Лекция

«Законы распределения случайных величин»

Санкт-Петербург

2019

Содержание

1 Генеральная совокупность и выборка

В предыдущей лекции мы рассмотрели такие первостепенные понятия теории вероятностей, как случайный эксперимент, случайное событие, определили понятие вероятности события. Кроме того, мы построили несколько вероятностных моделей, среди которых: классическая вероятностная схема, схема с неравновозможными исходами, схема с условной вероятностью, геометрическая вероятность и другие.

1.1 Понятие генеральной совокупности и выборки

В завершении лекции мы познакомились с таким, на первый взгляд, теоретическим понятием, как случайная величина, узнали, как можно найти ее математическое ожидание и дисперсию, а также за что отвечают эти параметры. Что же, можно считать, что нам повезло, если мы знаем распределение случайной величины. Почему повезло? Потому что в жизни, оказывается, все обстоит несколько иначе: это распределение (или характеристики этого распределения) как раз-таки и хочется найти. Смотрите.

Обычно исследователь имеет дело с набором данных, который был собран в результате некоторого случайного эксперимента (или процесса): количество литров бензина, истраченного каждый день в течение месяца, количество клиентов автомойки за день в течение недели, рост первых 100 призывников в военкомате с начала призывной кампании, и так далее. Все эти данные являются результатом некоторого наблюдения.

При этом крайне важно задаться вопросом: а есть ли какая-то закономерность в этих данных? Скажем, если сравнить рост первых 100 призывников этого года и прошлого, понятно, что точно таких же данных мы не получим, но будет ли картина похожей? А что изменится, и изменится ли, если собрать больше наблюдений, скажем, 1000 или 10000? Очевидно, что во всех случаях будет наблюдаться относительно небольшое число людей, обладающих очень высоким и очень низким ростом, а большинство наблюдений будут находиться в некотором «условно среднем» диапазоне.

Оказывается, что при решении прикладных задач, при наблюдении тех или иных явлений, многие закономерности носят вероятностных характер. А что означает последняя фраза?

Она означает, что изучаемая нами закономерность (явление) случайна в своих проявлениях и, тем самым, описывается случайной величиной ξ, которая, в свою очередь, имеет вероятностное распределение. Для случая призывников, вероятность иметь рост, мало отличающийся от среднего, высока, а чем дальше рост от среднего, тем вероятность обладать таким ростом меньше.

Зная же распределение ξ, можно вычислять такие полезные параметры, как: математическое ожидание $\mathsf{E}\xi$, дисперсию $\mathsf{D}\xi = \mathsf{E}\left(\xi - \mathsf{E}\xi\right)^2$, вероятности попадания в те или иные множества и многое-многое другое. Иными словами, зная распределение интересующей нас случайной величины, мы сразу знаем очень много, а также можем получать сразу достаточно большой объем полезной информации.

Итак, пусть у нас имеется n проявлений некоторой вероятностной закономерности. Иными словами, мы наблюдаем n значений некоторой случайной величины ξ, называемой еще генеральной совокупностью, имеющей какое-то (неизвестное нам) распределение. Тогда перед нами обычный числовой n-мерный вектор $X = (x_1, x_2, ..., x_n)$ – это выборка после эксперимента. Итак,

Определение 1.1.1 *Пусть ξ – рассматриваемая нами случайная величина. Выборкой (после эксперимента) $X = (x_1, x_2, ..., x_n) \in \mathbb{R}^n$ называется n независимых реализаций случайной величины ξ. Последнюю часто называют генеральной совокупностью.*

Например, пусть ξ – это рост случайно взятого человека, тогда

$$X = (175, 182, 168, 155, 192)$$

это выборка объема 5 из генеральной совокупности ξ. Хорошо, допустим теперь мы хотим вычислить среднее, а точнее предположить, чему равно математическое ожидание ξ. Как бы это сделать? Вполне логично посчитать среднее арифметическое, называемое выборочным средним:

$$\overline{X} = \frac{175 + 182 + 168 + 155 + 192}{5} = 174.4.$$

Но является ли полученное значение математическим ожиданием генеральной совокупности ξ? Интуитивно понятно, что не является, так как если выбрать других 5 человек, то и среднее значение, скорее всего, поменяется. Тогда, видимо, правильнее сказать, что выборочное среднее $\overline{X}$ является оценкой математического ожидания генеральной совокупности ξ.

Важно еще понимать и вот какой момент, называемый неоднородностью. Придя в детский сад и померив людей в нем, получим одно число (вероятно, сравнительно небольшое, даже рост воспитателей и нянечек не поможет), а измерив рост членов баскетбольной команды – разительно отличающееся. Дело в том, что выборки-то берутся из разных распределений! Поэтому мы всегда должны понимать, какое именно распределение мы изучаем по данной выборке. А то получится вроде ситуации с опросом в интернете с вопросом: «Пользуетесь ли вы интернетом»?

1.2 Эмпирическое распределение

С каждой конкретной выборкой, полученной в результате эксперимента, то есть с выборкой $X = (x_1, x_2, ..., x_n)$, разумно связать новую случайную величину ξ^*. Такую, которая каждое значение выборки x_i принимает с вероятностью $\frac{1}{n}$ – у нас же нет любимчиков в выборке. В итоге, можно написать следующую таблицу

$$\begin{array}{c|c|c|c|c} \xi^* & x_1 & x_2 & ... & x_n \\ \hline \widetilde{\mathsf{P}} & \frac{1}{n} & \frac{1}{n} & ... & \frac{1}{n} \end{array}.$$

Кстати, как легко видеть, случайная величина, а точнее ее распределение, конечно, зависит от n, от объема выборки, но мы не будем наделять ее дополнительным индексом.

Так как случайная величина ξ^* не является истинной случайной величиной ξ, то вероятность, с которой она принимает те или иные значения, будем обозначать $\widetilde{\mathsf{P}}$. Например, имея в результате эксперимента, показывающем количество голов, забитых в пяти матчах, выборку

$$X = (1, 2, 1, 3, 1)$$

объема 5, соответсвующее распределение новой случайной величины ξ^* записывается, как

$$\begin{array}{c|c|c|c} \xi^* & 1 & 2 & 3 \\ \hline \widetilde{\mathsf{P}} & \frac{3}{5} & \frac{1}{5} & \frac{1}{5} \end{array}.$$

Оказывается, что с ростом объема выборки n распределение ξ^* все лучше и лучше приближает истинное (нам неизвестное) распределение генеральной совокупности ξ.

Наверное, теперь ясна основная идея: может быть, в качестве оценки некоторой характеристики генеральной совокупности ξ, можно использовать соответствующую характеристику ξ^*, раз распределение последней так неплохо приближает истинное? Попробуем.

Тогда в качестве оценки $\widehat{\mathsf{E}\xi}$ математического ожидания ξ логично взять соответствующее математическое ожидание $\widetilde{\mathsf{E}}\xi^*$, то есть, по сути, выборочное среднее:

$$\widehat{\mathsf{E}\xi} = \widetilde{\mathsf{E}}\xi^* = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

В качестве оценки $\widehat{\mathsf{D}\xi}$ дисперсии генеральной совокупности ξ логично взять дисперсию случайной величины ξ^*, называемую выборочной дисперсией:

$$\widehat{\mathsf{D}\xi} = \widetilde{\mathsf{D}}\xi^* = \widetilde{\mathsf{E}}(\xi^* - \widetilde{\mathsf{E}}\xi^*)^2 = S^2 = \frac{1}{n}\sum_{i=1}^{n} \left(x_i - \overline{X}\right)^2.$$

Замечание 1.2.1 *В предположении, что с ростом объема выборки n распределение ξ^* все лучше и лучше приближает истинное распределение генеральной совокупности ξ, есть некоторая неточность, которая будет устранена в следующих лекциях.*

Пример 1.2.1 *Найдем выборочное среднее и выборочную дисперсию для ранее рассмотренной выборки*

$$X = (1, 2, 1, 3, 1)$$

и построенной по ней эмпирической случайной величины ξ^*, распределение которой задается таблицей:

ξ^*	1	2	3
$\widetilde{\mathsf{P}}$	$\frac{3}{5}$	$\frac{1}{5}$	$\frac{1}{5}$

Выборочное среднее $\overline{X}$ находится, как

$$\overline{X} = \frac{1+2+1+3+1}{5} = \frac{8}{5} = 1.6 = \widetilde{\mathsf{E}}\xi^*.$$

Что это означает? Это означает, что, согласно выборке, стоит ожидать в среднем $1-2$ гола за матч.

Выборочная дисперсия находится, как:

$$\widehat{\mathsf{D}\xi} = \frac{(1-1.6)^2 + (2-1.6)^2 + (1-1.6)^2 + (3-1.6)^2 + (1-1.6)^2}{5} =$$

$$= \frac{3 \cdot (-0.6)^2 + 0.4^2 + 2.4^2}{5} = 0.64 = \widetilde{\mathsf{D}}\xi^*,$$

а корень из нее, то есть оценка $\widehat{\sigma}$, равен 0.8.

Какой можно сделать вывод? Такой, что согласно выборке, наиболее вероятным стоит ожидать от $1.6 - 0.8 = 0.8$ *до* $1.6 + 0.8 = 2.4$ *голов, то есть, в среднем,* $1-3$ *гола. Это не значит, что команда не может забить, скажем,* 6 *голов, или не забить голов вовсе. Просто эти события имею маленькую вероятность.*

Замечание 1.2.2 *Напомним, что полученные значения являются оценками истинных значений математического ожидания и дисперсии некоторой генеральной совокупности (случайной величины) ξ, однако неизбежно возникает вопрос, а насколько хороши эти оценки, и существуют ли какие-то еще? Мы ответим на этот вопрос в одной из следующих лекций.*

2 Основные законы распределения дискретных случайных величин

Мы уже неоднократно говорили, что выборка – это некоторое количество независимых реализаций генеральной совокупности ξ, то есть какой-то конечный набор значений, которые приняла эта случайная величина в результате эксперимента. Очевидно, что закономерности, проявляющиеся в результатах экспериментов, могут описываться случайными величинами с различными распределениями. Оказывается, множество различных явлений, с которыми мы сталкиваемся ежедневно, описываются случайными величинами, имеющими достаточно ограниченный набор распределений. Рассмотрим самые часто используемые.

2.1 Распределение Бернулли

В этом пункте поговорим про так называемое распределение Бернулли. Начнем с такого примера, навеянного собранной статистикой.

Пример 2.1.1 *Ученые подсчитали, что на каждые* 100 *девочек рождается от* 104 *до* 107 *мальчиков (данные колеблются в зависимости от страны). Для удобства будем считать, что рождается* 106 *мальчиков. Можно предположить, основываясь на частоте, что вероятность рождения мальчика равна*

$$\mathsf{P} = \frac{106}{(100+106)} \approx 0.52.$$

Пусть в некоторой семье уже есть один ребенок, и это девочка. Тогда, если родится девочка, родителям не придется тратиться на одежду и игрушки, так как ими явно поделится старшая сестра. Если же родится мальчик, то дополнительных трат не избежать. Тогда распределение случайной величины ξ*, показывающей наличие или отсутствие трат на нового ребенка, можно задать таблицей*

ξ	0	1
P	0.52	0.48

Оказывается, примеров такого вида существует достаточно большое количество. Под описываемую нами схему подходит любой эксперимент, который завершается одним из двух возможных исходов: (условно) успехом, и ему ставится в соответствие 1, или неудачей, которой ставится в соответствие 0: попал в створ ворот или промахнулся, дали кредит или не дали, сдал экзамен или не сдал, голоден или нет и так далее.

Определение 2.1.1 *Говорят, что случайная величина ξ имеет распределение Бернулли с параметром p (пишут $\xi \sim \mathsf{B}_p$), если ее распределение задается таблицей*

$$\begin{array}{c|c|c} \xi & 0 & 1 \\ \hline \mathsf{P} & 1-p & p \end{array},$$

где $p \in (0,1)$.

Замечание 2.1.1 *Величину $(1-p)$ часто обозначают через (как) q. Кроме того, вероятность p условно называют «вероятностью успеха», а q – «вероятностью неудачи».*

Найдем математическое ожидание случайной величины ξ, имеющей распределение Бернулли. По определению математического ожидания,

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

Для нахождения дисперсии, воспользуемся формулой:

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

Если $\xi \sim \mathsf{B}_p$, то распределение случайной величины ξ^2 задается таблицей

$$\begin{array}{c|c|c} \xi^2 & 0 & 1 \\ \hline \mathsf{P} & 1-p & p \end{array}.$$

Тогда,

$$\mathsf{E}\xi^2 = 0 \cdot (1-p) + 1 \cdot p = p,$$

а значит

$$\mathsf{D}\xi = \mathsf{E}\xi^2 - (\mathsf{E}\xi)^2 = p - p^2 = p(1-p) = pq.$$

В итоге, если $\xi \sim \mathsf{B}_p$, то

$$\mathsf{E}\xi = p, \quad \mathsf{D}\xi = p(1-p) = pq.$$

2.2 Биномиальное распределение

Рассмотрим закон распределения случайной величины, имеющей непосредственное отношение к последовательностям испытаний по схеме Бернулли.

Пусть случайная величина ξ показывает число успехов в серии из n испытаний по схеме Бернулли с вероятностью успеха p в каждом испытании.

Определение 2.2.1 *Говорят, что случайная величина ξ имеет биномиальное распределение, и пишут $\xi \sim \mathsf{Bin}(n,p)$, $p \in (0,1)$, $n \in \mathbb{N}$, если она принимает значения $0,1,2,...,n$ с вероятностями*

$$\mathsf{P}(\xi = k) = C_n^k p^k (1-p)^{n-k}, \quad k \in \{0,1,...,n\}.$$

Таблица распределения случайной величины, имеющей биномиальное распределение, имеет вид

ξ	0	1	$\dots$	$n-1$	n
P	$(1-p)^n$	$C_n^1 p(1-p)^{n-1}$	$\dots$	$C_n^{n-1}p^{n-1}(1-p)$	p^n

Замечание 2.2.1 *Заметим, что для того, чтобы приведенная таблица задавала распределение случайной величины ξ, необходимо, чтобы*

$$\sum_{m=0}^{n} C_n^m p^m q^{n-m} = 1,$$

то есть чтобы сумма чисел, стоящих во второй строке таблицы, давала единицу. Последнее следует из формулы бинома Ньютона

$$(a+b)^n = \sum_{m=0}^{n} C_n^m a^m b^{n-m},$$

если положить $a = p$, $b = 1-p$ (в этом случае $(a+b)^n = (p+1-p)^n = 1$).

Пример 2.2.1 *Студент выполняет тест, содержащий 8 вопросов с четырьмя вариантами ответа каждый, из которых только один верный. Построим ожидаемое распределение количества правильно решенных заданий, если ответы выбираются наугад.*

Ясно, что рассматриваемая случайная величина имеет биномиальное распределение $\mathsf{Bin}(8, 0.25)$. Таблица распределения (значения второй строки которой округлены так, чтобы и правда получилось распределение, то есть чтобы сумма чисел, стоящих в этой строке, равнялась единице), имеет следующий вид

ξ	0	1	2	3	4	5	6	7	8
P	0.1	0.267	0.311	0.208	0.087	0.023	0.004	0	0

Вероятности соответствующих значений случайной величины вычисляются по формулам

$$\mathsf{P}(\xi = k) = C_8^k \cdot 0.25^k \cdot 0.75^{8-k}, \quad k \in \{0,1,2,...,8\}.$$

Ответим на вопрос: какова вероятность сдать тест, если для этого необходимо ответить не менее, чем на 5 *вопросов. Ясно, что*

$$\mathsf{P}\left(\xi \geq 5\right) = \mathsf{P}(\xi = 5) + \mathsf{P}(\xi = 6) + \mathsf{P}(\xi = 7) + \mathsf{P}(\xi = 8) =$$

$$= 0.023 + 0.004 = 0.027.$$

Давайте теперь вычислим математическое ожидание случайной величины ξ, имеющей биномиальное распределение. А для этого рассмотрим случайные величины:

$$\xi_i = \begin{cases} 1, & \text{если в } i\text{-том испытании произошел успех} \\ 0, & \text{если в } i\text{-том испытании произошла неудача,} \end{cases}$$

где $i \in \{1, 2, ..., n\}$

В наших условиях все эти случайные величины ξ_i имеют распределение Бернулли B_p с одинаковым параметром p, причем

$$\xi = \xi_1 + \xi_2 + \ldots + \xi_n,$$

так как число слагаемых, равных единице, равно количеству успехов в n испытаниях. Тогда, по свойству математического ожидания,

$$\mathsf{E}\xi = \mathsf{E}\left(\xi_1 + \xi_2 + ... + \xi_n\right) = \mathsf{E}\xi_1 + \mathsf{E}\xi_2 + ... + \mathsf{E}\xi_n = p + p + ... + p = np.$$

Для нахождения дисперсии воспользуемся тем, что испытания в схеме Бернулли независимы, а значит и случайные величины ξ_i будут независимы, тогда, по свойствам дисперсии,

$$\mathsf{D}\xi = \mathsf{D}\left(\sum_{i=1}^{n} \xi_i\right) = \sum_{i=1}^{n} \mathsf{D}\xi_i = \sum_{i=1}^{n} pq = npq.$$

Пример 2.2.2 *Продолжим пример про студента, выполняющего тест. Найдем математическое ожидание и дисперсию. Напомним, что число вопросов (испытаний)* $n = 8$*, вероятность успеха в каждом –* $p = 0.25$*, тогда*

$$\mathsf{E}\xi = np = 8 \cdot 0.25 = 2,$$

$$\mathsf{D}\xi = npq = 8 \cdot 0.25 \cdot 0.75 = 1.5.$$

$$\sigma = \sqrt{\mathsf{D}\xi} = \sqrt{1.5} \approx 1.22$$

Иными словами, в среднем из 8 *вопросов студент верно ответит на* 2*. При этом, в среднем, число верных ответов будет колебаться от* 0.78 *до* 3.22*.*

2.3 Распределение Пуассона

В этом пункте сначала введем формальное определение, а потом поясним область применения.

Определение 2.3.1 *Говорят, что случайная величина имеет распределение Пуассона, и пишут* $\xi \sim \Pi_\lambda$, $\lambda > 0$, *если она принимает значения* $0, 1, 2, 3, ..., n, ...$ *с вероятностями*

$$\mathsf{P}(\xi = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k \in \{0, 1, 2, 3, ..., n, ...\}.$$

Проверим, что введенное нами «распределение» действительно является распределением, а именно покажем, что

$$\sum_{k=0}^{\infty} \mathsf{P}(\xi = k) = 1.$$

Имеем,

$$\sum_{k=0}^{\infty} \mathsf{P}(\xi = k) = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} \cdot e^{\lambda} = 1,$$

так как получившийся ряд – не что иное, как ряд Маклорена для экспоненты

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

Замечание 2.3.1 *Распределение Пуассона хорошо описывает так называемые «редкие явления», когда при достаточно большом числе испытаний, которое заранее неизвестно, событие наступает редко, то есть имеет маленькую вероятность. Например, число ДТП в некотором городе относительно всех поездок всех водителей этого города за год, число рекламных сообщений, пришедших на телефон, за месяц, относительно всех сообщений, пришедших за этот месяц, число забитых мячей на чемпионате мира, относительно всех ударов по мячу.*

Можно показать, что математическое ожидание и дисперсия распределения Пуассона одинаковы и равны λ, то есть

$$\mathsf{E}\xi = \mathsf{D}\xi = \lambda.$$

Доказательство этого утверждения в рамках данного курса мы рассматривать не будем.

Пример 2.3.1 *Менеджер телекоммуникационной компании решил рассчитать вероятность того, что в некотором небольшом городе в течение пяти минут поступят $0, 1, 2, \ldots$ звонков. Он выбрал случайные интервалы в 5 минут и подсчитал число вызовов в каждом их из них. Оказалось, что в среднем поступало 4.3 звонка. Давайте вычислим вероятность того, что в течение пяти минут поступит ровно 7 звонков:*

$$\mathsf{P}(\xi = 7) = \frac{4.3^7}{7!} \cdot e^{-4.3} \approx 0.073.$$

Вероятность такого события мала, а значит менеджер может пока не заботиться о расширении компании, или покупке нового оборудования.

3 Функция распределения случайной величины

3.1 Определение и свойства функции распределения

Как мы уже неоднократно сказали, распределение дискретной случайной величины может быть задано таблицей, то есть перечислением значений случайной величины и вероятностей, с которыми эти значения принимаются. Однако, случайные величины, которые не являются дискретными, принципиально невозможно задать аналогичным образом. Существует общий способ задания распределения любой случайной величины. Этот способ опирается на задание так называемой функции распределения.

Определение 3.1.1 *Функцией распределения случайной величины ξ называется функция*

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

Итак, функция распределения случайной величины ξ в точке x показывает вероятность события, что $\xi \in (-\infty, x)$.

Функция распределения $F_\xi(x)$ случайной величины ξ обладает следующими свойствами, которые интуитивно совершенно понятны:

1. $F_\xi(x) \in [0, 1]$

2. $F_\xi(x)$ не убывает, то есть

$$\text{если} \;\; x_1 < x_2, \;\text{то}\;\; F_\xi(x_1) \le F_\xi(x_2).$$

3. $\lim\limits_{x \to -\infty} F_\xi(x) = 0.$

4. $\lim\limits_{x\to+\infty} F_\xi(x) = 1.$

Первое свойство следует из того, что функция распределения – это некоторая вероятность, и, так как вероятность находится в диапазоне от 0 до 1, то и значения функции распределения лежат там же. Второе свойство – это следствие монотонности вероятности, ведь при $x_1 < x_2$ событие $\{\xi < x_1\}$ влечет событие $\{\xi < x_2\}$. Последние два свойства опираются на то, что ξ – это случайная величина, то есть функция, а значит ее значения лежат в множестве $\mathbb{R} = (-\infty, +\infty)$. Так, третье свойство, грубо говоря, означает, что мы вычисляем вероятность события $\{\xi < -\infty\}$, что, конечно, событие невозможное, и его вероятность равна нулю. В последнем свойстве, наоборот, мы вычисляем вероятность события $\{\xi < +\infty\}$, что является достоверным событием, а значит и его вероятность равна 1.

Чем же нам помогает функция распределения?

Замечание 3.1.1 *Оказывается, что знание функции распределения позволяет найти вероятность попадания значений случайной величины в заданный промежуток, а именно:*

$$\mathsf{P}(a \le \xi < b) = F_\xi(b) - F_\xi(a),$$

$$\mathsf{P}(a < \xi \le b) = F_\xi(b+0) - F_\xi(a+0),$$
$$\mathsf{P}(a \le \xi \le b) = F_\xi(b+0) - F_\xi(a),$$
$$\mathsf{P}(a < \xi < b) = F_\xi(b) - F_\xi(a+0),$$

где в формулах $F_\xi(x+0)$ означает предел функции F_ξ в точке x справа. В качестве a, b могут выступать и бесконечности (соответствующих знаков).

Рассмотрим пример, в котором построим функцию распределения и вычислим с ее помощью интересующую нас вероятность.

Пример 3.1.1 *В лотерее на каждые* 100 *билетов в среднем приходится* 15 *выигрышных. Данные о количестве билетов и размере выигрышей (в рублях) приведены в таблице*

размер выигрыша	0	100	500	2000
количество билетов	85	10	4	1

Пусть ξ – случайная величина, показывающая размер выигрыша на один случайно выбранный билет. Тогда ее распределение задается следующей таблицей

ξ	0	100	500	2000
P	0.85	0.1	0.04	0.01

Построим функцию распределения $F_\xi(x) = \mathsf{P}(\xi < x)$ *данной случайной величины. Ясно, что ключевыми точками построения являются точки-значения случайной величины:* $0,\ 100,\ 500,\ 2000$. *Тогда*

$$F_\xi(x) = \begin{cases} 0, & x \le 0 \\ 0.85, & 0 < x \le 100 \\ 0.95, & 100 < x \le 500 \\ 0.99, & 500 < x \le 2000 \\ 1, & x > 2000 \end{cases}.$$

График функции распределения представлен на рисунке 1. Как получилась

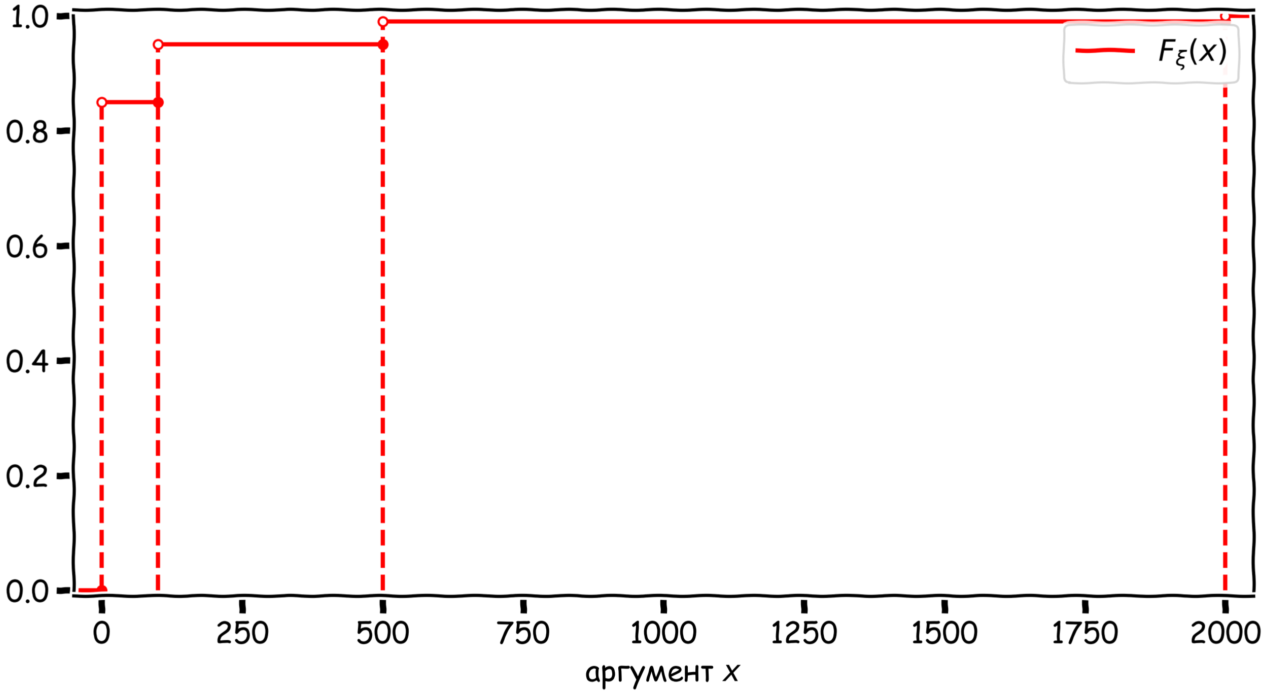

Рис. 1: Функция распределения $F_\xi(t)$

такая функция распределения? Ну смотрите, что происходит при $x \le 0$? *Нас интересует вероятность события, что случайная величина меньше, чем* x. *Но наша случайная величина* ξ *не принимает отрицательных значений, а значит вероятность события* $\{\xi < x,\ x \le 0\}$ *равна* 0, *тогда и значение функции распределения равно* 0.

Теперь рассмотрим промежуток $0 < x \le 100$. *При каждом* x *из такого промежутка событие* $\{\xi < x\}$ *состоит ровно из одного значения случайной величины, нулевого, его вероятность равна* 0.85, *что и записано в функции распределения.*

В заключении ответим на вопрос: какова вероятность события, что *случайный билет принесет выигрыш более* 100 *рублей?*

$$\mathsf{P}(\xi > 100) = \sum_{i:\ a_i > 100} \mathsf{P}(\xi = a_i)$$

Так как значения случайной величины, превышающие 100 *– это* 500 *и* 2000, *то последняя сумма состоит из двух слагаемых:*

$$\mathsf{P}(\xi > 100) = \mathsf{P}(\xi = 500) + \mathsf{P}(\xi = 2000) = 0.04 + 0.01 = 0.05.$$

Значит, искомая вероятность равна 0.05.

Используя функцию распределения, ту же самую вероятность можно вычислить, как

$$\mathsf{P}(\xi > 100) = \mathsf{P}(\xi \in (100, +\infty)) = F_\xi(+\infty) - F_\xi(100+0) = 1 - 0.95 = 0.05.$$

3.2 Плотность распределения непрерывной случайной величины

До сих пор мы рассматривали исключительно дискретные случайные величины, и у слушателя может возникнуть неверное представление о том, что других распределений в природе нет. На самом деле, дискретное распределение – это одна из двух крайностей в распределениях. Другая крайность – это так называемое непрерывное распределение, которое мы и рассмотрим.

Определение 3.2.1 *Говорят, что случайная величина* ξ *имеет непрерывное распределение, если существует такая неотрицательная функция* $f_\xi(x) : \mathbb{R} \to \mathbb{R}$, *что*

$$F_\xi(x) = \mathsf{P}(\xi < x) = \int\limits_{-\infty}^{x} f_\xi(t)dt.$$

Определение 3.2.2 *Функция* $f_\xi(x)$ *называется плотностью вероятности случайной величины* ξ.

Проведем аналогию со случаем дискретной случайной величины. Ясно, что непрерывная величина получена в некотором смысле «предельным переходом» от дискретной, что особенно хорошо видно из рассмотрения пары соотношений (сумма меняется на интеграл):

$$F_\xi(x) = \mathsf{P}(\xi < x) = \int\limits_{-\infty}^{x} f_\xi(t)dt \quad \leftrightarrow \quad F_\xi(x) = \mathsf{P}(\xi < x) = \sum_{i:\ a_i < x} \mathsf{P}(\xi = a_i).$$

Кроме того, так как $\lim\limits_{x\to+\infty} F_\xi(x) = 1$, то $\int\limits_{-\infty}^{+\infty} f_\xi(t)dt = 1$. В дискретном случае было равенство вида $\sum\limits_i \mathsf{P}(\xi = a_i) = 1$. Тем самым, мы получаем пару «связанных» соотношений

$$\int\limits_{-\infty}^{+\infty} f_\xi(t)dt = 1 \quad \leftrightarrow \quad \sum_i \mathsf{P}(\xi = a_i) = 1.$$

Геометрический смысл плотности заключается в следующем: пусть f_ξ – плотность некоторой случайной величины ξ (рисунок 2). Вся площадь под графиком, как мы уже отметили выше, равна $\int\limits_{-\infty}^{+\infty} f_\xi(t)dt = 1$. Так как геометрический смысл определенного интеграла на промежутку I от неотрицательной функции – это площадь под графиком этой функции над промежутком I, то ясно, что

$$\mathsf{P}(a \leq \xi < b) = \int\limits_a^b f_\xi(t)dt = S_1,$$

$$\mathsf{P}(\xi < c) = \int\limits_{-\infty}^{c} f_\xi(t)dt = S_2.$$

Итак, у непрерывной случайной величины вероятность попасть в некоторое множество $A \subset \mathbb{R}$ – это площадь под графиком плотности над множеством A.

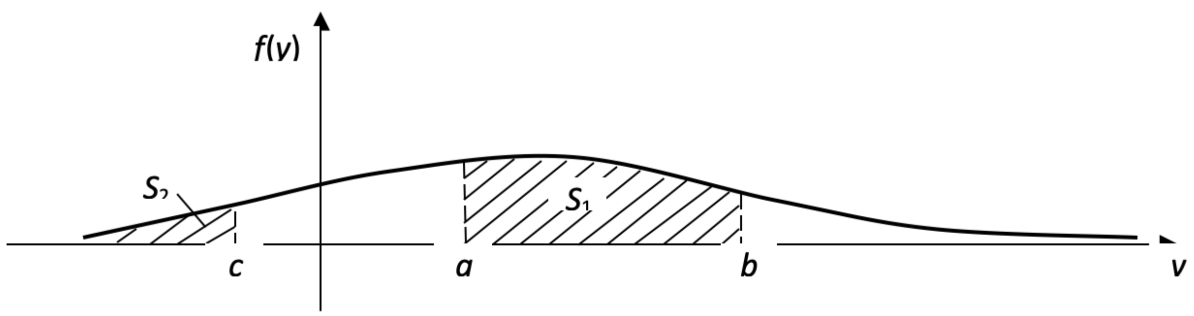

Рис. 2: Геометрический смысл плотности распределения

4 Числовые характеристики непрерывных случайных величин

Числовые характеристики дискретных случайных величин мы уже рассматривали. Введём теперь эти характеристики для непрерывных случайных величин.

4.1 Математическое ожидание

Определение 4.1.1 *Пусть случайная величина ξ имеет непрерывное распределение с плотностью $f_\xi(x)$. Математическим ожиданием случайной величины ξ называется число*

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

при условии, что написанный интеграл сходится абсолютно (то есть если существует $\mathsf{E}|\xi|$). Иначе говорят, что математического ожидания не существует.

Пример 4.1.1 *Пусть плотность случайной величины ξ имеет вид:*

$$f_\xi(t) = \begin{cases} 0, & t \le 0 \\ \frac{t}{2}, & 0 < t \le 2 \\ 0, & t > 2 \end{cases}.$$

Вычислим математическое ожидание, используя определение:

$$\mathsf{E}\xi = \int\limits_{-\infty}^{+\infty} t \cdot f_\xi(t) dt = \int\limits_{-\infty}^{0} t \cdot 0 dt + \int\limits_{0}^{2} \frac{1}{2} t^2 dt + \int\limits_{2}^{+\infty} t \cdot 0 dt = \frac{1}{2} \cdot \frac{2^3}{3} = \frac{4}{3}.$$

Оказывается, для математического ожидания справедливо свойство, которое будет очень полезно в дальнейшем.

Пусть $g(\xi)$ – случайная величина, построенная по случайной величине ξ. Тогда если ξ имеет непрерывное распределение, а математическое ожидание $g(\xi)$ определено, то

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

4.2 Дисперсия непрерывной случайной величины

Дисперсия любой случайной величины (в том числе непрерывной), согласно определению, равна

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

если существует $\mathsf{E}\xi^2$. Как уже ранее отмечалось, зачастую дисперсию удобнее вычислять, используя формулу

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

Это соотношение немедленно следует из свойств математического ожидания, а именно

$$\mathsf{E}\left(\xi - \mathsf{E}\xi\right)^2 = \mathsf{E}\left(\xi^2 - 2\xi\mathsf{E}\xi + (\mathsf{E}\xi)^2\right) = \mathsf{E}\xi^2 - 2\left(\mathsf{E}\xi\right)^2 + \left(\mathsf{E}\xi\right)^2 = \mathsf{E}\xi^2 - \left(\mathsf{E}\xi\right)^2,$$

так как математическое ожидание линейно, $\mathsf{E}\xi$ – константа, константа выносится за знак математического ожидания, а значит $\mathsf{E}\left(2\xi\mathsf{E}\xi\right) = 2\mathsf{E}\xi\mathsf{E}\xi = 2\left(\mathsf{E}\xi\right)^2$ и $\mathsf{E}\left(\mathsf{E}\xi\right)^2 = \left(\mathsf{E}\xi\right)^2$.

С учетом сделанного выше утверждения о вычислении математического ожидания функции от случайной величины, дисперсия непрерывной случайной величины (если она существует) будет выражаться через плотность распределения следующим образом:

$$\mathsf{D}\xi = \int\limits_{-\infty}^{+\infty} x^2 f_\xi(x)dx - \left(\int\limits_{-\infty}^{+\infty} x f_\xi(x)dx\right)^2 .$$

Замечание 4.2.1 *Для непрерывной случайной величины математическое ожидание и дисперсия обладают теми же свойствами, что и для дискретной случайной величины.*

Пример 4.2.1 *Рассмотрим случайную величину ξ из предыдущего примера:*

$$f_\xi(t) = \begin{cases} 0, & t \le 0 \\ \frac{t}{2}, & 0 < t \le 2 \\ 0, & t > 2 \end{cases}.$$

Мы уже нашли ее математическое ожидание $\mathsf{E}\xi = \frac{4}{3}$. Найдем ее дисперсию.

$$\mathsf{E}\xi^2 = \int\limits_{-\infty}^{+\infty} t^2 f_\xi(t)dt = \int\limits_{-\infty}^{0} t^2 \cdot 0dt + \int\limits_{0}^{2} t^2 \cdot \frac{t}{2}dt + \int\limits_{2}^{+\infty} t^2 \cdot 0dt = \frac{1}{2} \cdot \frac{2^4}{4} = 2.$$

Тогда

$$\mathsf{D}\xi = \mathsf{E}\xi^2 - (\mathsf{E}\xi)^2 = 2 - \left(\frac{4}{3}\right)^2 = \frac{2}{9}.$$

5 Основные законы распределения непрерывных случайных величин

Аналогично тому, как было сделано в дискретном случае, рассмотрим наиболее часто встречающиеся непрерывные распределения.

5.1 Равномерное распределение

Определение 5.1.1 *Говорят, что случайная величина ξ имеет равномерное распределение на отрезке $[a, b]$ и пишут $\xi \sim \mathsf{U}_{a,b}$, если ее плотность имеет вид*

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

График плотности представлен на рисунке 3.

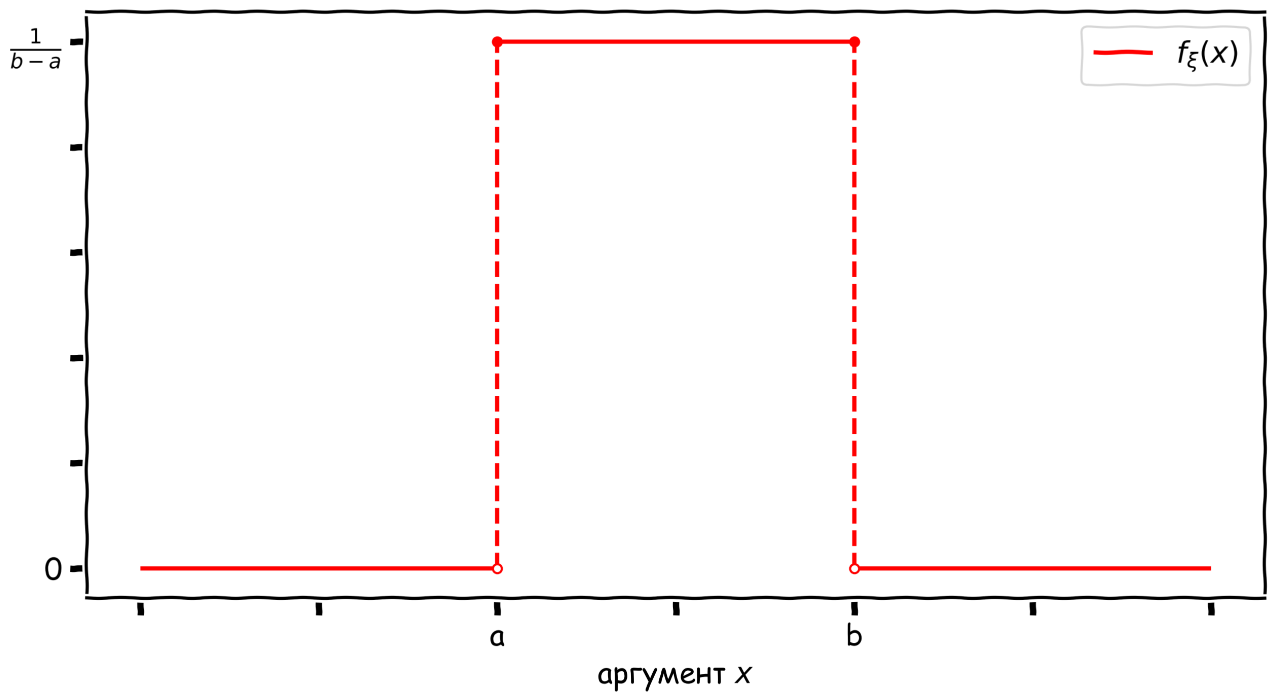

Рис. 3: Функция распределения случайной величины, имеющей равномерное распределение

Если говорить о смысле такого распределения, то можно считать, что рассматриваемая случайная величина есть не что иное, как координата точки, случайно брошенной на отрезок $[a, b]$. Плотность такой величины одинаково «размазана» по отрезку, поэтому и попадание в любую точку этого отрезка «равновозможно».

График функции распределения показан на рисунке 4.

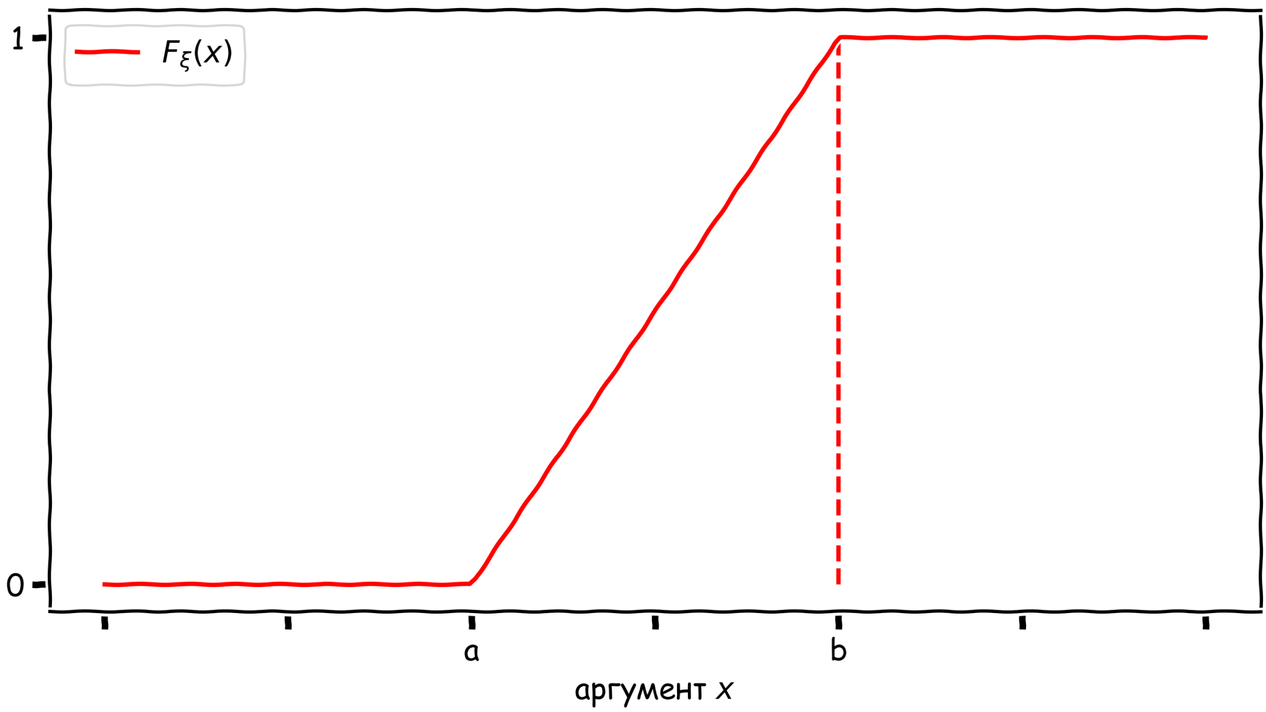

Рис. 4: Функция распределения случайной величины, имеющей равномерное распределение

Замечание 5.1.1 *Можно провести аналогию рассматриваемого распределения и геометрической вероятности. Если случайная величина ξ равномерно распределена на отрезке $[a, b]$, то вероятность попасть в интервал $A = (c, d) \subset [a, b]$ пропорциональна длине этого интервала, и равна*

$$\mathsf{P}(\cdot \in A) = \frac{\lambda(A)}{\lambda([a, b])} = \frac{d - c}{b - a}.$$

Пример 5.1.1 *Однородную нить длиной* 1 *метр тянут за концы, и происходит разрыв в случайной точке $\xi \in [0, 1]$. Найти вероятность того, что разрыв произойдет в интервале от* 10 *до* 12 *см. Так как нить однородная, можно считать, что случайная величина ξ имеет равномерное распределение на отрезке $[0, 1]$. В данном случае $a = 0$, $b = 1$, $c = 0.1$ и $d = 0.12$. Тогда*

$$\mathsf{P}(\xi \in (0.1, 0.12)) = \frac{d - c}{b - a} = \frac{0.12 - 0.1}{1} = 0.02.$$

Математическое ожидание равномерно распределенной величины ξ равно

$$\mathsf{E}\xi = \int\limits_{-\infty}^{+\infty} x f_\xi(x) dx = \int\limits_a^b \frac{x}{b-a} dx = \frac{b^2 - a^2}{2(b-a)} = \frac{a+b}{2}$$

Для вычисления дисперсии найдем $\mathsf{E}\xi^2$.

$$\mathsf{E}\xi^2 = \int\limits_{-\infty}^{+\infty} x^2 f_\xi(x) dx = \int\limits_a^b \frac{x^2}{b-a} dx = \frac{b^2 + ab + a^2}{3}.$$

Учитывая, что $\mathsf{E}\xi = \frac{a+b}{2}$, получим:

$$\mathsf{D}\xi = \mathsf{E}\xi^2 - (\mathsf{E}\xi)^2 = \frac{b^2 + ab + a^2}{3} - \frac{(a+b)^2}{4} = \frac{(b-a)^2}{12}.$$

5.2 Показательное распределение

Следующее распределение в некотором смысле является непрерывным аналогом распределения Пуассона.

Определение 5.2.1 *Говорят, что случайная величина ξ имеет показательное распределение с параметром $\lambda > 0$ и пишут $\xi \sim \mathsf{Exp}_\lambda$, если ее плотность имеет вид*

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

График плотности показательного распределения представлен на рисунке 5.

Функция распределения случайной величины ξ легко вычисляется и задается соотношением

$$F_\xi(x) = \begin{cases} 0, & x \leq 0 \\ 1 - e^{-\lambda x}, & x > 0 \end{cases}.$$

График последней функции представлен на рисунке 6. Оказывается, что длительности телефонных разговоров, промежутки времени между последовательными приходами клиентов на обслуживание, длительности обслуживания клиентов, время безотказной работы прибора и многое другое имеют показательное распределение.

Для случайной величины $\xi \sim \mathsf{Exp}_\lambda$ можно показать, что

$$\mathsf{E}\xi = \frac{1}{\lambda}, \quad \mathsf{D}\xi = \frac{1}{\lambda^2}.$$

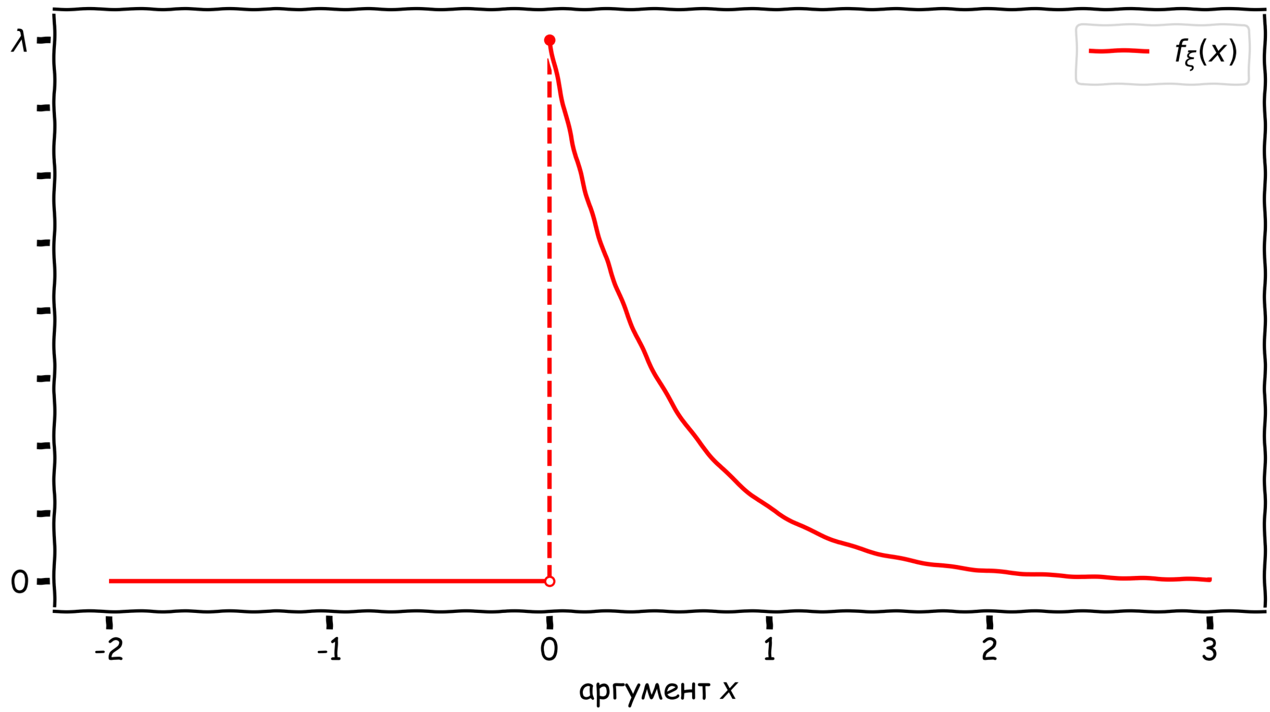

Рис. 5: Плотность случайной величины, имеющей показательное распределение

Пример 5.2.1 *Пусть ξ – случайная величина, показывающая время обслуживания (в минутах) покупателя в магазине, имеющая показательное распределение. Какова вероятность, что время, затраченное на обслуживание покупателя, будет находиться в интервале от 2 до 4 минут, если среднее время обслуживания составляет 2 минуты? Исходя из условия задачи:*

$$\mathsf{E}\xi = \frac{1}{\lambda} = 2 \Rightarrow \lambda = \frac{1}{2}.$$

Функция распределения случайно величины $\xi \sim \mathsf{Exp}_{\frac{1}{2}}$ для всех $x \geq 0$ имеет вид:

$$F_\xi(x) = 1 - e^{-\frac{x}{2}},$$

тогда

$$\mathsf{P}(2 \leq \xi \leq 4) = F_\xi(4) - F_\xi(2) = 1 - e^{-2} - \left(1 - e^{-1}\right) = e^{-1} - e^{-2} \approx 0.23$$

5.3 Нормальное распределение

Следующий пример распределения является одним из важнейших.

Определение 5.3.1 *Говорят, что случайная величина ξ имеет нормальное (гауссовское) распределение с параметрами $a \in \mathbb{R}$, σ^2, и пишут $\xi \sim$*

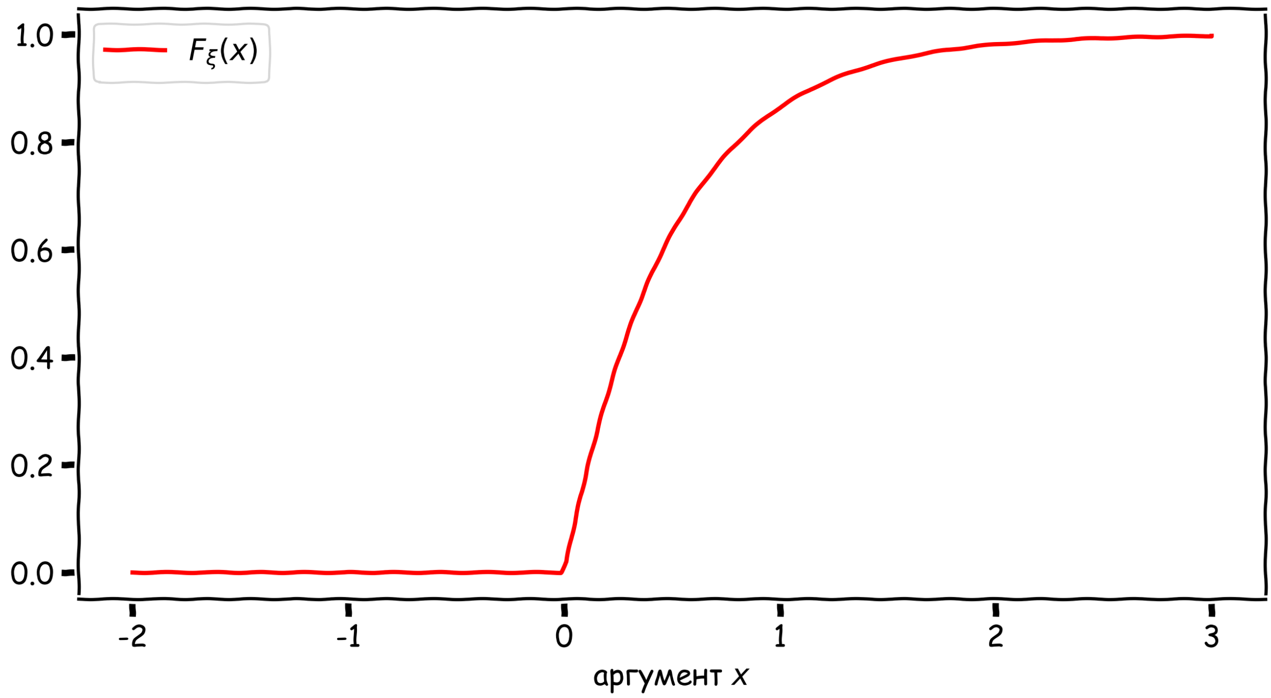

Рис. 6: Функция распределения случайной величины, имеющей показательное распределение

N_{a,σ^2}, *если ее плотность имеет вид*

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

График плотности нормального распределения при разных значениях a, σ^2 представлен на рисунке 7. Нормальное распределение используется очень часто – в нашем мире очень многое «нормально».

Определение 5.3.2 *Нормальное распределение с параметрами* $a = 0$, $\sigma^2 = 1$, *то есть распределение* $\mathsf{N}_{0,1}$, *называют стандартным нормальным.*

Ясно, что плотность стандартного нормального распределения имеет вид

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

Функция распределения нормального распределения, ввиду важности последнего, обозначается особым образом:

$$F_\xi(x) = \Phi_{a,\sigma^2} = \frac{1}{\sigma\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{-\frac{(t-a)^2}{2\sigma^2}} dt,$$

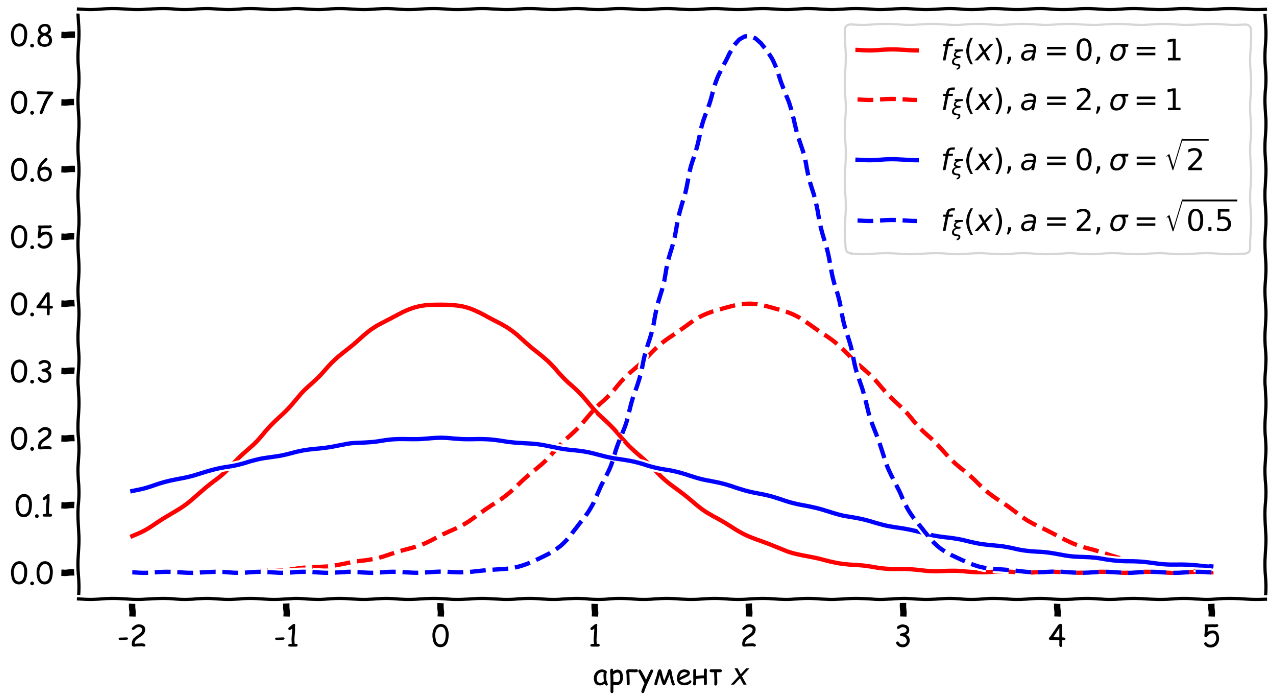

Рис. 7: Плотность случайной величины, имеющей нормальное распределение

не выражается в элементарных функциях и затабулирована (ее значения можно найти в таблицах). Точнее, затабулирована функция распределения стандартного нормального распределения, но с помощью замены переменной в интеграле легко убедиться, что

$$\Phi_{a,\sigma^2}(x) = \Phi_{0,1}\left(\frac{x-a}{\sigma}\right).$$

График функции распределения нормального закона представлен на рисунке 8. Отметим несколько важных свойств функции распределения

Лемма 5.3.1 *Функция распределения стандартного нормального распределения обладает следующими свойствами:*

1. $\Phi_{0,1}(0) = \frac{1}{2}$;
2. $\Phi_{0,1}(-x) = 1 - \Phi_{0,1}(x)$;

Как и в случае с показательным распределением примем без доказательства тот факт, что если $\xi \sim \mathsf{N}_{a,\sigma^2}$, то

$$\mathsf{E}\xi = a, \quad \mathsf{D}\xi = \sigma^2.$$

Пример 5.3.1 *Случайная величина ξ распределена по нормальному закону с параметрами $a = 2.5$, $\sigma = 2$. Определить $\mathsf{P}(|\xi| \le 3)$.*

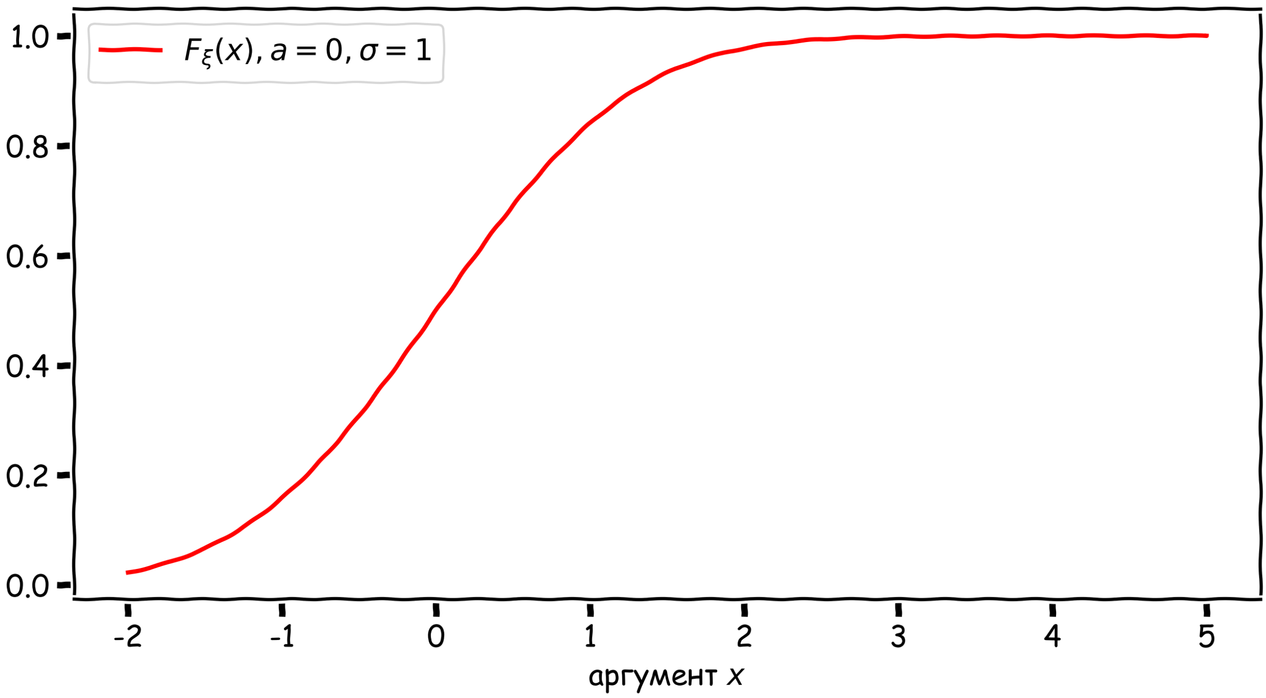

Рис. 8: Функция распределения случайной величины, имеющей стандартное нормальное распределение

Вероятность попадания случайной величины в отрезок $[-3, 3]$ *можно найти следующим образом:*

$$\begin{gathered}\mathsf{P}(|\xi| \leq 3) = \mathsf{P}(-3 \leq \xi \leq 3) = \\ = \Phi_{0,1}\left(\frac{3-2.5}{2}\right) - \Phi_{0,1}\left(\frac{-3-2.5}{2}\right) = \\ = \Phi_{0,1}\left(0.25\right) - \Phi_{0,1}\left(-2.75\right) \approx 0.596.\end{gathered}$$

Для случайной величины ξ, имеющей нормальное распределение N_{a,σ^2}, справедливо так называемое правило 3σ, которое гласит, что

$$\mathsf{P}(|\xi - a| < 3\sigma) = \mathsf{P}(\xi \in (a - 3\sigma, a + 3\sigma)) = 0.9972.$$

Конкретное значение помнить не обязательно, но важно понимать, что в случае, когда случайная величина имеет нормальное распределение N_{a,σ^2}, «практически вся плотность» расположена от среднего a на расстоянии 3σ.

6 Совместное распределение двух случайных величин

В реальных задачах часто приходится рассматривать не одну, а несколько случайных величин. Достаточно естественно, что распределение одной из них как правило зависит от распределения других. В этом разделе мы изучим вопросы, связанные с законом распределения векторных случайных величин, обсудим характер возможной зависимости координат случайного вектора, а также способы определения меры этой зависимости. В рамках нашего курса мы будем рассматривать только двумерные дискретные величины.

6.1 Совместное распределение двух дискретных случайных величин

Определение 6.1.1 *Совместным распределением случайных величин ξ_1 и ξ_2 называется набор вероятностей $\mathsf{P}(\xi_1 = a, \xi_2 = b)$, где числа a пробегают всевозможные значения $a_1, ..., a_n$ случайной величины ξ_1, а числа b – всевозможные значения $b_1, ..., b_k$ случайной величины ξ_2, причем*

$$\sum_{i=1}^{n}\sum_{j=1}^{k}\mathsf{P}(\xi_1 = a_i, \xi_2 = b_j) = 1.$$

Пусть случайная величина ξ_1 принимает значения $a_1, ..., a_n$, а случайная величина ξ_2 принимает значения $b_1, ..., b_k$. Совместное распределение случайных величин часто записывают в виде таблицы

$\xi_1 \setminus \xi_2$	b_1	b_2	...	b_k
a_1	$\mathsf{P}(\xi_1 = a_1, \xi_2 = b_1)$	$\mathsf{P}(\xi_1 = a_1, \xi_2 = b_2)$	...	$\mathsf{P}(\xi_1 = a_1, \xi_2 = b_k)$
a_2	$\mathsf{P}(\xi_1 = a_2, \xi_2 = b_1)$	$\mathsf{P}(\xi_1 = a_2, \xi_2 = b_2)$	...	$\mathsf{P}(\xi_1 = a_2, \xi_2 = b_k)$
...	...	...	...	...
a_n	$\mathsf{P}(\xi_1 = a_n, \xi_2 = b_1)$	$\mathsf{P}(\xi_1 = a_n, \xi_2 = b_2)$	...	$\mathsf{P}(\xi_1 = a_n, \xi_2 = b_k)$

Зная совместное распределение, можно восстановить так называемые маргинальные (они же – просто обычные одномерные) распределения случайных величин ξ_1 и ξ_2 по правилам

$$\mathsf{P}(\xi_1 = a_i) = \sum_{j=1}^{k}\mathsf{P}(\xi_1 = a_i, \xi_2 = b_j),\ i \in \{1, ..., n\},$$

$$\mathsf{P}(\xi_2 = b_j) = \sum_{i=1}^{n}\mathsf{P}(\xi_1 = a_i, \xi_2 = b_j),\ j \in \{1, ..., k\}.$$

Пример 6.1.1 *Дана таблица совместного распределения двух случайных величин ξ_1 и ξ_2:*

$\xi_2 \setminus \xi_1$	-1	-2	5
3	0.2	0.2	0
5	0.1	0.05	0.05
7	0.05	0.1	0.25

Составим таблицы маргинальных распределений.

Для нахождения маргинальных распределений просуммируем вероятности по столбцам для случайной величины ξ_1, и по строкам для случайной величины ξ_2. Например, для $\xi_1 = -1$ сложим вероятности, находящиеся в столбце, отвечающем значению -1 случайной величины ξ_1 и получим

$$0.2 + 0.1 + 0.05 = 0.35.$$

Для $\xi_2 = 7$ произведем аналогичную процедуру, но со строкой:

$$0.05 + 0.1 + 0.25 = 0.4.$$

Произведя такие операции со строками и столбцами, получим маргинальные распределения, задаваемые следующими таблицами:

ξ_1	-1	-2	5
P	0.35	0.35	0.3

,

ξ_2	3	5	7
P	0.4	0.2	0.4

.

Совместное распределение как раз-таки и помогает понять: зависимы случайные величины или нет, ведь именно благодаря ему можно определить выполняется равенство

$$\mathsf{P}(\xi_1 = a, \xi_2 = b) = \mathsf{P}(\xi_1 = a)\mathsf{P}(\xi_2 = b),$$

или нет.

Пример 6.1.2 *Проверим, являются ли независимыми случайные величины ξ_1 и ξ_2 из предыдущего примера. Независимость устанавливается в том случае, если равенство*

$$\mathsf{P}(\xi_1 = a, \xi_2 = b) = \mathsf{P}(\xi_1 = a)\mathsf{P}(\xi_2 = b),$$

выполняется для всех a и b. В противном случае случайные величины зависимы.

Таблицы совместного и маргинальных распределений имеют вид:

$\xi_2 \backslash \xi_1$	-1	-2	5
3	0.2	0.2	0
5	0.1	0.05	0.05
7	0.05	0.1	0.25

,

ξ_1	-1	-2	5
P	0.35	0.35	0.3

,

ξ_2	3	5	7
P	0.4	0.2	0.4

.

Тогда

$$\mathsf{P}(\xi_1 = -2, \xi_2 = 7) = 0.1.$$

С другой стороны

$$\mathsf{P}(\xi_1 = -2)\mathsf{P}(\xi_2 = 7) = 0.35 \cdot 0.4 = 0.14.$$

Так как $0.1 \neq 0.14$, *можно сделать вывод, что случайные величины зависимы.*

Отметим еще одну важную роль совместного распределения. Зная совместное распределение случайных величин, мы можем написать распределение различных функций от этих случайных величин, в частности распределение суммы, разности или произведения. Знание маргинальных распределений не позволяет этого сделать, как показывает следующий пример.

Пример 6.1.3 *Пусть задано совместное распределение случайных величин* ξ_1 *и* ξ_2 *следующей таблицей* ($r \in [0, 0.5]$)

$\xi_1 \setminus \xi_2$	0	1
0	r	$\frac{1}{2} - r$
1	$\frac{1}{2} - r$	r

.

Маргинальные распределения у случайных величин ξ_1 *и* ξ_2 *одинаковы и не зависят от* r:

ξ_1	0	1
P	$\frac{1}{2}$	$\frac{1}{2}$

.

ξ_2	0	1
P	$\frac{1}{2}$	$\frac{1}{2}$

.

Найдем распределение случайной величины $\xi_1 + \xi_2$. *Ясно, что сумма может принимать значения* $0, 1, 2$, *причем*

$$\mathsf{P}(\xi_1 + \xi_2 = 0) = \mathsf{P}(\xi_1 = 0, \xi_2 = 0) = r,$$

$$\mathsf{P}(\xi_1 + \xi_2 = 1) = \mathsf{P}((\xi_1 = 1, \xi_2 = 0) \cup (\xi_1 = 0, \xi_2 = 1)) =$$

$$\mathsf{P}(\xi_1 = 1, \xi_2 = 0) + \mathsf{P}(\xi_1 = 0, \xi_2 = 1) = 1 - 2r$$

и

$$\mathsf{P}(\xi_1 + \xi_2 = 2) = \mathsf{P}(\xi_1 = 1, \xi_2 = 1) = r.$$

Тем самым,

$\xi_1 + \xi_2$	0	1	2
P	r	$1 - 2r$	r

Видно, что распределение зависит от r *при неизменных маргинальных распределениях.*

6.2 Ковариация. Коэффициент корреляции

Обычно у рассматриваемого явления есть не один, а несколько признаков. Например, у человека есть рост, вес, возраст, уровень дохода, регион проживания и пр. У автомобиля максимальная скорость, пробег, объем двигателя и пр. В таком случае может возникнуть вопрос – есть ли какая-то связь между величинами? То есть можем ли мы сказать, что изменение одного признака скорее всего повлечет изменение другого? Например, если мы будем съедать по пять сдобных круассанов в день, то наш вес наверняка увеличится. Или стал расти курс доллара – жди изменений цены на евро. Снизилась ставка ипотечного кредитования – повысился спрос на квартиры.

Когда мы рассматриваем математические функции, то наблюдаем прямую зависимость значения функции от аргумента. Например, $y = 2x + 3$. Тут каждому значению x соответствует одно строго определенное значение y. Оперируя со случайными величинами, чаще всего такую прямую функциональную зависимость построить не удается. Мы можем рассмотреть корреляционную зависимость - это вероятностная зависимость между величинами, которая возникает тогда, когда одна из величин зависит не только от заданной второй величины, но, возможно, и еще от других случайных условий.

Зададимся подробнее математическим вопросом о зависимости или независимости двух случайных величин. Пусть ξ_1 и ξ_2 – дискретные случайные величины со значениями $a_1, ..., a_n$ и $b_1, ..., b_k$, соответсвенно, и известно их совместное распределение, тогда

$$\mathsf{E}(\xi_1\xi_2) = \sum_{i=1}^{n}\sum_{j=1}^{k} a_i b_j \mathsf{P}(\xi_1 = a_i, \xi_2 = b_j).$$

Если предположить, что ξ_1 и ξ_2 независимы, то

$$\mathsf{P}(\xi_1 = a_i, \xi_2 = b_j) = \mathsf{P}(\xi_1 = a_i)\mathsf{P}(\xi_2 = b_j),$$

а значит

$$\mathsf{E}(\xi_1\xi_2) = \sum_{i=1}^{n}\sum_{j=1}^{k} a_i b_j \mathsf{P}(\xi_1 = a_i, \xi_2 = b_j) = \sum_{i=1}^{n}\sum_{j=1}^{k} a_i b_k \mathsf{P}(\xi_1 = a_i)\mathsf{P}(\xi_2 = b_j) =$$

$$= \sum_{i=1}^{n} a_i \mathsf{P}(\xi_1 = a_i) \cdot \sum_{j=1}^{k} b_j \mathsf{P}(\xi_2 = b_j) = \mathsf{E}\xi_1 \cdot \mathsf{E}\xi_2.$$

Итого оказывается, что если случайные величины ξ_1 и ξ_2 независимы, то это отражается в том алгебраическом свойстве, что

$$\mathsf{E}(\xi_1\xi_2) = \mathsf{E}\xi_1\mathsf{E}\xi_2,$$

а если это равенство не выполнено, то величины оказываются заведомо зависимыми.

Определение 6.2.1 *Величина* $\mathsf{E}(\xi_1\xi_2) - \mathsf{E}\xi_1\mathsf{E}\xi_2$ *называется ковариацией случайных величин* ξ_1 *и* ξ_2 *и обозначается* $\mathsf{cov}(\xi_1, \xi_2)$.

С ковариацией есть одна проблема – она измеряется в квадратах единиц измерения случайных величин ξ_1 и ξ_2. Это плохо, ведь увеличение значений одной случайной величины в, скажем, 100 раз, влечет и увеличение ковариации в такое же количество раз, однако «сила» зависимости не меняется. Тут приходит на помощь коэффициент корреляции.

Определение 6.2.2 *Коэффициентом корреляции двух величин* ξ_1 *и* ξ_2 *с отличными от нуля дисперсиями, называется величина*

$$\rho(\xi_1, \xi_2) = \frac{\mathsf{cov}(\xi_1, \xi_2)}{\sqrt{\mathsf{D}\xi_1}\sqrt{\mathsf{D}\xi_2}} = \frac{\mathsf{cov}(\xi_1, \xi_2)}{\sigma_{\xi_1}\sigma_{\xi_2}}$$

Коэффициент корреляции обладает следующими свойствами:

1. Его абсолютное значение не превосходит единицы, то есть $|\rho(\xi_1, \xi_2)| \leq 1$;

2. Если ξ_1, ξ_2 – независимы, то $\rho(\xi_1, \xi_2) = 0$;

3. $|\rho(\xi_1, \xi_2)| = 1$ тогда и только тогда, когда $\xi_1 = a\xi_2 + b$, причем $a \cdot \rho(\xi_1, \xi_2) > 0$.

Отметим, что коэффициент корреляции показывает степень линейной зависимости случайных величин. Чем он ближе по модулю к единице, тем более зависимость близка к линейной.

Так, если он равен 1, то увеличение значений одной случайной величины гарантированно влечет увеличение другой, а если -1, то увеличение одной влечет уменьшение другой.

В то же время к трактовке «зависимости» двух случайных величин в статистике, когда корреляция не равна единице, нужно подходить с большой осторожностью. Предположим, что у нас есть данные о силе наводнения и количестве привлеченных спасателей службы МЧС. Ясно, что эти показатели коррелируют, причем коэффициент корреляции положительный. Однако из этого вовсе не следует, что увеличение числа привлеченных спасателей свидетельствует о каком-то ужасном бедствии. И уж тем более это не значит, что увольнение всех спасателей приведет к отсутствию наводнений и бедствий, с ними связанных.

Кроме того, равенство нулю корреляции не свидетельствует об отсутствии зависимости.

Пример 6.2.1 *Покажем, что условие*

$$\mathsf{E}(\xi_1\xi_2) = \mathsf{E}\xi_1\mathsf{E}\xi_2$$

не является достаточным для независимости случайных величин. Для этого предположим, что пространство элементарных исходов Ω *имеет вид* $\Omega = \{-\frac{\pi}{2}, 0, \frac{\pi}{2}\}$*, и на нем заданы две заведомо зависимые случайные величины* $\xi_1 = \sin\omega$ *и* $\xi_2 = \cos\omega$*. Запишем законы распределения случайных величин:*

ξ_1	-1	0	1
P	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

ξ_2	0	1
P	$\frac{2}{3}$	$\frac{1}{3}$

Ясно, что $\mathsf{E}\xi_1 = 0$, $\mathsf{E}\xi_2 = \frac{1}{3}$*, значит* $\mathsf{E}\xi_1\mathsf{E}\xi_2 = 0$*. Кроме того, на рассматриваемом пространстве* Ω *всегда* $\xi_1\xi_2 = 0$*, значит и* $\mathsf{E}(\xi_1\xi_2) = 0$*. Тем самым условие*

$$\mathsf{E}(\xi_1\xi_2) = \mathsf{E}\xi_1\mathsf{E}\xi_2$$

выполнено, но величины зависимы.

Пример 6.2.2 *Случайные величины* ξ_1 *и* ξ_2 *заданы таблицей совместного распределения. Определить ковариацию и коэффициент корреляции. Ответ округлить до сотых.*

$\xi_2 \backslash \xi_1$	2	3	5
-1	0.1	0.3	0.2
1	0.1	0.05	0
4	0	0.15	0.1

Ковариация двух случайных величин находится как:

$$\mathsf{cov}(\xi_1, \xi_2) = \mathsf{E}(\xi_1\xi_2) - \mathsf{E}\xi_1\mathsf{E}\xi_2.$$

Составим таблицу распределения для случайной величины $\xi_1\xi_2$. Для этого необходимо рассмотреть все возможные произведения значений случайных величин. Вероятность произведения для конкретного случая будет находиться на пересечении соответствующего столбца и строки. Если значения произведения получаются одинаковыми (в нашем случае такого нет), то необходимо сложить соответствующие вероятности. Значения, для которых вероятность равна 0, в таблицу можно не заносить

$\xi_1\xi_2$	-5	-3	-2	2	3	12	20
P	0.2	0.3	0.1	0.1	0.05	0.15	0.1

.

Составим таблицы маргинальных распределений ξ_1 и ξ_2.

ξ_1	2	3	5
P	0.2	0.5	0.3

,

ξ_2	-1	1	4
P	0.6	0.15	0.25

Найдем соответствующие математические ожидания:

$$\mathsf{E}(\xi_1\xi_2) = 2.05$$
$$\mathsf{E}\xi_1 = 3.4$$
$$\mathsf{E}\xi_2 = 0.55$$

Тогда

$$\mathsf{cov}(\xi_1, \xi_2) = 2.05 - 3.4 \cdot 0.55 = 0.18$$

Найдем коэффициент корреляции ρ для наших случайных величин. Напомним, что

$$\rho(\xi_1, \xi_2) = \frac{\mathsf{cov}(\xi_1, \xi_2)}{\sqrt{\mathsf{D}\xi_1} \cdot \sqrt{\mathsf{D}\xi_2}}$$

Найдем соответствующие дисперсии ξ_1 и ξ_2 и подставим в выражение для ρ:

$$\rho(\xi_1, \xi_2) = \frac{0.18}{\sqrt{1.24} \cdot \sqrt{4.4475}} \approx 0.08$$

Как видно, коэффициент корреляции отличен от 0, значит зависимость есть (мы это уже показывали в примере 6.1.2), при этом она достаточно далека от линейной.

Замечание 6.2.1 *Помимо дискретного совместного распределения случайных величины, в жизни очень часто встречаются и непрерывные совместные распределения случайных величин. Работа с последними требует более серьезной математической подготовки, поэтому в рамках данного курса рассматриваться не будет. Отметим лишь, что для двумерных непрерывных случайных величин сохраняются данные ранее определения, а также свойства ковариации и коэффициента корреляции.*