepsilon/2$ и имела скачки лишь в точках непрерывности F . Вне $[-M, N)$ полагаем g_ε равной нулю. Пусть, к примеру, $g_\varepsilon(y) = g(y_i)$ на интервале $[y_i, y_{i+1})$, где $y_0 = -M, \dots, y_k = N$, y_i — точки непрерывности F . Тогда

$$\left| \int_{-M}^N g_\varepsilon(y) dF_n(y) - \int_{-M}^N g(y) dF_n(y) \right| \leq \frac{\varepsilon}{2},$$

$$\left| \int_{-M}^N g_\varepsilon(y) dF(y) - \int_{-M}^N g(y) dF(y) \right| \leq \frac{\varepsilon}{2}.$$

Следовательно, для достаточно больших n интегралы $\int_{-\infty}^{\infty} g(y) dF_n(y)$ и $\int_{-\infty}^{\infty} g(y) dF(y)$

отличаются от $\int_{-M}^N g_\varepsilon(y) dF_n(y)$ и $\int_{-M}^N g_\varepsilon(y) dF(y)$ соответственно менее, чем на ε . Но

$$\begin{aligned} \int_{-M}^N g_\varepsilon(y) dF_n(y) &= \sum_{i=0}^{k-1} g(y_i) [F_n(y_{i+1}) - F_n(y_i)] \rightarrow \sum_{i=0}^{k-1} g(y_i) [F(y_{i+1}) - F(y_i)] \\ &= \int_{-M}^N g_\varepsilon(y) dF(y). \end{aligned}$$

Поэтому

$$\begin{aligned} \limsup_{n \rightarrow \infty} \int_{-\infty}^{\infty} g(y) dF_n(y) &\leq \varepsilon + \limsup_{n \rightarrow \infty} \int_{-M}^N g_\varepsilon(y) dF_n(y) = \varepsilon + \int_{-M}^N g_\varepsilon(y) dF(y) \\ &\leq 2\varepsilon + \int_{-\infty}^{\infty} g(y) dF(y), \end{aligned}$$

и аналогичное неравенство верно для нижнего предела. Теорема доказана.

Замечания

1. Если F непрерывна, то слабая сходимость $F_n \Rightarrow F$ эквивалентна сходимости $F_n(y) \rightarrow F(y)$ равномерно по всем y , т. е. $\sup_y |F_n(y) - F(y)| \rightarrow 0$.

Доказательство. Пусть точки $y_0 = -\infty, y_1, \dots, y_k = \infty$ таковы, что

$$F(y_i) - F(y_{i-1}) \leq \varepsilon/2.$$

Тогда найдется n_0 такое, что при $n \geq n_0$

$$|F_n(y_i) - F(y_i)| \leq \varepsilon/2 \quad \text{для всех } i = 0, \dots, k.$$

Заметим при этом, что $F_n(\pm\infty) = F(\pm\infty)$. Пусть теперь y произвольно. Для него найдется число i такое, что $y \in [y_{i-1}, y_i]$. Тогда

$$F_n(y) - F(y) \leq F_n(y_i) - F(y_{i-1}) \leq F(y_i) + \frac{\varepsilon}{2} - F(y_{i-1}) \leq \varepsilon,$$

$$F_n(y) - F(y) \geq F_n(y_{i-1}) - F(y_i) \geq F(y_{i-1}) - \frac{\varepsilon}{2} - F(y_i) \geq -\varepsilon,$$

откуда следует наше утверждение.

2. Если F_n и F дискретны и имеют скачки в одних и тех же точках y_1, y_2, \dots , то слабая сходимость $F_n \Rightarrow F$ эквивалентна сходимости $F_n(y_i + 0) - F_n(y_i) \rightarrow F(y_i + 0) - F(y_i)$.

3. Слабая сходимость $F_n \Rightarrow F$ ровным счетом ничего не говорит о сходимости случайных величин $X_n \in F_n$, они вообще могут быть заданы на разных вероятностных пространствах. Даже если они заданы на одном пространстве, то следующий пример показывает, что ни о какой сходимости случайных величин не может быть речи.

Пусть при однократном бросании монеты X_1 равно числу выпавших гербов (X_1 принимает только значения 1 и 0 с равными вероятностями), а X_2 равно числу выпавших решек. Ясно, что последовательность $X_1, X_2, X_1, X_2, \dots$ ни к чему не сходится, хотя все члены этой последовательности имеют одно и то же распределение.

В качестве примера слабо сходящейся последовательности функций распределения можно рассмотреть $F_n(y) = U_{-\frac{1}{n}, 0}(y)$. Здесь $F_n(y) \rightarrow I_0(y)$ для всех y , кроме точки разрыва $y = 0$.

Теорема. Если $X_n \xrightarrow{P} X$, то $F_{X_n} \Rightarrow F_X$. Обратное неверно. Однако если $F_X = I_a$ — вырожденное распределение, то из $F_{X_n} \Rightarrow I_a$ следует $X_n \xrightarrow{P} a$.

Доказательство. Пусть g — непрерывная ограниченная функция, тогда $g(X_n) \xrightarrow{P} g(X)$, и в силу теоремы о мажорируемой сходимости $\mathbf{E}g(X_n) \rightarrow \mathbf{E}g(X)$, что доказывает первую часть теоремы. Далее, если $F_{X_n} \Rightarrow I_a$, то

$$\begin{aligned} \mathbf{P}(|X_n - a| < \varepsilon) &= \mathbf{P}(a - \varepsilon < X_n < a + \varepsilon) \geq \mathbf{P}(a - \frac{\varepsilon}{2} \leq X_n < a + \varepsilon) = \\ &= F_{X_n}(a + \varepsilon) - F_{X_n}(a - \frac{\varepsilon}{2}) \rightarrow 1, \end{aligned}$$

поскольку $a + \varepsilon$ и $a - \frac{\varepsilon}{2}$ являются точками непрерывности функции распределения F_a . Теорема доказана.

Таким образом, закон больших чисел также является одним из примеров слабой сходимости распределений $F_{S_n/n} \Rightarrow F_a$.

4.5. Характеристические функции

Наряду с вещественнозначными случайными величинами можно рассматривать и комплекснозначные вида $X_1 + iX_2$. По определению полагаем

$$\mathbf{E}(X_1 + iX_2) = \mathbf{E}X_1 + i\mathbf{E}X_2.$$

Независимость случайных величин $X_1 + iX_2$ и $Y_1 + iY_2$ можно понимать как независимость σ -алгебр $\sigma(X_1, X_2)$ и $\sigma(Y_1, Y_2)$, порожденных случайными векторами (X_1, X_2) и (Y_1, Y_2) . Нетрудно проверить также для комплекснозначных случайных величин выполнение всех основных свойств математических ожиданий (кроме, разумеется, тех, где сравниваются величины без модуля), включая свойство $\mathbf{E}(XY) = \mathbf{E}X\mathbf{E}Y$ для независимых случайных величин.

Определение. *Характеристической функцией* (х. ф.) случайной величины X называется

$$\varphi_X(t) = \mathbf{E}e^{itX} = \int_{-\infty}^{\infty} e^{ity} dF_X(y) = \int_{-\infty}^{\infty} \cos ty dF_X(y) + i \int_{-\infty}^{\infty} \sin ty dF_X(y), \quad -\infty < t < \infty.$$

Если F_X абсолютно непрерывна, то х. ф. есть преобразование Фурье плотности распределения. В соответствии с правилом вычисления математических ожиданий функций от случайных величин, имеем

$$\varphi_X(t) = \begin{cases} \sum_{k=1}^{\infty} e^{ity_k} \mathbf{P}(X = y_k) & \text{(для дискретных распределений),} \\ \int_{-\infty}^{\infty} e^{ity} f_X(y) dy & \text{(для распределений с плотностью).} \end{cases}$$

Х.ф. всегда существует, так как

$$|\varphi_X(t)| = \left| \int_{-\infty}^{\infty} e^{ity} dF_X(y) \right| \leq \int_{-\infty}^{\infty} |e^{ity}| dF_X(y) = 1.$$

Свойства характеристических функций.

1. $\varphi_X(0) = 1$, $|\varphi_X(t)| \leq 1$.

2. $\varphi_{aX+b}(t) = e^{itb} \varphi_X(at)$, $\overline{\varphi_X(t)} = \varphi_X(-t) = \varphi_{-X}(t)$.

3. Если случайные величины X_1, \dots, X_n независимы, то

$$\varphi_{X_1+\dots+X_n}(t) = \varphi_{X_1}(t) \dots \varphi_{X_n}(t).$$

4. Если существует $\mathbf{E}|X|^k < \infty$, $k \geq 1$, то существует непрерывная k -я производная $\varphi_X^{(k)}(t)$ и $\varphi_X^{(k)}(0) = i^k \mathbf{E}X^k$.

Доказательство. В силу того, что равномерно по t выполняется

$$\left| \int_{-\infty}^{\infty} iy e^{ity} dF_X(y) \right| \leq \int_{-\infty}^{\infty} |y| dF_X(y) < \infty,$$

можно дифференцировать под знаком интеграла:

$$\varphi'_X(t) = \left(\int_{-\infty}^{\infty} e^{ity} dF_X(y) \right)' = \int_{-\infty}^{\infty} iy e^{ity} dF_X(y).$$

Дальше действуем по индукции. Пусть при $l < k$

$$\varphi_X^{(l)}(t) = \int_{-\infty}^{\infty} (iy)^l e^{ity} dF_X(y),$$

тогда

$$\varphi_X^{(l+1)}(t) = \int_{-\infty}^{\infty} (iy)^{l+1} e^{ity} dF_X(y)$$

в силу равномерной по t сходимости интеграла в правой части. Итак,

$$\varphi_X^{(k)}(t) = \int_{-\infty}^{\infty} (iy)^k e^{ity} dF_X(y),$$

и остается положить $t = 0$ в этом выражении.

Установим непрерывность производной.

$$|\varphi_X^{(k)}(t+h) - \varphi_X^{(k)}(t)| = \left| i^k \int_{-\infty}^{\infty} y^k (e^{i(t+h)y} - e^{ity}) dF_X(y) \right| \leq \int_{-\infty}^{\infty} |y|^k |e^{ihy} - 1| dF_X(y) \rightarrow 0$$

в силу теоремы о мажорируемой сходимости, т. к. $e^{iXh} - 1 \rightarrow 0$ при $h \rightarrow 0$, и, кроме того,

$$|X|^k |e^{ihX} - 1| \leq 2|X|^k, \quad \mathbf{E}|X|^k < \infty.$$

Случай $k = 0$ не исключается, т. е. мы установили кроме всего прочего, что х. ф. является равномерно непрерывной функцией.

5. Если $\mathbf{E}|X|^k < \infty$, то в окрестности точки $t = 0$

$$\varphi_X(t) = 1 + \sum_{j=1}^k \frac{(it)^j}{j!} \mathbf{E}X^j + o(|t|^k).$$

Примеры.

1. Если $X \in I_a$, то $\varphi_a(t) = e^{iat}$.

2. Если $X \in \Pi_\lambda$, то

$$\varphi_X(t) = \sum_{k=0}^{\infty} \frac{e^{itk} \lambda^k}{k!} e^{-\lambda} = \exp\{\lambda(e^{it} - 1)\}.$$

3. Пусть $X \in \Phi_{\alpha, \sigma^2}$. Воспользуемся тем, что $X = \sigma Y + \alpha$, где $Y \in \Phi_{0,1}$. Значит, $\varphi_X(t) = e^{i\alpha t} \varphi_Y(\sigma t)$. Найдем $\varphi_Y(t)$. Продифференцировав эту функцию по t , получим

$$\begin{aligned} \varphi'_Y(t) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} iye^{ity - \frac{y^2}{2}} dy = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} ie^{ity} d(-e^{-\frac{y^2}{2}}) = \\ &= \frac{1}{\sqrt{2\pi}} \left[-ie^{ity - \frac{y^2}{2}} \Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} te^{ity - \frac{y^2}{2}} dy \right] = -t\varphi_Y(t), \end{aligned}$$

т. е. $[\ln \varphi_Y(t)]' = -t$. Отсюда получаем

$$\ln \varphi_Y(t) = -\frac{t^2}{2} + C.$$

Поскольку $\varphi_Y(0) = 1$, то $C = 0$,

$$\varphi_Y(t) = e^{-\frac{t^2}{2}}, \quad \varphi_X(t) = e^{i\alpha t - \frac{\sigma^2 t^2}{2}}.$$

Пользуясь этой формулой, удобно находить моменты случайной величины X .

Далее мы установим взаимно однозначное соответствие между функциями распределения и характеристическими функциями.

Теорема (формула обращения). *Если F — функция распределения, а φ — соответствующая ей х. ф., то для любых точек непрерывности x и y функции F*

$$F(y) - F(x) = \frac{1}{2\pi} \lim_{\sigma \rightarrow 0} \int_{-\infty}^{\infty} \frac{e^{-itx} - e^{-ity}}{it} \varphi(t) e^{-t^2 \sigma^2} dt.$$

Замечание. Если функция $\varphi(t)/t$ интегрируема на бесконечности, то становится законным предельный переход под знаком интеграла в этой формуле, и можно записать

$$F(y) - F(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{-itx} - e^{-ity}}{it} \varphi(t) dt.$$

Доказательство теоремы. Предположим сначала, что существует плотность $f(y) = F'(y)$, а х.ф. $\varphi(t)$ интегрируема. Тогда можно воспользоваться формулой для обратного преобразования Фурье

$$f(u) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itu} \varphi(t) dt$$

и

$$\begin{aligned} F(y) - F(x) &= \int_x^y f(u) du = \frac{1}{2\pi} \int_x^y \int_{-\infty}^{\infty} e^{-itu} \varphi(t) dt du = \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_x^y e^{-itu} du \varphi(t) dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{-itx} - e^{-ity}}{it} \varphi(t) dt. \end{aligned}$$

Перемена порядка интегрирования здесь возможна в силу интегрируемости функции $\varphi(t)$.

Пусть теперь F — произвольная функция распределения, и пусть (X, Y) — случайный вектор, имеющий совместную функцию распределения $F_{X,Y}(u, v) = F(u)\Phi_{0,1}(v)$, т.е. X и Y независимы, $X \in F$, $Y \in \Phi_{0,1}$. Введем случайную величину

$$Z_k = \sigma_k \sqrt{2}Y \in \Phi_{0,2\sigma_k^2}.$$

Если $\sigma_k \rightarrow 0$, то почти наверное $Z_k \rightarrow 0$ и $X + Z_k \rightarrow X$. Следовательно, $F_{X+Z_k} \Rightarrow F_X$. С другой стороны, случайная величина $X + Z_k$ будет обладать плотностью и ее х. ф., равная $\varphi(t)e^{-t^2\sigma_k^2}$, интегрируема. Поэтому

$$F_{X+Z_k}(y) - F_{X+Z_k}(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{-itx} - e^{-ity}}{it} \varphi(t)e^{-t^2\sigma_k^2} dt.$$

Если x и y — точки непрерывности F , то $F(y) - F(x) = \lim_{\sigma_k \rightarrow 0} (F_{X+Z_k}(y) - F_{X+Z_k}(x))$. Отсюда следует утверждение теоремы.

Следствие 1. *Х. ф. случайной величины однозначно определяет ее функцию распределения.*

Доказательство следует из формулы обращения и из того, что разности $F(y) - F(x)$ однозначно определяют F . Достаточно для этого в формуле обращения устремить $x \rightarrow -\infty$.

Следствие 2. *Пусть $X_1 \in \Pi_{\lambda_1}$, $X_2 \in \Pi_{\lambda_2}$, X_1 и X_2 независимы, тогда $X_1 + X_2 \in \Pi_{\lambda_1 + \lambda_2}$.*

Этот факт уже доказывался ранее прямыми вычислениями. Теперь мы приводим новое доказательство:

$$\varphi_{X_1+X_2}(t) = \exp\{\lambda_1(e^{it} - 1)\} \exp\{\lambda_2(e^{it} - 1)\} = \exp\{(\lambda_1 + \lambda_2)(e^{it} - 1)\},$$

что соответствует распределению $\Pi_{\lambda_1 + \lambda_2}$.

Следствие 3. *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.

$$\varphi_{X_1+X_2}(t) = e^{i\alpha_1 t - \frac{\sigma_1^2 t^2}{2}} e^{i\alpha_2 t - \frac{\sigma_2^2 t^2}{2}} = e^{i(\alpha_1 + \alpha_2)t - \frac{(\sigma_1^2 + \sigma_2^2)t^2}{2}},$$

что соответствует распределению $\Phi_{\alpha_1 + \alpha_2, \sigma_1^2 + \sigma_2^2}$.

Если X — целочисленная случайная величина, то удобнее работать с производящей функцией, которая получается из х. ф. заменой $z = e^{it}$:

$$\psi_X(z) = \mathbf{E}z^X = \sum_k z^k \mathbf{P}(X = k), \quad |z| = 1.$$

Очевидно, производящая функция представляет собой ряд Лорана. Для нахождения его коэффициентов можно воспользоваться известной формулой

$$\mathbf{P}(X = k) = \frac{1}{2\pi i} \int_{|z|=1} \psi_X(z) z^{-k-1} dz,$$

которая также может восприниматься как вариант формулы обращения.

Теорема о непрерывном соответствии. *Пусть $\{F_n\}$ — последовательность функций распределения, а $\{\varphi_n\}$ — соответствующая ей последовательность х. ф.*

Тогда для сходимости $F_n \Rightarrow F$, где F — функция распределения, необходимо и достаточно, чтобы $\varphi_n(t) \rightarrow \varphi(t)$ при каждом t , где φ — х. ф., соответствующая распределению F .

Доказательство. Необходимость устанавливается просто: при каждом t функции $\cos ty$ и $\sin ty$ непрерывны и ограничены, поэтому

$$\int_{-\infty}^{\infty} \cos ty dF_n(y) \rightarrow \int_{-\infty}^{\infty} \cos ty dF(y),$$

$$\int_{-\infty}^{\infty} \sin ty dF_n(y) \rightarrow \int_{-\infty}^{\infty} \sin ty dF(y),$$

и, следовательно,

$$\varphi_n(t) = \int_{-\infty}^{\infty} \cos ty dF_n(y) + i \int_{-\infty}^{\infty} \sin ty dF_n(y) \rightarrow \int_{-\infty}^{\infty} \cos ty dF(y) + i \int_{-\infty}^{\infty} \sin ty dF(y) = \varphi(t).$$

Доказательство достаточности проведем в несколько этапов. Обозначим C^0 множество непрерывных финитных (т. е. обращающихся в 0 вне некоторого отрезка) функций и будем писать просто \int , если интегрирование проводится по всей вещественной оси.

Лемма 1. Пусть $\int g dF_n \rightarrow \int g dF$ для любой функции $g \in C^0$. Тогда $F_n \Rightarrow F$.

Доказательство. Для произвольного $\varepsilon > 0$ выберем $g_\varepsilon \in C^0$ так, чтобы

$$0 \leq g_\varepsilon(y) \leq 1, \quad \int (1 - g_\varepsilon(y)) dF(y) \leq \varepsilon.$$

Это всегда можно сделать, положив $g_\varepsilon(y) = 1$ при $y \in [-N, N]$ и $0 \leq g_\varepsilon(y) \leq 1$ при $|y| > N$ (с сохранением финитности), где число N настолько велико, что $F(-N) + 1 - F(N) < \varepsilon$.

Возьмем теперь произвольную непрерывную ограниченную функцию g , $|g(y)| \leq K < \infty$. Тогда $gg_\varepsilon \in C^0$ и

$$\left| \int (1 - g_\varepsilon)g dF_n \right| \leq K \left(1 - \int g_\varepsilon dF_n \right) \rightarrow K \left(1 - \int g_\varepsilon dF \right) \leq K\varepsilon.$$

Поэтому

$$\begin{aligned} & \limsup \left| \int g dF_n - \int g dF \right| \leq \\ & \leq \limsup \left\{ \left| \int (1 - g_\varepsilon)g dF_n \right| + \left| \int gg_\varepsilon dF_n - \int gg_\varepsilon dF \right| + \left| \int (1 - g_\varepsilon)g dF \right| \right\} \leq 2K\varepsilon. \end{aligned}$$

Лемма доказана.

Обозначим C_k^0 множество всех финитных k раз непрерывно дифференцируемых функций, $k \geq 1$.

Лемма 2. Пусть $\int g dF_n \rightarrow \int g dF$ для любой функции $g \in C_k^0$ при некотором $k \geq 1$. Тогда $F_n \Rightarrow F$.

Доказательство. Любую функцию $g \in C^0$ можно равномерно приблизить функцией $\tilde{g} \in C_k^0$. К примеру, можно воспользоваться следующим процессом. Поскольку

для любого $\varepsilon > 0$ найдется $\delta = \delta(\varepsilon) > 0$ такое, что $|g(u) - g(y)| < \varepsilon$ при $|u - y| < \delta$ (свойство равномерной непрерывности g), то положив

$$\tilde{g}(y) = \frac{1}{2\delta} \int_{y-\delta}^{y+\delta} g(u) du,$$

будем иметь

$$|\tilde{g}(y) - g(y)| \leq \frac{1}{2\delta} \int_{y-\delta}^{y+\delta} |g(u) - g(y)| du \leq \varepsilon.$$

Функция \tilde{g} уже будет непрерывно дифференцируемой и по-прежнему финитной. Повторив эту операцию дважды, получим дважды непрерывно дифференцируемую финитную функцию и т. д. Итак, пусть $g \in C^0$, $\tilde{g} \in C_k^0$ такие, что $|\tilde{g}(y) - g(y)| \leq \frac{\varepsilon}{3}$, тогда для всех достаточно больших n

$$\left| \int g dF_n - \int g dF \right| \leq \left| \int (g - \tilde{g}) dF_n \right| + \left| \int \tilde{g} dF_n - \int \tilde{g} dF \right| + \left| \int (g - \tilde{g}) dF \right| \leq \varepsilon.$$

Лемма доказана.

Лемма 3 (равенство Парсеваля). Пусть $g \in C_k^0$ при некотором $k \geq 2$, $X \in F$, тогда

$$\mathbf{E}g(X) = \int g(y) dF(y) = \frac{1}{2\pi} \int \varphi(-t) \hat{g}(t) dt,$$

где $\varphi(t) = \mathbf{E}e^{itX}$, $\hat{g}(t) = \int e^{ity} g(y) dy$.

Если F имеет плотность $f = F'$, то мы получаем равенство Парсеваля в более привычном виде

$$\int g(y) f(y) dy = \frac{1}{2\pi} \int \varphi(-t) \hat{g}(t) dt.$$

Доказательство. Для функции $g \in C_k^0$ имеем

$$\begin{aligned} \hat{g}(t) &= \int_{-\infty}^{\infty} e^{ity} g(y) dy = \int_{-\infty}^{\infty} g(y) d\left(\frac{e^{ity}}{it}\right) = \left[\frac{g(y)e^{ity}}{it} \Big|_{-\infty}^{\infty} - \frac{1}{it} \int_{-\infty}^{\infty} e^{ity} g'(y) dy \right] = \\ &= -\frac{1}{it} \int_{-\infty}^{\infty} e^{ity} g'(y) dy = \dots = \frac{(-1)^k}{(it)^k} \int_{-\infty}^{\infty} e^{ity} g^{(k)}(y) dy. \end{aligned}$$

Отсюда

$$|\hat{g}(t)| \leq \frac{C}{|t|^k}, \quad k \geq 2,$$

поэтому $\hat{g}(t)$ — абсолютно интегрируемая на бесконечности функция. Воспользуемся формулой обратного преобразования Фурье:

$$g(y) = \frac{1}{2\pi} \int e^{-ity} \hat{g}(t) dt,$$

откуда

$$\mathbf{E}g(X) = \mathbf{E} \frac{1}{2\pi} \int e^{-itX} \hat{g}(t) dt = \frac{1}{2\pi} \int \mathbf{E} e^{-itX} \hat{g}(t) dt = \frac{1}{2\pi} \int \varphi(-t) \hat{g}(t) dt.$$

Лемма доказана.

Перейдем теперь непосредственно к доказательству теоремы о непрерывном соответствии. Пусть $\varphi_n(t) \rightarrow \varphi(t)$ при всех t . Тогда для всякой $g \in C_k^0$ при $k \geq 2$ имеет место

$$\int g(y) dF_n(y) = \frac{1}{2\pi} \int \varphi_n(-t) \hat{g}(t) dt,$$

и функция $|\varphi_n(-t) \hat{g}(t)| \leq |\hat{g}(t)|$ интегрируема. По теореме о мажорируемой сходимости правая часть сходится к

$$\frac{1}{2\pi} \int \varphi(-t) \hat{g}(t) dt = \int g(y) dF(y).$$

Теорема доказана.

Вернемся к закону больших чисел. Мы его доказали ранее при условии, что случайные величины X_n обладают конечным вторым моментом. Пользуясь аппаратом характеристических функций, мы можем получить более сильное утверждение.

Теорема Хинчина (закон больших чисел). Пусть случайные величины X_1, X_2, \dots независимы и одинаково распределены, причем $\mathbf{E}|X_1| < \infty$. 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. Поскольку речь идет о сходимости к константе, достаточно доказать, что $F_{S_n/n} \Rightarrow I_a$, а для этого, в свою очередь, необходимо и достаточно, чтобы $\varphi_{S_n/n}(t) \rightarrow e^{ita}$.

Имеем для каждого фиксированного t

$$\varphi_{S_n/n}(t) = \varphi_{S_n}(t/n) = \left[1 + \frac{ita}{n} + o\left(\frac{t}{n}\right) \right]^n,$$

$$\ln \varphi_{S_n/n}(t) = n \ln \left(1 + \frac{ita}{n} + o\left(\frac{t}{n}\right) \right) = n \left(\frac{ita}{n} + o\left(\frac{t}{n}\right) \right) \rightarrow ita.$$

Теорема доказана.

4.6. Центральная предельная теорема

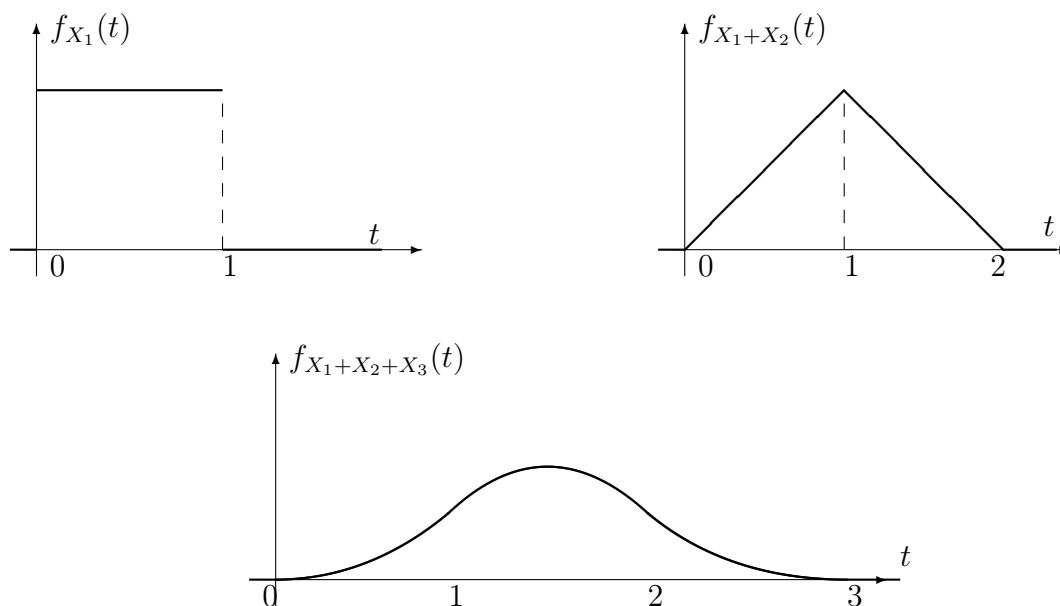
Как и ранее, будем иметь дело с последовательностью X_1, X_2, \dots независимых одинаково распределенных случайных величин, $S_n = X_1 + \dots + X_n$.

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. Достаточно доказать, что х. ф. случайной величины $\frac{S_n - na}{\sigma\sqrt{n}}$ сходится к $\exp\{-t^2/2\}$, т. е. к х. ф. стандартного нормального распределения. Обозначим для удобства $Y_k = X_k - a$, $k = 1, 2, \dots$. Тогда $\mathbf{E}Y_k = 0$, $\mathbf{D}Y_k = \mathbf{D}X_k = \sigma^2$, $S_n - na = Y_1 + \dots + Y_n$,

$$\varphi_{\frac{Y_1 + \dots + Y_n}{\sigma\sqrt{n}}}(t) = \varphi_{Y_1}^n\left(\frac{t}{\sigma\sqrt{n}}\right) = \left(1 - \frac{\sigma^2 t^2}{2\sigma^2 n} + o\left(\frac{t^2}{n}\right)\right)^n,$$

$$\ln \varphi_{Y_1}^n\left(\frac{t}{\sigma\sqrt{n}}\right) = n \ln\left(1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right)\right) = n\left(-\frac{t^2}{2n} + o\left(\frac{t^2}{n}\right)\right) \rightarrow -\frac{t^2}{2}.$$

Теорема доказана.

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. 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 $n \rightarrow \infty$. Как мы видели ранее, это следует из непрерывности предельной функции распределения.

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 (приводится без доказательства).

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

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

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. Частным случаем ЦПТ является следующая *интегральная теорема Муавра–Лапласа*, которая исторически появилась раньше ЦПТ.

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

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

Здесь S_n равно сумме независимых случайных величин, распределенных по закону Бернулли; $a = p$, $\sigma^2 = p(1-p)$.

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,

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

В качестве еще одного примера применения ЦПТ рассмотрим задачу вычисления интегралов методом Монте-Карло.

Пусть функция $g(y_1, \dots, y_k)$ определена на k -мерном единичном кубе V . Требуется вычислить интеграл

$$a = \int \dots \int_V g(y_1, \dots, y_k) dy_1 \dots dy_k.$$

Предположим, что известна константа C такая, что $|g(y_1, \dots, y_k)| \leq C$ при $(y_1, \dots, y_k) \in V$. Обозначим $Y = (Y_1, \dots, Y_k)$ случайный вектор, имеющий равномерное распределение на V . Тогда

$$\mathbf{E}g(Y_1, \dots, Y_k) = \int \dots \int_V g(y_1, \dots, y_k) dy_1 \dots dy_k = a,$$

$$\mathbf{D}g(Y_1, \dots, Y_k) = \int \dots \int_V (g(y_1, \dots, y_k) - a)^2 dy_1 \dots dy_k \leq 4C^2.$$

Пусть у нас имеется n независимых случайных векторов $(Y_1^{(i)}, \dots, Y_k^{(i)})$, $i = 1, \dots, n$, одинаково распределенных с (Y_1, \dots, Y_k) , и пусть $X_i = g(Y_1^{(i)}, \dots, Y_k^{(i)})$. Все случайные величины X_1, \dots, X_n независимы и одинаково распределены. Тогда в силу ЗБЧ при больших n

$$\frac{S_n}{n} = \frac{X_1 + \dots + X_n}{n} \approx \mathbf{E}X_1 = a.$$

Это соображение и лежит в основании метода Монте-Карло. Векторы $(Y_1^{(i)}, \dots, Y_k^{(i)})$ обычно моделируют с помощью генератора случайных чисел. Вопрос стоит о выборе числа n . Предположим, что нас вполне устроит, чтобы с вероятностью, близкой к единице, погрешность такого приближения не превышала малой величины Δ . Тогда в силу ЦПТ

$$\begin{aligned} \mathbf{P}\left(\left|\frac{S_n}{n} - a\right| < \Delta\right) &= \mathbf{P}\left(\left|\frac{S_n - na}{\sigma\sqrt{n}}\right| < \frac{\Delta\sqrt{n}}{\sigma}\right) \geq \mathbf{P}\left(\left|\frac{S_n - na}{\sigma\sqrt{n}}\right| < \frac{\Delta\sqrt{n}}{2C}\right) \approx \\ &\approx \Phi_{0,1}\left(\frac{\Delta\sqrt{n}}{2C}\right) - \Phi_{0,1}\left(-\frac{\Delta\sqrt{n}}{2C}\right) = 1 - 2\Phi_{0,1}\left(-\frac{\Delta\sqrt{n}}{2C}\right). \end{aligned}$$

Выбирая число n достаточно большим, можем сделать полученную вероятность как угодно близкой к единице. Заметим, что при выводе этой оценки мы воспользовались весьма грубой оценкой для дисперсии.

4.7. Оценка точности в теореме Пуассона.

В этом разделе мы докажем сформулированную ранее оценку скорости сходимости в теореме о приближении Пуассона для биномиального распределения.

Пусть S_n — число успехов в схеме Бернулли, p — вероятность успеха в одном испытании.

Теорема. Для любого множества $A \subset \{0, 1, 2, \dots\}$

$$\left| \mathbf{P}(S_n \in A) - \sum_{k \in A} \frac{(np)^k}{k!} e^{-np} \right| < np^2.$$

Доказательство. Воспользуемся так называемым методом одного вероятностного пространства. Он состоит в следующем. Коль скоро в теореме речь идет о сближении распределений и ничего не говорится о том, где заданы соответствующие случайные величины, то мы можем подобрать удобное для нас вероятностное пространство и задать на нем одновременно S_n и случайную величину с пуассоновским распределением. После чего оценим близость распределений этих случайных величин.

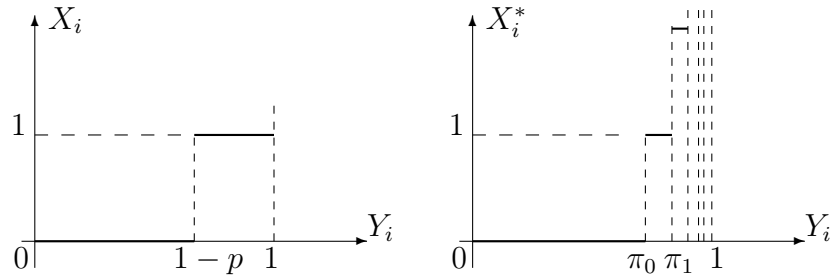
Пусть $\Omega = [0, 1]$, $\mathcal{S} = \mathcal{B}(\Omega)$. Для всякого интервала $A \subset \Omega$ положим $\mathbf{P}(A) = \lambda(A)$, где $\lambda(A)$ — длина интервала. Предположим, что на $[0, 1]$ независимо друг от друга бросаются n точек, обозначим их координаты Y_1, \dots, Y_n . Это независимые случайные величины, имеющие распределение $U_{0,1}$. Построим на этом же пространстве новые случайные величины. Для $i = 1, \dots, n$ положим

$$X_i = \begin{cases} 0, & \text{если } Y_i < 1 - p, \\ 1, & \text{если } Y_i \geq 1 - p, \end{cases}$$

$$X_i^* = \begin{cases} 0, & \text{если } Y_i < e^{-p} = \pi_0, \\ k \geq 1, & \text{если } Y_i \in [\pi_{k-1}, \pi_k), \end{cases}$$

где

$$\pi_k = e^{-p} \left(1 + \dots + \frac{p^k}{k!} \right), \quad \mathbf{P}(X_i^* = k) = \pi_k - \pi_{k-1} = e^{-p} \frac{p^k}{k!}.$$



Заметим, что $\pi_0 = e^{-p} > 1 - p$, поэтому $X_i \neq X_i^*$ только если $Y_i \in [1 - p, \pi_0)$ или $Y_i \in [\pi_1, 1]$. Значит, при каждом i

$$\mathbf{P}(X_i \neq X_i^*) = (e^{-p} - 1 + p) + (1 - e^{-p} - pe^{-p}) = p(1 - e^{-p}) < p^2.$$

Очевидно, случайные величины X_i независимы и имеют распределение Бернулли B_p , $S_n = X_1 + \dots + X_n \in B_{n,p}$. Случайные величины X_i^* также независимы и $S_n^* = X_1^* + \dots + X_n^* \in \Pi_{np}$. Имеем

$$\mathbf{P}(S_n \neq S_n^*) \leq \mathbf{P}(\cup_i \{X_i \neq X_i^*\}) \leq np^2.$$

Поэтому для любого множества $A \subset \{0, 1, 2, \dots\}$

$$\begin{aligned} \mathbf{P}(S_n \in A) &= \mathbf{P}(S_n \in A, S_n = S_n^*) + \mathbf{P}(S_n \in A, S_n \neq S_n^*) = \\ &= \mathbf{P}(S_n^* \in A) - \mathbf{P}(S_n^* \in A, S_n \neq S_n^*) + \mathbf{P}(S_n \in A, S_n \neq S_n^*), \end{aligned}$$

то есть

$$\begin{aligned} |\mathbf{P}(S_n \in A) - \mathbf{P}(S_n^* \in A)| &= |\mathbf{P}(S_n \in A, S_n \neq S_n^*) - \mathbf{P}(S_n^* \in A, S_n \neq S_n^*)| \leq \\ &\leq \max(\mathbf{P}(S_n^* \in A, S_n \neq S_n^*), \mathbf{P}(S_n \in A, S_n \neq S_n^*)) \leq \mathbf{P}(S_n \neq S_n^*) \leq np^2. \end{aligned}$$

Теорема доказана.

5. Цепи Маркова

5.1. Основные определения

В предыдущих разделах мы изучали последовательности независимых испытаний (например, в схеме Бернулли) и связанные с ними последовательности независимых случайных величин. Теперь рассмотрим простейший вариант зависимых испытаний.

Пусть некоторый объект в каждый момент времени может находиться в одном из состояний E_k , где $k = 0, \pm 1, \pm 2, \dots$; с течением времени он может переходить из одного состояния в другое. Время будем рассматривать дискретное: $n = 0, 1, 2, \dots$. Переходы из состояния в состояние происходят неким случайным образом, однако номер каждого последующего состояния зависит, кроме всего прочего, и от номера предыдущего состояния.

Рассмотрим некоторые примеры.

1. Объект — население города, состояние — число больных гриппом, отмечаемое ежедневно. Число больных завтра будет определяться числом больных сегодня, а также случайными факторами (кто-то заболел за сутки, кто-то выздоровел).

2. Капитал игрока после очередной игры. Он складывается из имеющегося капитала до игры плюс выигрыш (проигрыш можно считать выигрышем со знаком минус, так что капитал может принимать отрицательные значения).

3. Число особей в биологической популяции.

4. Число клиентов в банке.

5. Количество самолетов в аэропорту на каждый час. Оно складывается из числа самолетов, находившихся в аэропорту час назад, плюс число прилетевших и минус число улетевших в течение часа.

Чтобы перейти к точному определению, рассмотрим последовательность случайных величин $\{X_n, n = 0, 1, \dots\}$, которые принимают целые значения. Будем полагать $X_n = k$, если объект в момент времени n находится в состоянии E_k , $k = 0, \pm 1, \pm 2, \dots$. Таким образом, значение X_n равно номеру состояния в момент времени n .

Определение. Последовательность $\{X_n, n = 0, 1, \dots\}$ называется *цепью Маркова*, если для любых моментов времени $0 \leq n_1 < n_2 < \dots < n_k < m < n$ и для любых целых чисел $i_1, i_2, \dots, i_k, i, j$ выполняется равенство

$$\mathbf{P}(X_n = j / X_{n_1} = i_1, \dots, X_{n_k} = i_k, X_m = i) = \mathbf{P}(X_n = j / X_m = i).$$

Чтобы понять суть этого определения, представим себе, что момент m — это настоящее, моменты n_1, n_2, \dots, n_k находятся в прошлом, а n — момент времени, относящийся к будущему. Приведенное определение означает, что если известна предыстория эволюции объекта в моменты времени n_1, n_2, \dots, n_k и известно состояние объекта в настоящее время, то для будущего предыстория оказывается несущественной. Влияние оказывает только состояние объекта в настоящий момент времени.

Такого сорта зависимость характерна для приведенных выше примеров. Ее называют *марковской* по имени известного русского математика А.А.Маркова (1856 - 1922), в трудах которого впервые систематически изучалась такая зависимость.

Цепь называется *однородной*, если вероятности перехода $\mathbf{P}(X_n = j / X_{n-1} = i)$ не зависят от n . Мы будем изучать только однородные цепи Маркова.

Будем обозначать через $p_{ij} = \mathbf{P}(X_n = j / X_{n-1} = i)$ вероятности перехода из i -го состояния в j -е за один шаг и $p_{ij}(n) = \mathbf{P}(X_{n+k} = j / X_k = i) = \mathbf{P}(X_n = j / X_0 = i)$ —

вероятности перехода за n шагов (эти вероятности от k не зависят для однородных цепей).

Пусть задано распределение случайной величины X_0 , его называют *начальным распределением* цепи:

$$\pi_j^0 = \mathbf{P}(X_0 = j), \quad \sum_j \pi_j^0 = 1,$$

и пусть заданы также вероятности перехода $\{p_{ij}\}$. Этого достаточно, чтобы найти распределение цепи $\pi_j^n = \mathbf{P}(X_n = j)$ для любого момента времени n . Действительно, для любого j по формуле полной вероятности получаем

$$\pi_j^1 = \mathbf{P}(X_1 = j) = \sum_i \mathbf{P}(X_0 = i) \mathbf{P}(X_1 = j / X_0 = i) = \sum_i \pi_i^0 p_{ij}, \quad (6)$$

и аналогично

$$\pi_j^n = \mathbf{P}(X_n = j) = \sum_i \mathbf{P}(X_{n-1} = i) \mathbf{P}(X_n = j / X_{n-1} = i) = \sum_i \pi_i^{n-1} p_{ij}.$$

Предположим для простоты, что цепь имеет конечное множество состояний E_1, E_2, \dots, E_r . Тогда совокупность вероятностей перехода $\{p_{ij}\}$ образует матрицу $r \times r$, которую мы обозначим \mathbb{P} . Она, очевидно, обладает следующими свойствами:

- 1) $p_{ij} \geq 0$ при всех $i = 1, \dots, r; j = 1, \dots, r$;
- 2) $\sum_{j=1}^r p_{ij} = 1$ (т.е. сумма элементов любой строки равна 1).

Матрицы с указанными двумя свойствами называются *стохастическими*. Обозначим вектор-строку распределения X_n через $\pi^n = (\pi_1^n, \dots, \pi_r^n)$, тогда для вектора π^1 имеем в силу (6)

$$\pi^1 = \pi^0 \mathbb{P},$$

и аналогично для любого n

$$\pi^n = \pi^{n-1} \mathbb{P} = \dots = \pi^0 \mathbb{P}^n.$$

Кроме того,

$$\pi_j^n = \sum_{i=1}^r \mathbf{P}(X_0 = i) \mathbf{P}(X_n = j / X_0 = i) = \sum_{i=1}^r \pi_i^0 p_{ij}(n).$$

Это означает, что числа $p_{ij}(n)$ являются элементами матрицы \mathbb{P}^n .

Таким образом, знание начального распределения π^0 и матрицы переходных вероятностей \mathbb{P} позволяет вычислить распределение X_n в произвольный момент времени n .

Если множество состояний бесконечно, то и матрица \mathbb{P} будет бесконечной, но приведенные соотношения сохраняются.

Примеры.

1. Последовательность $Y_n, n \geq 0$, независимых целочисленных случайных величин, очевидно, является цепью Маркова.

2. Пусть $Y_n, n \geq 0$ — независимые целочисленные случайные величины. Тогда последовательность сумм $X_n = Y_0 + \dots + Y_n, n \geq 0$, образует цепь Маркова. Действительно, для любых моментов времени $0 \leq n_1 < n_2 < \dots < n_k < m < n$ и для

любых целых чисел $i_1, i_2, \dots, i_k, i, j$ имеем

$$\begin{aligned}
& \mathbf{P}(X_n = j / X_{n_1} = i_1, \dots, X_{n_k} = i_k, X_m = i) = \\
&= \frac{\mathbf{P}(X_n = j, X_{n_1} = i_1, \dots, X_{n_k} = i_k, X_m = i)}{\mathbf{P}(X_{n_1} = i_1, \dots, X_{n_k} = i_k, X_m = i)} = \\
&= \frac{\mathbf{P}(Y_{m+1} + \dots + Y_n = j - i, X_{n_1} = i_1, \dots, X_{n_k} = i_k, X_m = i)}{\mathbf{P}(X_{n_1} = i_1, \dots, X_{n_k} = i_k, X_m = i)} = \\
&= \frac{\mathbf{P}(Y_{m+1} + \dots + Y_n = j - i) \mathbf{P}(X_{n_1} = i_1, \dots, X_{n_k} = i_k, X_m = i)}{\mathbf{P}(X_{n_1} = i_1, \dots, X_{n_k} = i_k, X_m = i)} = \\
&= \mathbf{P}(Y_{m+1} + \dots + Y_n = j - i) = \mathbf{P}(X_n = j / X_m = i).
\end{aligned}$$

Если случайные величины Y_n в этих примерах вдобавок ко всему одинаково распределены, то цепи Маркова будут однородными. Нетрудно найти для них вероятности перехода за один шаг. Пусть $\mathbf{P}(Y_n = j) = p_j$. В первом из примеров

$$\mathbf{P}(Y_n = j / Y_{n-1} = i) = \mathbf{P}(Y_n = j) = p_j,$$

во втором из них

$$\mathbf{P}(X_n = j / X_{n-1} = i) = \mathbf{P}(Y_n = j - i) = p_{j-i}.$$

3. Предположим, что каждый день на склад завозится некоторое случайное число мешков с мукой, и должно вывозиться ежедневно также некоторое случайное число мешков. Считаем, что движение продукции в разные дни не связано друг с другом. Обозначим через X_n количество мешков с мукой на складе к концу n -го дня. Поскольку вместимость склада ограничена (скажем, числом M мешков), то, очевидно,

$$X_n = \begin{cases} X_{n-1} + Y_n, & \text{если } 0 \leq X_{n-1} + Y_n \leq M \\ 0, & \text{если } X_{n-1} + Y_n < 0, \\ M, & \text{если } X_{n-1} + Y_n > M, \end{cases}$$

где через Y_n обозначено предполагаемое приращение продукции (поступление минус вывоз) в n -й день. Нетрудно видеть, что последовательность X_n также образует цепь Маркова.

5.2. Возвратность состояний

Обозначим через

$$q_j(n) = \mathbf{P}(X_n = j, X_{n-1} \neq j, \dots, X_1 \neq j / X_0 = j)$$

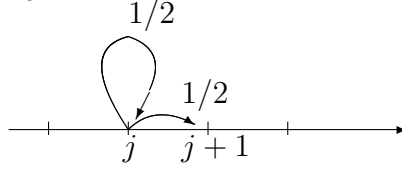
вероятность того, что, выйдя из состояния с номером j , наша цепь впервые вернется в него на n -м шаге. Пусть

$$Q_j = \sum_{n=1}^{\infty} q_j(n)$$

— вероятность того, что, выйдя из состояния с номером j , цепь когда-либо вернется в него.

Определение. Состояние E_j называется *возвратным*, если $Q_j = 1$, и *невозвратным*, если $Q_j < 1$.

Пример. Частица блуждает по целочисленным точкам вещественной оси, осуществляя с вероятностью $1/2$ в каждый момент времени $n = 1, 2, \dots$ прыжок вправо на единицу, и оставаясь на месте с вероятностью $1/2$. Это соответствует тому, что $p_{jj} = p_{j,j+1} = 1/2$ для любого j .



Ясно, что $q_j(1) = 1/2$ и $q_j(n) = 0$ при $n > 1$. Поэтому $Q_j = 1/2$ для любого j , и все состояния цепи невозвратны.

Теорема. Состояние E_j *возвратно тогда и только тогда, когда*

$$P_j = \sum_{n=1}^{\infty} p_{jj}(n) = \infty.$$

Для невозвратного E_j имеет место $Q_j = \frac{P_j}{1 + P_j}$.

Доказательство.

Имеет место соотношение

$$p_{jj}(n) = q_j(1)p_{jj}(n-1) + \dots + q_j(n-1)p_{jj}(1) + q_j(n). \quad (7)$$

Смысл его в следующем. Вероятность вернуться в j -е состояние за n шагов разбивается на перебор взаимно исключающих случаев в зависимости от того, за какое число шагов цепь *впервые* вернется в состояние E_j . Если, к примеру, впервые цепь вернется за k шагов (вероятность этого равна $q_j(k)$), то затем ей нужно где-то “погулять” и вернуться назад за оставшиеся $n - k$ шагов. Перебор вариантов по всем k осуществляется суммированием. Формально (7) получается из цепочки равенств

$$\begin{aligned} \mathbf{P}(X_n = j / X_0 = j) &= \frac{\mathbf{P}(X_n = j, X_0 = j)}{\mathbf{P}(X_0 = j)} = \\ &= \frac{1}{\mathbf{P}(X_0 = j)} \sum_{k=1}^n \mathbf{P}(X_0 = j, X_1 \neq j, \dots, X_{k-1} \neq j, X_k = j, X_n = j) = \\ &= \sum_{k=1}^n \frac{\mathbf{P}(X_0 = j, X_1 \neq j, \dots, X_{k-1} \neq j, X_k = j)}{\mathbf{P}(X_0 = j)} \times \\ &\times \frac{\mathbf{P}(X_0 = j, X_1 \neq j, \dots, X_{k-1} \neq j, X_k = j, X_n = j)}{\mathbf{P}(X_0 = j, X_1 \neq j, \dots, X_{k-1} \neq j, X_k = j)} = \sum_{k=1}^n q_j(k) p_{jj}(n-k). \end{aligned}$$

Далее нам потребуется понятие *производящей функции*. Ранее вводилось понятие производящей функции для распределения, сосредоточенного на решетке целых чисел. Пусть теперь $\{a_n, n \geq 1\}$ — произвольная ограниченная числовая последовательность, свойства которой требуется изучить. Производящей функцией этой последовательности называется сумма ряда

$$g(z) = \sum_{n=1}^{\infty} a_n z^n.$$

Этот ряд сходится абсолютно при каждом z из множества $|z| < 1$ и является там непрерывной (и даже аналитической) функцией. Если, к тому же, $\sum_{n=1}^{\infty} |a_n| < \infty$, то непрерывность будет иметь место при $|z| \leq 1$. Зная функцию $g(z)$, можно однозначно восстановить все коэффициенты a_n ; по поведению функции $g(z)$ можно определить многие свойства этих коэффициентов. Исследование свойств последовательности a_n через ее производящую функцию во многих случаях является весьма эффективным.

Мы воспользуемся этим инструментом.

Введем производящие функции

$$P_j(z) = \sum_{n=1}^{\infty} p_{jj}(n)z^n, \quad |z| < 1,$$

$$Q_j(z) = \sum_{n=1}^{\infty} q_j(n)z^n, \quad |z| \leq 1.$$

Умножим обе части равенства (7) на z^n и просуммируем по n (здесь $|z| < 1$):

$$\begin{aligned} P_j(z) &= \sum_{n=1}^{\infty} z^n \sum_{k=1}^n q_j(k)p_{jj}(n-k) = \\ &= \sum_{k=1}^{\infty} z^k \sum_{n=k}^{\infty} z^{n-k} q_j(k)p_{jj}(n-k) = \\ &= \sum_{k=1}^{\infty} z^k q_j(k) \sum_{m=0}^{\infty} z^m p_{jj}(m) = Q_j(z)(1 + P_j(z)). \end{aligned}$$

Отсюда получаем

$$Q_j(z) = \frac{P_j(z)}{1 + P_j(z)}, \quad P_j(z) = \frac{Q_j(z)}{1 - Q_j(z)}. \quad (8)$$

Пусть теперь $P_j = \infty$. Покажем, что в этом случае $P_j(z) \rightarrow \infty$ при $z \nearrow 1$. Действительно, для любого как угодно большого числа M можно найти число N такое, что $\sum_{n=1}^N p_{jj}(n) \geq 2M$. Это следует из расходимости ряда $P_j = \sum_{n=1}^{\infty} p_{jj}(n)$. Для положительных чисел z , достаточно близких к 1, будет выполняться $z^N \geq 1/2$ и

$$P_j(z) \geq \sum_{n=1}^N p_{jj}(n)z^n \geq \sum_{n=1}^N p_{jj}(n)z^N \geq M.$$

Итак, раз $P_j(z) \rightarrow \infty$, то $Q_j(z) \rightarrow 1$, это следует из первого равенства в (8). Поскольку функция $Q_j(z)$ непрерывна в точке $z = 1$, то при $z \rightarrow 1$

$$Q_j(z) \rightarrow Q_j(1) = Q_j = 1,$$

то есть состояние возвратно.

Обратно, пусть $Q_j = 1$. Тогда опять в силу непрерывности $Q_j(z) \rightarrow Q_j(1) = 1$ при $z \rightarrow 1$, а значит $P_j(z) \rightarrow \infty$ — это следует из второго равенства в (8). Отсюда следует, что $P_j = \infty$, поскольку в противном случае ряд

$$P_j(z) = \sum_{n=1}^{\infty} p_{jj}(n)z^n$$

сходился бы при $|z| \leq 1$ и был бы непрерывной функцией в точке $z = 1$, то есть было бы $P_j(z) \rightarrow P_j(1) < \infty$, что невозможно.

Если $P_j < \infty$, то при $z \rightarrow 1$ получаем из (8) $Q_j = \frac{P_j}{1 + P_j}$. Теорема доказана.

Определение. Состояния E_i и E_j называются *сообщающимися*, если $p_{ij}(m) > 0$ и $p_{ji}(k) > 0$ при некоторых $m \geq 1$ и $k \geq 1$.

Теорема солидарности. *Сообщающиеся состояния одновременно оба возвратны или оба невозвратны.*

Доказательство. Пусть для состояний E_i и E_j выполняется $p_{ij}(m) > 0$ и $p_{ji}(k) > 0$ при некоторых $m \geq 1$ и $k \geq 1$. Тогда при $n \geq 1$

$$p_{ii}(n + m + k) \geq p_{ij}(m)p_{jj}(n)p_{ji}(k), \quad p_{jj}(n + m + k) \geq p_{ji}(k)p_{ii}(n)p_{ij}(m),$$

поэтому ряды $\sum_{n=1}^{\infty} p_{ii}(n)$ и $\sum_{n=1}^{\infty} p_{jj}(n)$ сходятся или расходятся одновременно.

5.3. Эргодическая теорема

Мы видели, что цепи Маркова могут рассматриваться в качестве математических моделей, которые описывают эволюцию во времени того или иного объекта. Для приложений очень важно бывает выяснить условия, при которых объект с течением времени впадает в стационарный режим, то есть он по-прежнему может находиться в разных состояниях, но вероятности $\mathbf{P}(X_n = j)$ перестают зависеть от n . Теоремы, устанавливающие условия сходимости к стационарному режиму, обычно называют *эргодическими*. Мы рассмотрим одну из них.

Теорема (эргодическая). *Пусть цепь Маркова имеет конечное число r состояний, и при некотором $n_0 \geq 1$ все элементы $p_{ij}(n_0)$ матрицы \mathbb{P}^{n_0} положительны. Тогда существуют пределы*

$$\lim_{n \rightarrow \infty} p_{ij}(n) = p_j, \quad i, j = 1, \dots, r.$$

Предельные вероятности p_j не зависят от начального состояния i и являются единственным решением системы

$$\sum_{k=1}^r p_k p_{kj} = p_j, \quad j = 1, \dots, r, \quad \sum_{j=1}^r p_j = 1. \quad (9)$$

Замечание. Если выполнены условия теоремы, то поведение цепи с течением времени действительно стабилизируется: для каждого $j = 1, \dots, r$

$$\pi_j^n = \mathbf{P}(X_n = j) = \sum_{i=1}^r \pi_i^0 p_{ij}(n) \rightarrow \sum_{i=1}^r \pi_i^0 p_j = p_j.$$

По этой причине совокупность вероятностей $\{p_1, \dots, p_r\}$ называется *стационарным* распределением цепи. Если его взять в качестве начального распределения, то есть положить $\pi_j^0 = p_j$, $j = 1, \dots, r$, то из (9) будет следовать, что вектор вероятностей π^1 совпадает с π^0 , а значит и с π^n при всех $n \geq 1$. Это соответствует тому, что с самого начала цепь будет находиться в стационарном режиме.

Доказательство. Обозначим

$$M_k(n) = \max_i p_{ik}(n), \quad m_k(n) = \min_i p_{ik}(n).$$

Поскольку $m_k(n) \leq p_{ik}(n) \leq M_k(n)$ для всех i , то из соотношений

$$p_{ik}(n+1) = \sum_{l=1}^r p_{il}p_{lk}(n), \quad m_k(n) \sum_{l=1}^r p_{il} \leq \sum_{l=1}^r p_{il}p_{lk}(n) \leq M_k(n) \sum_{l=1}^r p_{il}$$

следует, что $m_k(n) \leq p_{ik}(n+1) \leq M_k(n)$ при всех i . Отсюда заключаем, что

$$m_k(n) \leq m_k(n+1) \leq M_k(n+1) \leq M_k(n).$$

Таким образом, существуют пределы последовательностей $m_k(n)$ и $M_k(n)$ при $n \rightarrow \infty$. Докажем, что эти пределы совпадают.

Пусть индексы i и j таковы, что

$$p_{ik}(n_0+n) = M_k(n_0+n), \quad p_{jk}(n_0+n) = m_k(n_0+n).$$

Из равенства $\mathbb{P}^{n_0+n} = \mathbb{P}^{n_0}\mathbb{P}^n$ следует

$$M_k(n_0+n) = p_{ik}(n_0+n) = \sum_{l=1}^r p_{il}(n_0)p_{lk}(n),$$

$$m_k(n_0+n) = p_{jk}(n_0+n) = \sum_{l=1}^r p_{jl}(n_0)p_{lk}(n).$$

Вычитая одно равенство из другого, получим

$$M_k(n_0+n) - m_k(n_0+n) = \sum_{l=1}^r (p_{il}(n_0) - p_{jl}(n_0))p_{lk}(n).$$

Пусть $A = \{l : p_{il}(n_0) - p_{jl}(n_0) \geq 0\}$, $B = \{l : p_{il}(n_0) - p_{jl}(n_0) < 0\}$. Очевидно, множество A не пусто. Тогда

$$\begin{aligned} M_k(n_0+n) - m_k(n_0+n) &= \sum_{l \in A} (p_{il}(n_0) - p_{jl}(n_0))p_{lk}(n) + \sum_{l \in B} (p_{il}(n_0) - p_{jl}(n_0))p_{lk}(n) \leq \\ &\leq M_k(n) \sum_{l \in A} (p_{il}(n_0) - p_{jl}(n_0)) + m_k(n) \sum_{l \in B} (p_{il}(n_0) - p_{jl}(n_0)) = \\ &= (M_k(n) - m_k(n)) \sum_{l \in A} (p_{il}(n_0) - p_{jl}(n_0)). \end{aligned}$$

Здесь мы воспользовались тем, что

$$\sum_{l \in B} (p_{il}(n_0) - p_{jl}(n_0)) = - \sum_{l \in A} (p_{il}(n_0) - p_{jl}(n_0)),$$

поскольку

$$\sum_{l=1}^r (p_{il}(n_0) - p_{jl}(n_0)) = 0 = \sum_{l \in A} (p_{il}(n_0) - p_{jl}(n_0)) + \sum_{l \in B} (p_{il}(n_0) - p_{jl}(n_0)).$$

Обозначим

$$d_{ij} = \sum_{l \in A} (p_{il}(n_0) - p_{jl}(n_0)).$$

Из условия теоремы следует, что $d_{ij} < 1$ при всех i, j , поэтому $d = \max_{i,j} d_{ij} < 1$.

Таким образом, мы приходим к неравенству: для всякого $n \geq 1$

$$M_k(n_0 + n) - m_k(n_0 + n) \leq d(M_k(n) - m_k(n)).$$

Устремляя n к бесконечности в этом соотношении, получим

$$\lim_{n \rightarrow \infty} (M_k(n_0 + n) - m_k(n_0 + n)) = \lim_{n \rightarrow \infty} (M_k(n) - m_k(n)) \leq d \lim_{n \rightarrow \infty} (M_k(n) - m_k(n)),$$

что возможно только при $\lim_{n \rightarrow \infty} (M_k(n) - m_k(n)) = 0$.

Напомним, что $m_k(n) \leq p_{ik}(n) \leq M_k(n)$, значит, при всех i вероятности $p_{ik}(n)$ сходятся к одному и тому же пределу p_k при $n \rightarrow \infty$.

Далее, переходя в равенстве

$$p_{ij}(n+1) = \sum_{k=1}^r p_{ik}(n)p_{kj}$$

к пределу при $n \rightarrow \infty$, получим

$$p_j = \sum_{k=1}^r p_k p_{kj}.$$

Кроме того, $\sum_{j=1}^r p_{ij}(n) = 1$. Переходя к пределу при $n \rightarrow \infty$ получаем $\sum_{j=1}^r p_j = 1$.

Нам осталось доказать, что числа $\{p_1, \dots, p_r\}$ являются единственным решением указанной системы уравнений. Предположим, что нашелся другой вектор $x = (x_1, \dots, x_r)$, для которого $\sum x_j = 1$ и $x_j = \sum_{k=1}^r x_k p_{kj}$, $j = 1, \dots, r$. Последнее означает, что $x = x\mathbb{P} = x\mathbb{P}^2 = \dots = x\mathbb{P}^n$ для любого $n \geq 1$. В покоординатной записи это выглядит так:

$$x_j = \sum_{k=1}^r x_k p_{kj}(n).$$

Переходя в этом равенстве к пределу при $n \rightarrow \infty$, получим

$$x_j = \sum_{k=1}^r x_k p_j = p_j.$$

Теорема доказана.

Если задана матрица \mathbb{P} вероятностей перехода, то для проверки условия теоремы нужно искать показатель степени n_0 , при котором все элементы матрицы \mathbb{P}^{n_0} отличны от нуля. Возведение матрицы в степень является весьма трудоемкой операцией. Ее можно избежать, если воспользоваться простой графической иллюстрацией. По матрице \mathbb{P} можно построить диаграмму, в которой состояния изображаются отдельными точками, а наличие положительной вероятности перехода из состояния в состояние показывается стрелочкой. Пусть, к примеру, $r = 4$,

$$\mathbb{P} = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1/3 & 0 & 1/3 & 1/3 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

Построим диаграмму.

Случайные величины $X_1^{(j)}, X_2^{(j)}, \dots$ равны числу потомков частиц из $(j-1)$ -го поколения, эти потомки формируют j -е поколение. Мы полагаем здесь, что располагаем независимыми последовательностями независимых случайных величин

$$\{X_n^{(1)}\}, \{X_n^{(2)}\}, \dots, \quad n \geq 1,$$

где при всех j и n

$$\mathbf{P}(X_n^{(j)} = k) = p_k, \quad k = 0, 1, \dots$$

Таким образом, мы построили модель простейшего ветвящегося случайного процесса (в литературе такие модели получили название процессов Гальтона-Ватсона). Последовательность Y_n является цепью Маркова, однако матрица переходных вероятностей для нее трудна для вычислений, поэтому для исследования ветвящихся процессов разработаны свои собственные методы. Обычно интересуются распределением числа частиц в n -м поколении, его предельным поведением при $n \rightarrow \infty$, вероятностью вырождения процесса в какой-то момент времени.

Весьма удобным инструментом исследования ветвящихся процессов являются производящие функции.

Пусть

$$g(z) = \mathbf{E} z^{X_1^{(1)}} = \sum_{k=0}^{\infty} z^k p_k, \quad |z| \leq 1,$$

— производящая функция потомства одной частицы (не важно, какой именно; случайные величины $X_n^{(j)}$ одинаково распределены) и пусть $g_n(z)$ означает n -ю итерацию функции $g(z)$, то есть

$$g_1(z) = g(z), \quad g_2(z) = g(g(z)), \quad \dots,$$

$$g_n(z) = g(g_{n-1}(z)) = g_{n-1}(g(z)).$$

Следующая теорема устанавливает вид производящей функции числа частиц в n -м поколении.

Теорема. Для любого $n \geq 1$

$$\mathbf{E} z^{Y_n} = \sum_{k=0}^{\infty} z^k \mathbf{P}(Y_n = k) = g_n(z).$$

Доказательство. Воспользуемся методом математической индукции. При $n = 1$ имеем $\mathbf{E} z^{Y_1} = g(z) = g_1(z)$. Предположим, что $\mathbf{E} z^{Y_{n-1}} = g_{n-1}(z)$. Тогда

$$\begin{aligned} \sum_{k=0}^{\infty} z^k \mathbf{P}(Y_n = k) &= \sum_{k=0}^{\infty} z^k \sum_{m=0}^{\infty} \mathbf{P}(Y_n = k, Y_{n-1} = m) = \\ &= \sum_{k=0}^{\infty} z^k \sum_{m=0}^{\infty} \mathbf{P}(X_1^{(n)} + \dots + X_m^{(n)} = k, Y_{n-1} = m) = \\ &= \sum_{k=0}^{\infty} z^k \sum_{m=0}^{\infty} \mathbf{P}(X_1^{(n)} + \dots + X_m^{(n)} = k) \mathbf{P}(Y_{n-1} = m) = \\ &= \sum_{m=0}^{\infty} \mathbf{P}(Y_{n-1} = m) \sum_{k=0}^{\infty} z^k \mathbf{P}(X_1^{(n)} + \dots + X_m^{(n)} = k) = \end{aligned}$$

$$\begin{aligned}
&= \sum_{m=0}^{\infty} \mathbf{E} \left(z^{X_1^{(n)} + \dots + X_m^{(n)}} \right) \mathbf{P}(Y_{n-1} = m) = \sum_{m=0}^{\infty} \mathbf{E} \left(z^{X_1^{(n)}} z^{X_2^{(n)}} \dots z^{X_m^{(n)}} \right) \mathbf{P}(Y_{n-1} = m) = \\
&= \sum_{m=0}^{\infty} \mathbf{E} z^{X_1^{(n)}} \mathbf{E} z^{X_2^{(n)}} \dots \mathbf{E} z^{X_m^{(n)}} \mathbf{P}(Y_{n-1} = m) = \\
&= \sum_{m=0}^{\infty} g^m(z) \mathbf{P}(Y_{n-1} = m) = g_{n-1}(g(z)) = g_n(z).
\end{aligned}$$

Теорема доказана.

Знание производящей функции $g_n(z)$ позволяет найти все коэффициенты при z^k (например, $\mathbf{P}(Y_n = 0) = g_n(0)$, $\mathbf{P}(Y_n = 1) = g'_n(0)$, $\mathbf{P}(Y_n = 2) = g''_n(0)/2$ и т.д.), а также изучать свойства этих вероятностей.

Мы рассмотрим далее вопрос о вероятности вырождения процесса.

Процесс вырождается, если $Y_n = 0$ при некотором n . Обозначим A_n событие, состоящее в том, что $Y_n = 0$. Тогда вырождению процесса будет соответствовать событие

$$A = \bigcup_{n=1}^{\infty} \{Y_n = 0\} = \bigcup_{n=1}^{\infty} A_n.$$

Обозначим через $r = \mathbf{P}(A)$ вероятность вырождения.

Теорема. Вероятность вырождения r равна наименьшему корню уравнения

$$z = g(z) \tag{10}$$

на отрезке $[0, 1]$.

Доказательство.

Очевидно, $A_1 \subset A_2 \subset \dots$, поэтому по свойству непрерывности вероятности

$$r = \mathbf{P}(A) = \lim_{n \rightarrow \infty} \mathbf{P}(A_n) = \lim_{n \rightarrow \infty} g_n(0).$$

Устремим n к бесконечности в равенстве

$$g_{n+1}(0) = g(g_n(0)),$$

тогда пределом левой части будет число r , а правая часть стремится к $g(r)$ в силу непрерывности функции $g(z)$, то есть действительно вероятность вырождения удовлетворяет соотношению $r = g(r)$. Однако у уравнения (10) могут быть и другие корни, поэтому осталось доказать, что r совпадает с наименьшим корнем этого уравнения на $[0, 1]$.

Тривиальный случай: если $\mathbf{P}(X_1^{(1)} = 1) = 1$, то $Y_n = 1$ при всех n , то есть вырождения не происходит и, естественно, $r = 0$. Поскольку в этом случае $g(z) \equiv z$, то уравнение (10) превращается в тождество $z = z$, наименьшим решением которого на $[0, 1]$ является нуль.

Пусть теперь $\mathbf{P}(X_1^{(1)} = 1) < 1$. Выясним, как выглядит график функции $g(z)$ на $[0, 1]$. Функция является выпуклой вниз, поскольку $g''(z) = \sum k(k-1)z^{k-2}p_k \geq 0$. Кроме того, $g(1) = 1$. Обозначим $a = \mathbf{E} X_1^{(1)}$ и заметим, что $a = \sum k p_k = g'(1)$. Рассмотрим два случая.

1) Предположим, что $a \leq 1$. Если $\mathbf{P}(X_1^{(1)} > 1) = 0$, то графиком функции $g(z)$ будет прямая $g(z) = p_0 + zp_1$ (Рис. 1а), причем $p_0 > 0$. Поскольку $g(1) = 1$ и $g'(1) < 1$, то единственным решением уравнения (10) на $[0, 1]$ будет число $z = 1$, то есть в этом случае $r = 1$. Если же $\mathbf{P}(X_1^{(1)} > 1) > 0$, то кривая $y = g(z)$ также будет пересекать прямую $y = z$ только при $z = 1$ (Рис. 1б), то есть и в этом случае $r = 1$.

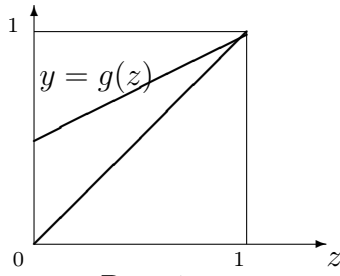


Рис. 1а

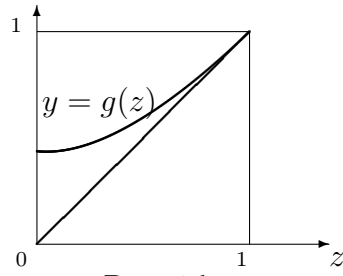


Рис. 1б

2) Пусть теперь $a > 1$. Тогда, разумеется, $\mathbf{P}(X_1^{(1)} > 1) > 0$, и уравнение (10) имеет ровно два корня $r_1 < 1$ и $r_2 = 1$ (См. Рис. 2).

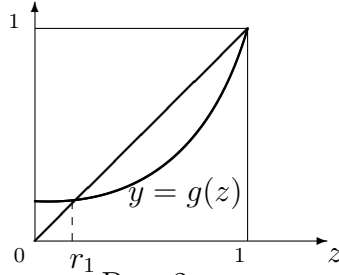


Рис. 2

Предположим, что $r = r_2 = 1$. Тогда $\delta_n = 1 - g_n(0) \rightarrow 1 - r = 0$ при $n \rightarrow \infty$ и, следовательно, $g(1 - \delta_n) < 1 - \delta_n$ при достаточно больших n . В этом случае

$$\delta_{n+1} = 1 - g_{n+1}(0) = 1 - g(g_n(0)) = 1 - g(1 - \delta_n) > 1 - (1 - \delta_n) = \delta_n,$$

что противоречит сходимости $\delta_n \rightarrow 0$. Значит, $r = r_1$. Теорема доказана.

Итак, мы видим, что возможность вырождения процесса определяется значением среднего числа потомков одной частицы. Если исключить из рассмотрения упомянутый выше тривиальный случай, то при $a \leq 1$ процесс вырождается с вероятностью единица, а при $a > 1$ вероятность вырождения меньше единицы (она обращается в нуль при $g(0) = 0 = \mathbf{P}(X_1^{(1)} = 0)$).

Ветвящийся процесс принято называть *докритическим*, если $a < 1$, *критическим* при $a = 1$, и *надкритическим*, если $a > 1$.

7. Случайные процессы с непрерывным временем

7.1. Общие определения. Винеровский процесс

До сих пор мы рассматривали семейства случайных величин, у которых множество индексов конечно или счетно, то есть мы рассматривали последовательности случайных величин $\{X_n, n \geq 1\}$. Во многих случаях значения индекса n интерпретировались как дискретные моменты времени.

Определение. *Случайным процессом* называется произвольное семейство случайных величин $\{X_t, t \in T \subset \mathbb{R}\}$, заданных на одном вероятностном пространстве.

В отличие от последовательностей случайных величин при рассмотрении случайных процессов чаще всего предполагают, что $T = [a, b]$ или $T = [0, \infty)$. Параметр t интерпретируется как время.

Отметим, что при фиксированном t мы имеем случайную величину $X_t(\omega)$, а при фиксированном ω получаем функцию $\{X_t, t \in T\}$, называемую обычно *траекторией процесса*.

Если зафиксируем t_1, \dots, t_n — некоторые значения параметра t , то им будет соответствовать случайный вектор $(X_{t_1}, X_{t_2}, \dots, X_{t_n})$. Распределения всевозможных таких векторов, когда $t_1 \in T, \dots, t_n \in T$, называются *конечномерными распределениями* процесса.

Предположим, что $X_0(\omega) = 0$.

Определение. Случайный процесс $\{X_t, t \geq 0\}$ называется *процессом с независимыми приращениями*, если для любых $0 \leq t_0 < t_1 < \dots < t_n$ случайные величины $X_{t_0}, X_{t_1} - X_{t_0}, \dots, X_{t_n} - X_{t_{n-1}}$ независимы.

Определение. Случайный процесс с независимыми приращениями называется *однородным*, если при любых $t_0 < t_1$ распределение $X_{t_1} - X_{t_0}$ определяется только длиной интервала $t_1 - t_0$ и не зависит от t_0 .

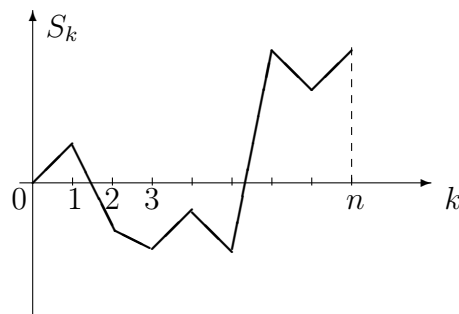
Среди однородных процессов с независимыми приращениями особую роль играет *винеровский* процесс. Детальное изучение его свойств требует привлечения весьма сложной математической техники, что выходит за рамки нашего курса. Мы ограничимся описанием качественной картины.

Итак, однородный случайный процесс X_t с независимыми приращениями называется *винеровским* (по имени известного математика Н. Винера), если $X_t \in \Phi_{0,t}$. Этот процесс называют также *процессом броуновского движения*, потому что его траектории наилучшим образом описывают движение броуновской частицы. Здесь имеется в виду одномерное движение частицы вдоль оси ординат, а по оси абсцисс по-прежнему откладывается время. Разумеется, для описания движения броуновской частицы на плоскости потребуется вводить двумерный винеровский процесс, чего мы делать не будем.

Траектории винеровского процесса устроены весьма сложным образом. Каждая траектория является непрерывной функцией переменной t , однако ни в одной точке производная этой функции не существует. Грубо говоря, траектория имеет изломы в каждой точке. Это является отражением того факта, что броуновская частица в каждый промежуток времени испытывает огромное число столкновений, меняющих направление ее движения.

Отметим еще одно необычное свойство траекторий. Если проложить вдоль траектории броуновской частицы (скажем, при $0 \leq t \leq 1$) ниточку, которая повторяет все изломы и изгибы траектории, то длина этой ниточки окажется бесконечной.

Вернемся к рассмотрению сумм независимых одинаково распределенных случайных величин $S_n = X_1 + \dots + X_n$. Предположим для простоты, что $\mathbf{E} X_1 = 0$, $\mathbf{E} X_1^2 = 1$, и пусть $S_0 = 0$. Если на координатной плоскости соединить отрезками прямых точки с координатами (k, S_k) , $k = 0, 1, \dots, n$, то получится ломаная, называемая траекторией случайного блуждания.



Известные предельные теоремы теории вероятностей (закон больших чисел, центральная предельная теорема) изучают предельное поведение при $n \rightarrow \infty$ распределения S_n , то есть ординаты конца этой ломаной. Однако можно изучать предельное поведение всей ломаной. Если сжать ее по оси абсцисс в n раз, а по оси ординат в \sqrt{n} раз, то получим ломаную, заданную уже на отрезке $[0,1]$, при этом ее звенья уменьшатся в размерах. Оказывается, при $n \rightarrow \infty$ эта сжатая ломаная будет все более походить на траекторию винеровского процесса. Другими словами, винеровский процесс является в некотором смысле предельным для траекторий случайного блуждания — и в этом заключается его дополнительная ценность. Мы можем изучать свойства траекторий случайного блуждания, пользуясь предельным переходом к винеровскому процессу, и наоборот, зная свойства винеровского процесса, делать соответствующие выводы для траекторий случайного блуждания.

Ниже мы рассмотрим более подробно один из наиболее простых и в то же время достаточно важных для приложений процессов с независимыми приращениями — пуассоновский.

7.2. Процесс Пуассона

Предположим, что в случайные моменты времени одно за другим происходят некоторые события. Нас интересует число таких событий, произошедших в промежутке времени $[0, t]$. Обозначим X_t это число.

Примерами таких ситуаций могут быть число частиц, зафиксированных прибором, число станков, вышедших из строя, число судов, прибывших в порт и т.д.

Относительно процесса появления событий будем предполагать следующее.

I. X_t — однородный процесс с независимыми приращениями.

Это означает, во-первых, что вероятность появления k событий в любом промежутке времени $[\tau, \tau + t]$ зависит только от t и не зависит от τ ; во-вторых, это все происходит вне зависимости от того, сколько событий и как появлялись до момента τ .

II. Обозначим $P_k(t) = \mathbf{P}(X_t = k)$, $k = 0, 1, \dots$, и будем предполагать, что при $h \rightarrow 0$

$$\mathbf{P}(X_h \geq 2) = \sum_{k=2}^{\infty} P_k(h) = o(h).$$

Это условие означает практическую невозможность появления двух или более событий за малый промежуток времени h .

Наша задача — найти в этих условиях вероятности $P_k(t)$. Мы покажем, что за исключением некоторых тривиальных случаев имеет место

$$P_k(t) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}, \quad k = 0, 1, \dots,$$

при некотором $\lambda > 0$.

Наши действия разобьем на несколько этапов.

1. Покажем, что за исключением некоторых простых ситуаций при некотором $\lambda > 0$ выполняется $P_0(t) = e^{-\lambda t}$.

Действительно, пусть $p = P_0(1)$. Разобьем отрезок времени $[0,1]$ на n равных частей; отсутствие событий за единицу времени означает, что на каждом из маленьких

промежутков времени длины $1/n$ происходит 0 событий. В силу независимости получаем $p = (P_0(1/n))^n$, откуда $P_0(1/n) = p^{1/n}$. Отсюда сразу же следует $P_0(k/n) = p^{k/n}$ при любом $k \geq 1$.

Покажем теперь, что вообще $P_0(t) = p^t$ при всех $t \geq 0$. Для каждого такого числа t и произвольного натурального n найдется число $k \geq 1$ такое, что

$$\frac{k-1}{n} \leq t < \frac{k}{n}.$$

Функция $P_0(t)$ не возрастает по t , поэтому

$$P_0\left(\frac{k-1}{n}\right) \geq P_0(t) \geq P_0\left(\frac{k}{n}\right),$$

или

$$p^{\frac{k-1}{n}} \geq P_0(t) \geq p^{\frac{k}{n}}.$$

Устремив $n \rightarrow \infty$ так, что $k/n \rightarrow t$, получим $P_0(t) = p^t$.

Возможны три случая: а) $p = 0$, б) $p = 1$, и в) $0 < p < 1$.

В первом из них $P_0(t) = 0$ для любого $t > 0$, то есть с вероятностью 1 в любом как угодно малом промежутке времени происходит хотя бы одно событие, а это эквивалентно тому, что в промежутке времени любой длины происходит бесконечно много событий. Это можно представлять себе как цепную реакцию при атомном взрыве, мы не будем останавливаться на этой крайности. В случае б) $P_0(t) = 1$, то есть события никогда не появляются. Таким образом, интерес вызывает только случай в). Положим $p = e^{-\lambda}$, здесь $0 < \lambda < \infty$. Тем самым получили

$$P_0(t) = e^{-\lambda t}.$$

2. Покажем, что при $h \rightarrow 0$

$$P_1(h) = \lambda h + o(h).$$

Для этого воспользуемся тем, что $P_0(h) = e^{-\lambda h} = 1 - \lambda h + o(h)$, и

$$P_0(h) + P_1(h) + \sum_{k=2}^{\infty} P_k(h) = 1.$$

Отсюда

$$P_1(h) = 1 - P_0(h) - \sum_{k=2}^{\infty} P_k(h) = \lambda h + o(h).$$

3. В этом пункте мы покажем, что вероятности $P_k(t)$ удовлетворяют некоторой системе дифференциальных уравнений.

Для $t \geq 0$ и $h > 0$ имеем

$$P_k(t+h) = \sum_{j=0}^k P_j(t)P_{k-j}(h)$$

(мы перебираем здесь все возможности о том, сколько событий произошло за время t и за последующий промежуток времени длины h).

Если $h \rightarrow 0$, то

$$\sum_{j=0}^{k-2} P_j(t)P_{k-j}(h) \leq \sum_{j=0}^{k-2} P_{k-j}(h) = \sum_{i=2}^k P_i(h) = o(h),$$

поэтому при $k \geq 1$

$$\begin{aligned} P_k(t+h) &= P_k(t)P_0(h) + P_{k-1}(t)P_1(h) + o(h) = \\ &= P_k(t)(1 - \lambda h + o(h)) + P_{k-1}(t)(\lambda h + o(h)) + o(h) = \\ &= P_k(t)(1 - \lambda h) + P_{k-1}(t)\lambda h + o(h). \end{aligned}$$

Отсюда получаем

$$\frac{P_k(t+h) - P_k(t)}{h} = -\lambda P_k(t) + \lambda P_{k-1}(t) + \frac{o(h)}{h}, \quad k = 1, 2, \dots$$

Устремим $h \rightarrow 0$. Поскольку при этом предел правой части существует, то он будет существовать и для левой части. В результате получаем

$$P'_k(t) = -\lambda P_k(t) + \lambda P_{k-1}(t), \quad k = 1, 2, \dots$$

К этой системе можно добавить соотношение

$$P'_0(t) = -\lambda P_0(t),$$

которое следует из формулы $P_0(t) = e^{-\lambda t}$. Выберем начальные условия — они диктуются логикой здравого смысла:

$$P_0(0) = 1, \quad P_k(0) = 0 \quad \text{при } k \geq 1.$$

4. Решение системы уравнений.

Воспользуемся методом производящих функций. Обозначим

$$g(z, t) = \sum_{k=0}^{\infty} z^k P_k(t), \quad |z| \leq 1. \quad (11)$$

Умножим полученные нами уравнения для $P_k(t)$ на z^k и просуммируем по k :

$$\sum_{k=0}^{\infty} z^k P'_k(t) = -\lambda \sum_{k=0}^{\infty} z^k P_k(t) + \lambda \sum_{k=1}^{\infty} z^k P_{k-1}(t),$$

или, что то же самое,

$$\frac{\partial g(z, t)}{\partial t} = -\lambda g(z, t) + \lambda z g(z, t) = \lambda(z-1)g(z, t).$$

Перепишем полученное уравнение в виде

$$\frac{\partial \ln g(z, t)}{\partial t} = \lambda(z-1),$$

откуда

$$\ln g(z, t) = \lambda(z-1)t + C.$$

В силу выбранных начальных условий при $t = 0$ имеем $g(z, 0) = 1$, то есть $C = 0$. В итоге получаем

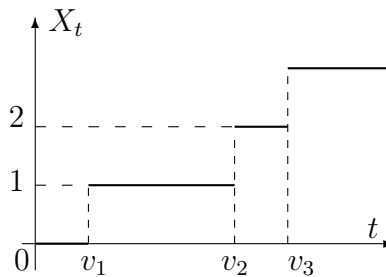
$$g(z, t) = e^{\lambda(z-1)t} = e^{-\lambda t} e^{\lambda z t} = e^{-\lambda t} \sum_{k=0}^{\infty} \frac{(\lambda t)^k}{k!} z^k. \quad (12)$$

Сравнивая коэффициенты разложений (11) и (12), обнаруживаем, что

$$P_k(t) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}, \quad k = 0, 1, \dots,$$

то есть мы получили вероятности, соответствующие распределению Пуассона с параметром λt . По этой причине изучаемый процесс называется пуассоновским. Иногда говорят о пуассоновском *потоке событий*.

Обозначим через τ_1 длину промежутка времени от нуля до первого появления события. Очевидно, $\mathbf{P}(\tau_1 > t) = P_0(t) = e^{-\lambda t}$, то есть случайная величина τ_1 распределена по показательному закону с параметром λ . Пусть также τ_2 — длина промежутка времени между первым и вторым событиями, τ_3 — длина промежутка времени между вторым и третьим событиями, и так далее. Можно показать, что все случайные величины τ_1, τ_2, \dots независимы и одинаково распределены по показательному закону с параметром λ . Траектории процесса X_t выглядят следующим образом.



Здесь обозначено $v_k = \tau_1 + \dots + \tau_k$. Момент v_k k -го появления события равен сумме независимых случайных величин, имеющих показательное распределение с параметром λ , поэтому $v_k \in \Gamma_{\lambda, k}$:

$$f_{v_k}(t) = \begin{cases} \frac{\lambda^k}{(k-1)!} t^{k-1} e^{-\lambda t}, & t > 0, \\ 0, & \text{иначе.} \end{cases}$$

Этот частный случай гамма-распределения принято называть распределением *Эрланга*.

Нетрудно вычислить среднее число событий, происходящих за время t :

$$\mathbf{E} X_t = \sum_{k=0}^{\infty} k P_k(t) = \lambda t.$$

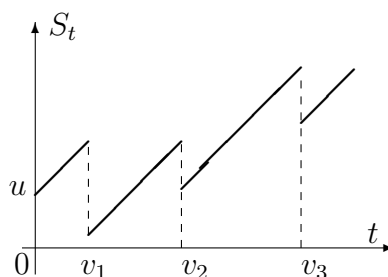
Среднее число событий, происходящих за единицу времени, называется *интенсивностью* пуассоновского процесса; мы видим, что она совпадает с параметром λ распределения промежутка времени между двумя последовательными событиями.

Вернемся к анализу траекторий процесса Пуассона. Они являются кусочно постоянными. При возрастании t величина X_t остается постоянной в течение промежутка времени, имеющего показательное распределение, а затем скачком увеличивается на единицу. Можно рассмотреть более общую конструкцию, когда X_t в момент времени $t = v_1$ увеличивается на случайную величину Y_1 , в момент времени v_2 происходит скачок на величину Y_2 и так далее, где Y_1, Y_2, \dots — последовательность независимых одинаково распределенных случайных величин, не зависящая также от последовательности τ_1, τ_2, \dots . Полученный таким образом случайный процесс Z_t называется *обобщенным* процессом Пуассона. Разумеется, его изучать труднее, чем обычный

процесс Пуассона, и мы не будем этого делать. Отметим только, что если к построенному процессу добавить еще линейный снос (то есть рассмотреть процесс вида $S_t = u + vt + Z_t$), то получится процесс, имеющий многочисленные приложения в теории страхования.

Рассмотрим более подробно эту модель.

Предположим, что страховая компания начинает в момент времени $t = 0$ свою деятельность, имея стартовый капитал u . Доход компании формируется из страховых взносов. Мы будем считать их постоянными во времени, то есть за время t суммарные поступления взносов составляют величину vt при некотором $v > 0$. Через случайные промежутки времени τ_1, τ_2, \dots происходят некоторые события, вынуждающие компанию делать страховые выплаты Y_1, Y_2, \dots ; при этом капитал компании скачкообразно уменьшается.



При изучении таких процессов наибольший интерес вызывает нахождение *вероятности разорения*, то есть вероятности того, что траектория процесса когда-либо коснется оси абсцисс. Эта задача не проста, и ее решение выходит за рамки нашего курса. Однако исследование таких процессов во многом опирается на установленные нами свойства процесса Пуассона.

Список использованной литературы

- Боровков А.А. Теория вероятностей. М.: Наука, 1986.
 Гнеденко Б. В. Курс теории вероятностей. М.: Наука, 1988.
 Розанов Ю.А. Случайные процессы. Краткий курс. М.: Наука, 1979.
 Севастьянов Б.А. Курс теории вероятностей и математической статистики. М.: 1982.