DIGITAL CULTURE
ITMO UNIVERSITY

Random Variables. Convergence of Random Variables

St. Petersburg
2019

Contents

1 Population and Sample

In the previous module, we discussed such fundamental concepts of probability theory as a random experiment and random event. We also introduced the definition of event probability. Moreover, we built several probability models, including the classical probability scheme, scheme with outcomes that are not equally likely, conditional probability scheme, geometric probability, and so on.

1.1 Concepts of Population and Sample

By the end of the previous module, you've learned the theoretical concept of a random variable. You've also learned how to find its expected value and variance, as well as explain what these parameters mean. Well, we're lucky if we know the distribution of a random variable. You may wonder why it is so. It turns out that there's a difference between theory and practice. In practice, we want to find the distribution itself (or its characteristics). Look.

An average researcher deals with a dataset collected as a result of some random experiment (or process). The dataset includes, for example, liters of gasoline per day within a month, the number of car wash clients per day within a week, the height of the first 100 draftees, and so on. The data is the result of an observation.

It is extremely important to ask whether there is a pattern in the data. Say, if we compare the height of the first 100 draftees of this and previous year, no doubt, there will be some differences. Nevertheless, the overall picture will be pretty the same. However, what will change (if something will change) if there are more observations, for example, one or ten thousand? Obviously, in all these cases, we will observe a relatively small number of people who are very short or tall, while most of the observations will be in some conditionally-average range.

In applied problems, many regularities happen to have a probability nature. But what does this mean?

It means that the studied regularity (event) is random in all its forms and, therefore, described by random variable ξ that, in turn, has a probabilistic distribution. In the draftee case, the probability to be of the height that slightly differs from the average is high, the farther the height from the average, the lower the probability to be of that height.

Based on the known distribution of ξ, we can calculate such useful parameters as expected value $\mathsf{E}\xi$, variance $\mathsf{D}\xi = \mathsf{E}\left(\xi - \mathsf{E}\xi\right)^2$, probabilities of falling in certain sets, and more. In other words, the known distribution of the random variable of interest tells us a lot. It allows us to extract a ton of information we can use.

Well, let us have n forms of some probabilistic regularity. We observe n values of some random variable ξ that is also called the population having some (unknown) distribution. Hence, it is a typical n-dimensional vector

$X = (x_1, x_2, ..., x_n)$ being a sample obtained after the experiment. Thus,

Definition 1.1.1 *Let ξ be a random variable of interest. Sample (after the experiment) $X = (x_1, x_2, ..., x_n) \in \mathbb{R}^n$ is n independent realizations of random variable ξ. The latter is often called a population.*

For example, assume that ξ is the height of a random person, then,

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

is the sample of size 5 from population ξ. Suppose we want to calculate the average, in particular, predict expected value ξ. How do we do it? It seems appropriate to calculate the arithmetic mean, called the sample mean:

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

But is the obtained value the expected value of population ξ? We inherently understand that it is not true because if we select 5 other persons, the mean value can also change. Therefore, we'd better say that sample mean $\overline{X}$ is the estimate of the expected value of population ξ.

You should also consider heterogeneity. The height of people in a kindergarten (even including the staff height) and the height of a basketball team will be dramatically different. The thing is, the samples are taken from different distributions. Therefore, we ought to have a clear understanding what distribution we study for that sample. Or it would be similar to the online poll asking people if they use the internet.

1.2 Empirical Distribution

It is reasonable to associate new random variable ξ^* with each sample obtained as a result of the experiment, that is, with the sample $X = (x_1, x_2, ..., x_n)$. In particular, this random variable should take each sample value x_i with probability $\frac{1}{n}$ because we treat all the values the same way. Thus, we can fill out the table with the following values:

ξ^*	x_1	x_2	...	x_n
$\widetilde{\mathsf{P}}$	$\frac{1}{n}$	$\frac{1}{n}$	...	$\frac{1}{n}$

By the way, you can see that the random variable, or rather its distribution depends on n, the sample size, but we will not assign an additional index to it.

Since random variable ξ^* is not true random variable ξ, the probability with which it takes values will be denoted by $\widetilde{\mathsf{P}}$. For example, if, as a result of an experiment showing the number of goals scored in five matches, we have a sample

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

of size 5, the corresponding distribution of new random variable ξ^* will be written as

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

It turns out that, as sample size n increases, distribution ξ^* better approximates the true (but unknown) distribution of population ξ.

Perhaps, the idea is clear. We can use the corresponding characteristic ξ^* as an estimate of some characteristic of population ξ, since the distribution of the latter well approximates the true one. Let's give it a try.

Then, as estimate $\widehat{\mathsf{E}\xi}$ of expected value ξ, it is logical to take the corresponding expected value $\widetilde{\mathsf{E}}\xi^*$, that is, in fact, a sample mean:

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

As estimate $\widehat{\mathsf{D}\xi}$ of the variance of population ξ, it is logical to take the variance of random variable ξ^* called sample variance:

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

Remark 1.2.1 *The assumption that, as sample size n increases, distribution ξ^* better approximates the true distribution of population ξ, is inaccurate, but we will deal with in the next modules.*

Example 1.2.1 *Let's find the sample mean and sample variance for the previously considered sample*

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

and empirical random variable ξ^* constructed for the sample. The random variable distribution is given by the table:

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

Sample mean $\overline{X}$ is found as follows:

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

What does it mean? Well, according to the sample, we should expect $1-2$ goals per match on average.

Sample variance is calculated as follows:

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

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

and the root of it, that is, estimate $\widehat{\sigma}$, *equals* 0.8.

What conclusion can be made? According to the sample, the most probable outcomes are from $1.6 - 0.8 = 0.8$ *to* $1.6 + 0.8 = 2.4$ *goals, that is,* $1 - 3$ *goals on average. This does not mean that a team cannot score, say,* 6 *goals, or score nothing at all. These events have a small probability.*

Remark 1.2.2 *Recall that the obtained values are the estimates of the true values of the expected value and variance of a certain population (random variable)* ξ. *However, here's the question. How accurate are these estimates, and are there any others? We will answer this question in the next modules.*

2 Main Laws of Distribution of Discrete Random Variables

As has been noted before, a sample is a certain number of independent realizations of population ξ, that is, some finite set of values that the random variable has taken as a result of the experiment. Regularities that occur in the results of experiments can be described by random variables with different distributions. It turns out that many different events that occur daily are described by random variables having a somewhat limited set of distributions. Let's consider the most common ones.

2.1 Bernoulli Distribution

Let's talk about the so-called Bernoulli distribution. We will start with the example inspired by the collected statistics.

Example 2.1.1 *Worldwide, there are from* 104 *to* 107 *boys born for every* 100 *girls (numbers vary by country). For convenience, we will suppose that* 106 *boys are born. We can assume that, based on the ratio, a pregnant woman's chance of giving birth to a boy is*

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

Suppose some family already has a child or, to be more specific, a girl. Then, if the second child is a girl, parents will not have to spend money on clothes and *toys, as the older sister will share them. If it's a boy, extra expenses cannot be avoided. Then, the distribution of random variable ξ reflecting the expenses on the second child can be given by the table:*

$$\begin{array}{c|c|c} \xi & 0 & 1 \\ \hline \mathsf{P} & 0.52 & 0.48 \end{array}.$$

It turns out that there are many similar examples. Any experiment that ends with one of two possible outcomes suits the scheme we are describing. These outcomes are a success corresponding to 1, or a failure corresponding to 0. Other examples are hitting or missing the target, being or not being approved for a loan, passed an exam or did not pass, hungry or not, and so on.

Definition 2.1.1 *Random variable ξ is said to have a Bernoulli distribution with parameter p (written as $\xi \sim \mathsf{B}_p$) if its distribution is given by the table:*

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

where $p \in (0, 1)$.

Remark 2.1.1 *Value $(1 - p)$ is often designated by q. Additionally, p is called the probability of a success, and q is the probability of a failure.*

Let's find the expected value of random variable ξ having a Bernoulli distribution. According to the definition of the expected value,

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

The variance can be calculated by the following formula:

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

If $\xi \sim \mathsf{B}_p$, the distribution of random variable ξ^2 will be given by the table

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

Then,

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

therefore,

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

Thus, if $\xi \sim \mathsf{B}_p$,

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

2.2 Binomial Distribution

Let's consider the law of distribution of a random variable that is closely related to the trial sequences associated with the Bernoulli process.

Suppose that random variable ξ shows the number of successes in a series of n trials according to the Bernoulli process with the probability of success p in each trial.

Definition 2.2.1 *Random variable ξ is said to have a binomial distribution and written as $\xi \sim \mathsf{Bin}(n, p)$, $p \in (0, 1)$, $n \in \mathbb{N}$ if it takes values $0, 1, 2, ..., n$ with the probabilities*

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

The distribution table of a random variable having a binomial distribution is as follows:

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

Remark 2.2.1 *Note that, in order for the given table to specify the distribution of random variable ξ, it is necessary that*

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

that is, the sum of the numbers in the second row should be one. The latter follows from the Newton binomial formula:

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

if we assume that $a = p$, $b = 1 - p$ (in this case, $(a + b)^n = (p + 1 - p)^n = 1$).

Example 2.2.1 *A student takes a test with 8 questions. The test has multiple choice questions with four choices with one correct answer each. We can write the expected distribution of the number of correctly answered questions if the answers are chosen at random.*

It is clear that the random variable under consideration has binomial distribution $\mathsf{Bin}(8, 0.25)$. The distribution table has the following form (the values of the second row are rounded so that the distribution is obtained, in other words, so that the sum of the numbers of this row is one):

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

The probabilities of the corresponding values of the random variable are calculated by the formulas:

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

We pose a question. What is the probability of passing the test if, to do so, one should have at least 5 *correct answers? Clearly,*

$$\mathsf{P}\left(\xi \geq 5\right) = \mathsf{P}(\xi = 5) + \mathsf{P}(\xi = 6) + \mathsf{P}(\xi = 7) + \mathsf{P}(\xi = 8) =$$
$$= 0.023 + 0.004 = 0.027.$$

Let's calculate the expected value of random variable ξ with a binomial distribution. For this, we will consider random variables:

$$\xi_i = \begin{cases} 1, & \text{if trial } i \text{ has been successful,} \\ 0, & \text{if trial } i \text{ has failed,} \end{cases}$$

where $i \in \{1, 2, ..., n\}$

In our conditions, all these random variables ξ_i have a Bernoulli distribution B_p with the same parameter p, and

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

since the number of terms equal to one is equal to the number of successes in n trials. Then, according to the property of expected value,

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

To find the variance, we will use the fact that the tests in the Bernoulli scheme are independent, and, therefore, random variables ξ_i will be independent, then, according to the variance properties,

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

Example 2.2.2 *Let's continue with the student's test example. Our task is to find the expected value and variance. Recall that the number of questions (tests)* $n = 8$*, the probability of success in each* $p = 0.25$*, then,*

$$\mathsf{E}\xi = np = 8 \cdot 0.25 = 2,$$
$$\mathsf{D}\xi = npq = 8 \cdot 0.25 \cdot 0.75 = 1.5.$$
$$\sigma = \sqrt{\mathsf{D}\xi} = \sqrt{1.5} \approx 1.22$$

In other words, a student will answer correctly 2 *of* 8 *questions on average. Moreover, on average, the number of correct answers will range from* 0.78 *to* 3.22.

2.3 Poisson Distribution

First, we need to introduce a formal definition and then explain when to apply it.

Definition 2.3.1 *A random variable is said to have a Poisson distribution and written as $\xi \sim \Pi_\lambda$, $\lambda > 0$ if it takes values $0, 1, 2, 3, ..., n, ...$ with the probabilities*

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

Let's prove that the introduced distribution is indeed a distribution, namely, we show that

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

We have

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

since the resulting series is nothing but the Maclaurin series for the exponent

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

Remark 2.3.1 *The Poisson distribution well describes the so-called rare events when an event rarely occurs within a sufficiently large number of trials which is not known in advance, in other words, it has a low probability. For example, the number of car accidents in a city with respect to all trips of all drivers of this city during a year, the number of marketing messages sent to the phone number per month with respect to all messages received during this month, the number of goals scored at the World Cup with respect to all shots.*

We can show that the expected value and variance of the Poisson distribution are the same and equal to λ, that is,

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

This course will not cover the proof of this statement.

Example 2.3.1 *A manager of the telecom company decided to calculate the probability that, in a small town, $0, 1, 2, \ldots$ calls will be picked up within five minutes. The manager has chosen random five-minute intervals and counted the number of* *calls in each of them. It turned out that* 4.3 *calls were picked up on average. Let's calculate the probability that exactly* 7 *calls will be picked up within five minutes:*

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

The probability of such an event is low. Therefore, the manager doesn't need to hire new employees or buy new equipment.

3 Distribution Function of Random Variable

3.1 Distribution Function Definition and Properties

It has been well noted that the distribution of a discrete random variable can be given by the table, that is, by listing the values of a random variable and probabilities to take these values. However, random variables that are not discrete cannot be given this way. There is a way to set the distribution of a random variable. It is based on the definition of the so-called distribution function.

Definition 3.1.1 *The distribution function of random variable ξ is the function*

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

So, the distribution function of random variable ξ at point x shows the probability of the event that $\xi \in (-\infty, x)$.

Distribution function $F_\xi(x)$ of random variable ξ has the following intuitive properties:

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

2. $F_\xi(x)$ is not decreasing, that is,

$$\text{if } \ x_1 < x_2, \ \text{then, } \ F_\xi(x_1) \leq F_\xi(x_2).$$

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

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

The first property follows from the fact that the distribution function is some probability, and, since the probability is between 0 and 1, the values of the distribution function lie there too. The second property follows from the monotony of probability because, given $x_1 < x_2$, event $\{\xi < x_1\}$ leads to $\{\xi < x_2\}$. The last two properties are based on the fact that ξ is a random variable, that is, a function, which means its values are in set $\mathbb{R} = (-\infty, +\infty)$. Roughly speaking, the third property means that we calculate the probability of event $\{\xi < -\infty\}$ that, of course, is an impossible event, and its probability is zero. For the latter property, we calculate the probability of event $\{\xi < +\infty\}$ that is a sure event. Therefore, its probability is 1.

How does the distribution function help us?

Remark 3.1.1 *It turns out that, with knowledge of the distribution function, we can find the probability that random variable values will fall within a given interval, that is:*

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

where $F_\xi(x+0)$ in the formulas means the limit of function F_ξ at point x on the right. a, b can be infinities (with the corresponding signs).

Let's consider an example in which we construct the distribution function to calculate the probability of interest.

Example 3.1.1 *Suppose that, among each* 100 *lottery tickets, there are* 15 *winning tickets on average. The data on the number of tickets and prize amounts (in rubles) is shown on the screen.*

prize	0	100	500	2.000
number of tickets	85	10	4	1

Let ξ be a random variable reflecting the prize amount per one ticket chosen at random. Then, its distribution is given by the following table:

ξ	0	100	500	2.000
P	0.85	0.1	0.04	0.01

Let's construct the distribution function $F_\xi(x) = \mathsf{P}(\xi < x)$ of that random variable. It's clear that key construction points are the point values of the random variable: 0, 100, 500, 2000. *Then,*

$$F_\xi(x) = \begin{cases} 0, & x \le 0 \\ 0.85, & 0 < x \le 100 \\ 0.95, & 100 < x \le 500 \\ 0.96, & 500 < x \le 2.000 \\ 1, & x > 2.000 \end{cases}.$$

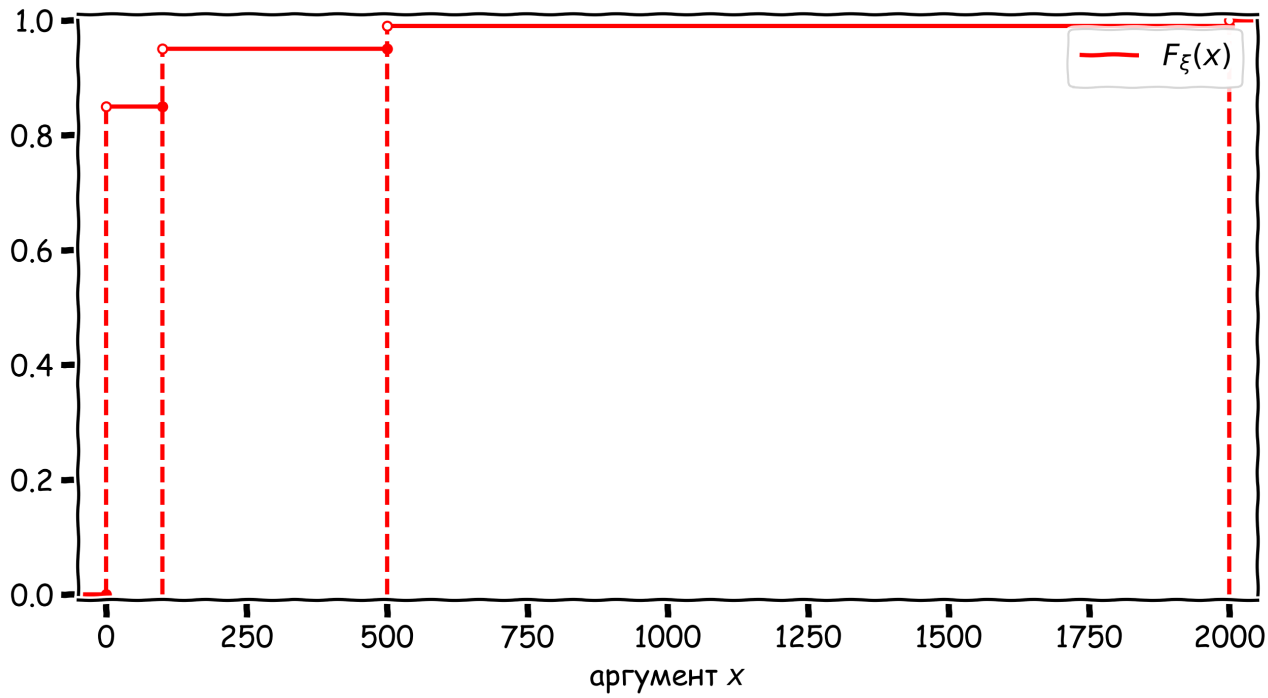

Figure 1: Distribution function $F_\xi(t)$

The graph of the distribution function is shown on the screen. How did we get this distribution function? Well, what happens when $x \leq 0$? *We are interested in the probability of an event that the random variable is less than* x. *However, our random variable* ξ *does not take negative values. Thus, the probability of event* $\{\xi < x,\ x \leq 0\}$ *is* 0, *and the value of the distribution function is* 0 *too.*

Let's consider interval $0 < x \leq 100$. *At any* x *in this interval, event* $\{\xi < x\}$ *consists of exactly one, zero, value of the random variable, and its probability equals* 0.85, *which is written in the distribution function.*

In conclusion, let's pose a question. What is the probability of the event that a random ticket will win more than 100 *rubles?*

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

Since 500 *and* 2.000 *are the random variable values that are more than* 100, *the last sum consists of two terms:*

$$\mathsf{P}\left(\xi > 100\right) = \mathsf{P}\left(\xi = 500\right) + \mathsf{P}\left(\xi = 2.000\right) = 0.04 + 0.01 = 0.05.$$

Hence, the probability of interest equals 0.05.

Based on the distribution function, we can calculate the same probability as follows:

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

3.2 Distribution Density of Continuous Random Variable

We have considered so far discrete random variables only, and some students might have an incorrect notion that there are no other distributions. Actually, discrete distribution is one of two distribution extremes. Another extreme is the so-called continuous distribution. Let's elaborate upon it.

Definition 3.2.1 *Random variable ξ is said to have a continuous distribution if there is such non-negative function $f_\xi(x) : \mathbb{R} \to \mathbb{R}$ that*

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

Definition 3.2.2 *Function $f_\xi(x)$ is called the density function of random variable ξ.*

It is like in the case of the discrete random variable. A continuous variable is obtained from a discrete one by the passage to the limit in some sense. The ratio pair clearly shows that (the sum changes to the integral):

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

Moreover, since $\lim\limits_{x\to+\infty} F_\xi(x) = 1$, then $\int\limits_{-\infty}^{+\infty} f_\xi(t)dt = 1$. In the discrete case, there was equation $\sum\limits_i \mathsf{P}(\xi = a_i) = 1$. Thus, we obtain a pair of related ratios.

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

The geometric meaning of density is as follows. Let f_ξ be the density of some random variable ξ. As has been noted, the area under the graph equals $\int\limits_{-\infty}^{+\infty} f_\xi(t)dt = 1$. Since the geometric meaning of the definite integral in interval I of nonnegative function is the area under the graph of this function above interval I, then

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

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

The probability that a continuous random variable will be in a certain set $A \subset \mathbb{R}$ is the area under the density graph above set A.

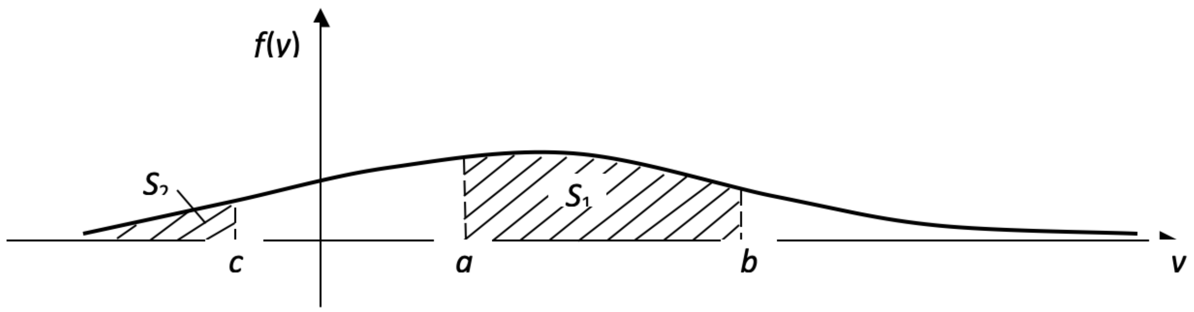

Figure 2: Geometric meaning of distribution density

4 Numerical Characteristics of Continuous Random Variables

We have already considered the numerical characteristics of discrete random variables. So let's introduce these characteristics for continuous random variables.

4.1 Expected Value

Definition 4.1.1 *Let random variable ξ have a continuous distribution with density $f_\xi(x)$. The expected value of random variable ξ is the number*

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

given that the integral converges absolutely (that is, if there's $E|\xi|$). Otherwise, the expected value does not exist.

Example 4.1.1 *Let the density of random variable ξ be as follows:*

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

Let's calculate the expected value according to the definition:

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

It turns out that one potentially useful property is true for the expected value.

Let $g(\xi)$ be a random variable constructed from random variable ξ. If ξ has a continuous distribution, and the expected value $g(\xi)$ is determined, then

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

4.2 Variance of Continuous Random Variable

According to the definition, the variance of any random variable (including that of a continuous one) equals

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

if there's $\mathsf{E}\xi^2$. We have already noted that variance is easier to calculate using the formula:

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

This ratio follows from the properties of the expected value, namely

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

since the expected value is linear, $\mathsf{E}\xi$ is a constant, and the constant is brought outside of the expected value sign, therefore, $\mathsf{E}(2\xi\mathsf{E}\xi) = 2\mathsf{E}\xi\mathsf{E}\xi = 2(\mathsf{E}\xi)^2$ and $\mathsf{E}(\mathsf{E}\xi)^2 = (\mathsf{E}\xi)^2$.

Based on this statement about calculating the expected value of the function of the random variable, the variance of a continuous random variable (if it exists) will be expressed in terms of the distribution density as follows:

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

Remark 4.2.1 *For a continuous random variable, the expected value and variance have the same properties as for a discrete random variable.*

Example 4.2.1 *Let's consider random variable ξ from the previous example:*

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

We have already found its expected value $\mathsf{E}\xi = \frac{4}{3}$. Now we can find its variance.

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

Then,

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

5 The Main Laws of Distribution of Continuous Random Variables

As in the discrete case, we will consider the most common continuous distributions.

5.1 Uniform Distribution

Definition 5.1.1 *Random variable ξ is said to have a uniform distribution over line segment $[a, b]$, which is written as $\xi \sim \mathsf{U}_{a,b}$, if its density is given by*

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

The density graph is shown on the screen.

As to the meaning of this distribution, we can assume that the random variable is the coordinate of a point randomly thrown onto line segment $[a, b]$. The variable density is uniformly distributed on the line segment, therefore, the chances of hitting any point on it are equally likely.

The graph of the distribution function is shown on the screen.

Remark 5.1.1 *We can draw an analogy between the considered distribution and geometric probability. If random variable ξ is uniformly distributed on line segment $[a, b]$, the probability of being within interval $A = (c, d) \subset [a, b]$ is proportional to the length of this interval and is equal to*

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

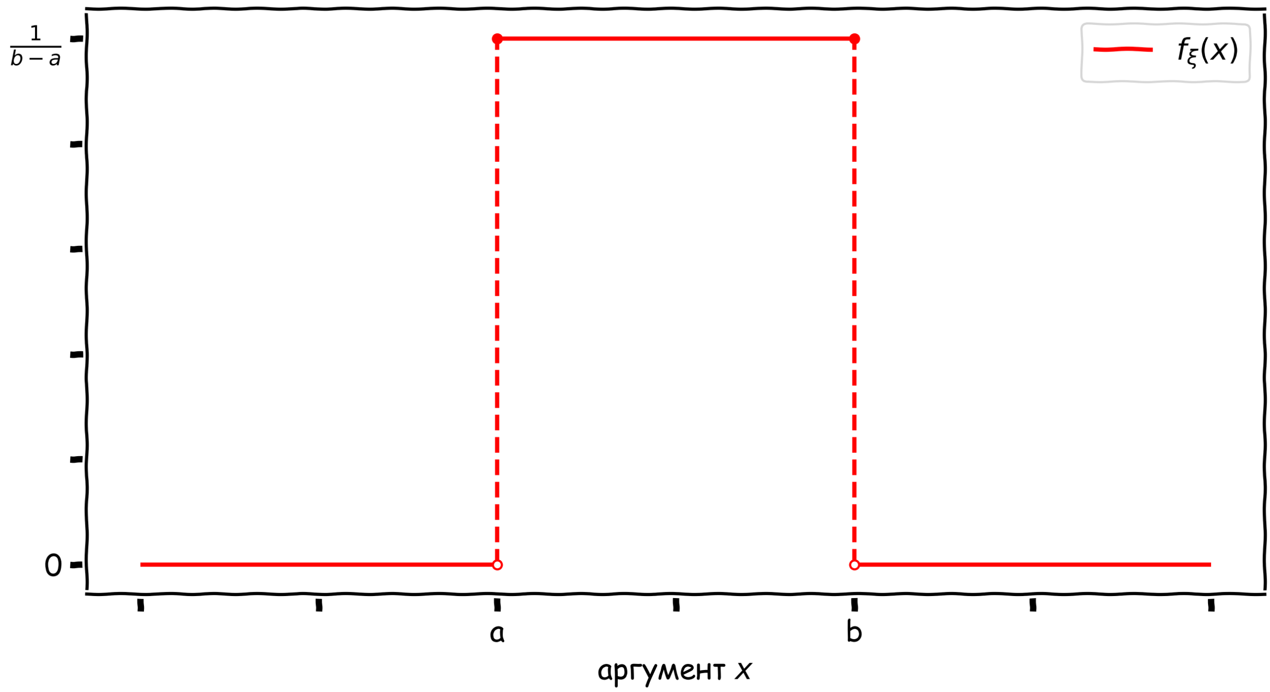

Figure 3: The distribution function of the random variable with the uniform distribution

Example 5.1.1 *Two people are pulling on the ends of a uniform* 1*-meter rope, and it breaks at some point* $\xi \in [0,1]$. *We need to find the probability that it will break within the interval between* 10 *and* 12 *cm. Since the rope is uniform, we will assume that random variable* ξ *has a uniform distribution over line segment* $[0,1]$. *In this case,* $a = 0$, $b = 1$, $c = 0.1$, *and* $d = 0.12$. *Then,*

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

The expected value of random variable ξ with the uniform distribution equals

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

To compute the variance, we will find $\mathsf{E}\xi^2$.

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

Considering that $\mathsf{E}\xi = \frac{a+b}{2}$, we obtain:

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

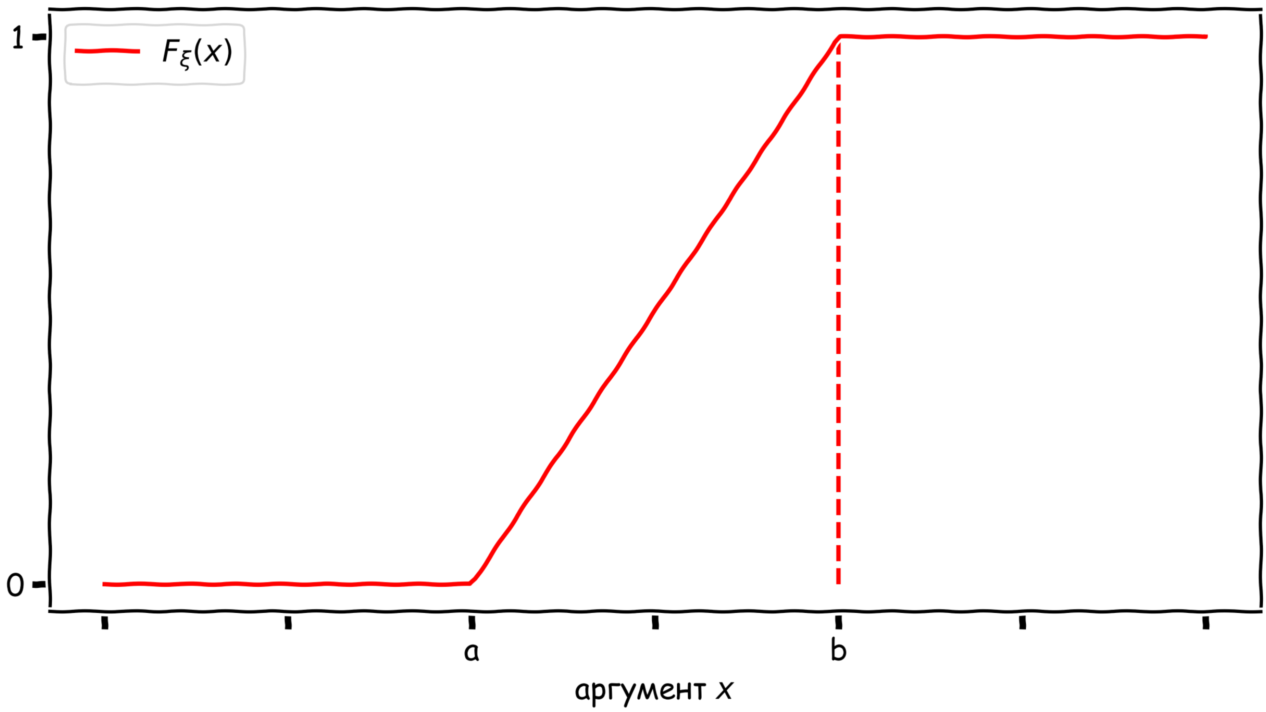

Figure 4: The distribution function of the random variable with the uniform distribution

5.2 Exponential Distribution

The next distribution is, in a sense, a continuous analog of the Poisson distribution.

Definition 5.2.1 *Random variable ξ is said to have an exponential distribution with parameter $\lambda > 0$, which is written as $\xi \sim \mathsf{Exp}_\lambda$, if its density is given by*

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

The graph of the exponential distribution is shown on the screen.

The distribution function of random variable ξ is easily calculated and defined as the ratio:

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

The graph of the last function is shown on the screen. It turns out that the duration of calls, the time gaps between customer visits, the time of customer service, the periods of faultless operation of the device, and many similar things have an exponential distribution.

For random variable $\xi \sim \mathsf{Exp}_\lambda$, we can show that

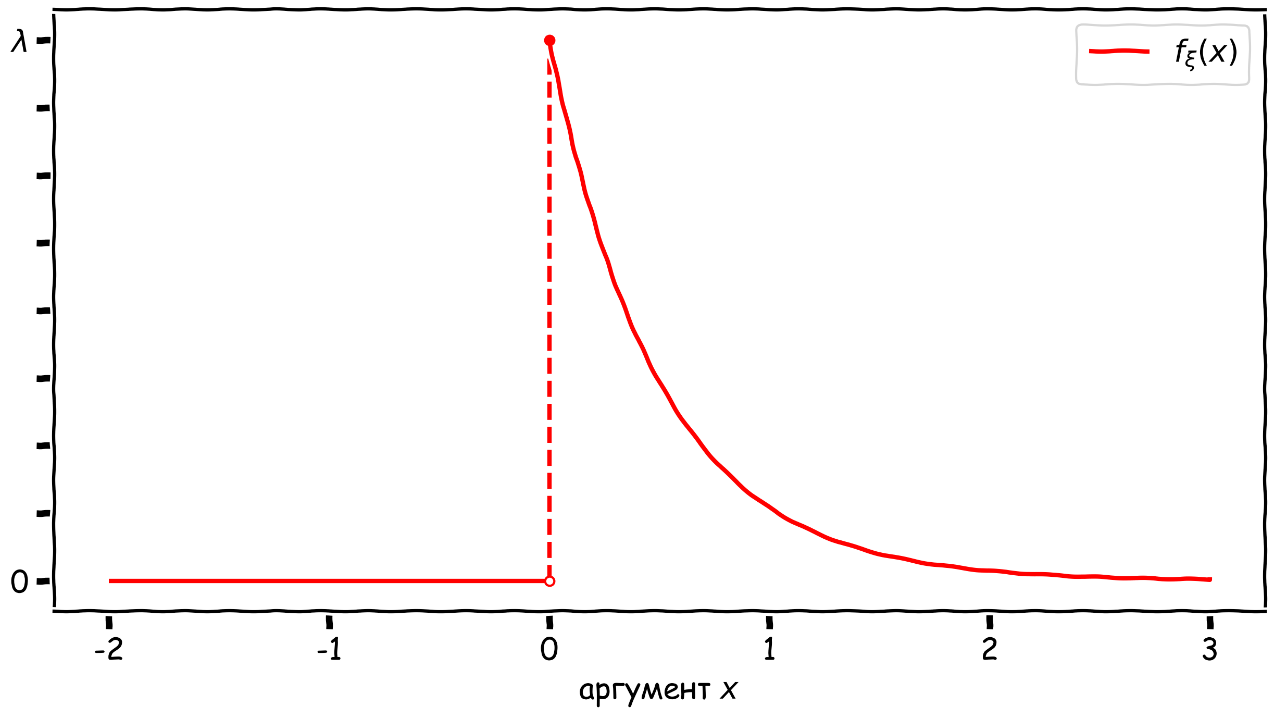

Figure 5: The density of the random variable with the exponential distribution.

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

Example 5.2.1 *Let ξ be a random variable for the service time (in minutes) in the store. It has an exponential distribution. What is the probability that the customer service time will take from 2 to 4 minutes if the average service time is 2 minutes? Based on the problem setting:*

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

The distribution function of random variable $\xi \sim \mathsf{Exp}_{\frac{1}{2}}$ for all $x \geq 0$ takes the following form:

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

then,

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

5.3 Normal Distribution

The next distribution example is one of the most important.

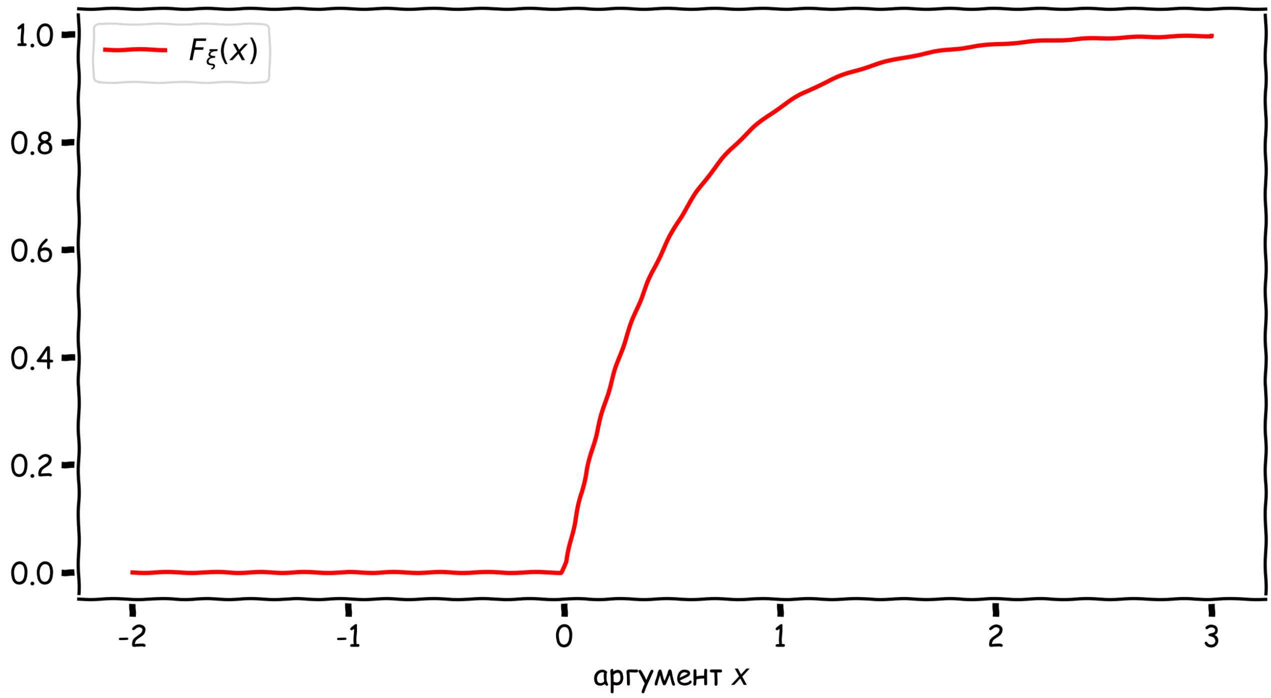

Figure 6: The distribution function of the random variable with the exponential distribution.

Definition 5.3.1 *Random variable ξ is said to have a normal (Gaussian) distribution with parameters $a \in \mathbb{R}$, σ^2, which is written as $\xi \sim \mathsf{N}_{a,\sigma^2}$, if its density is given by*

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

The density graph of the normal distribution given different values of a, σ^2 is shown on the screen. The normal distribution is common because it fits many natural phenomena.

Definition 5.3.2 *The normal distribution with parameters $a = 0$, $\sigma^2 = 1$, i.e. a distribution $\mathsf{N}_{0,1}$, is called the standard normal distribution.*

The density of the standard normal distribution has the form

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

The distribution function of the normal distribution, because of the importance of the latter, is designated by:

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

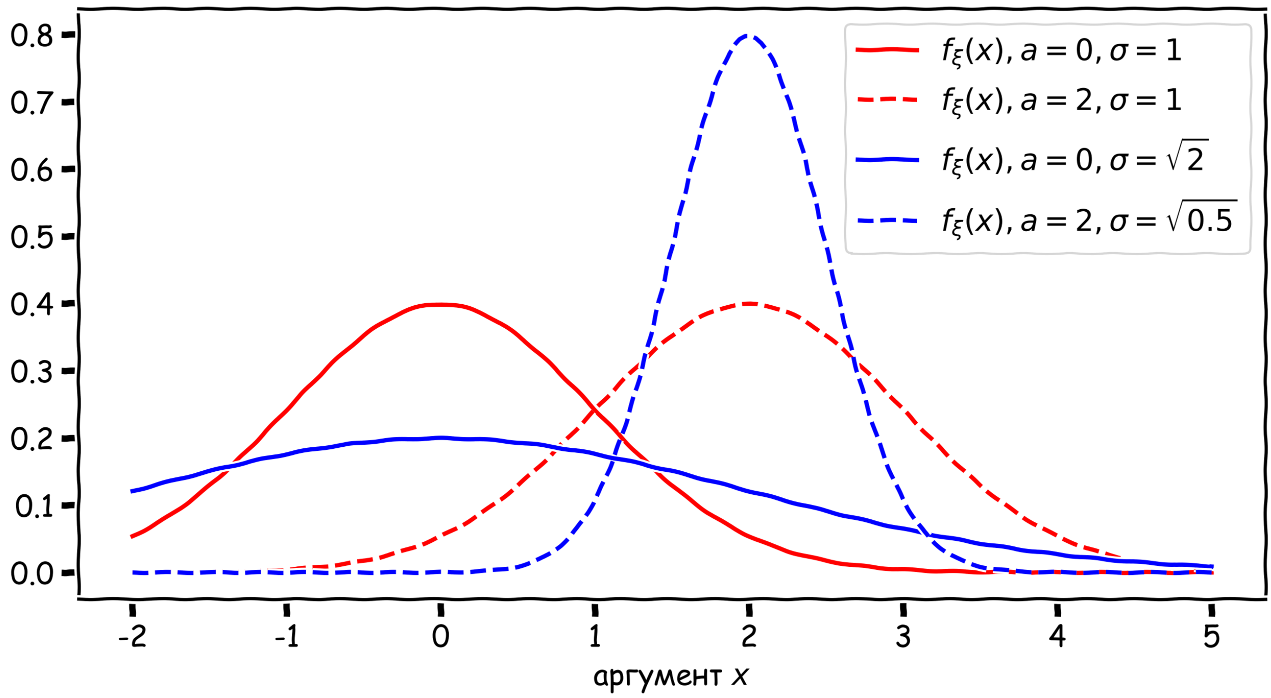

Figure 7: The density of the random variable with the normal distribution.

It can't be expressed in elementary functions, and its values are provided in the tables. More precisely, the distribution function of the standard normal distribution is tabulated, but, by substituting the variable in the integral, we can easily prove that

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

The graph of the normal distribution law is shown on the screen. Note several important properties of the distribution function.

Lemma 5.3.1 *The distribution function of the standard normal distribution has the following properties:*

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

As in the case of the exponential distribution, we accept without proof the fact that if $\xi \sim \mathsf{N}_{a,\sigma^2}$, then

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

Example 5.3.1 *Random variable ξ is distributed according to the normal law with parameters $a = 2.5$, $\sigma = 2$. We need to define $\mathsf{P}(|\xi| \leq 3)$.*

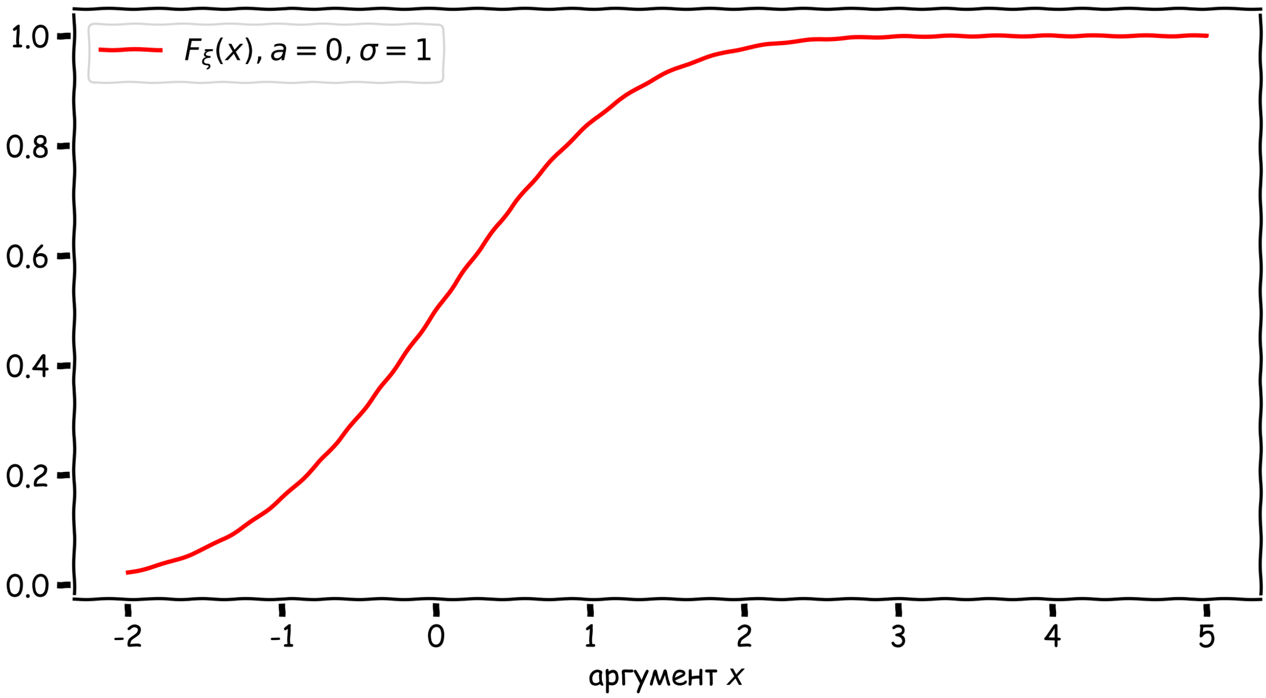

Figure 8: The distribution function of the random variable with the standard normal distribution.

The probability that a random variable will fall on line segment $[-3, 3]$ *can be found as follows:*

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

To random variable ξ with normal distribution N_{a,σ^2}, the 3-sigma (3σ) rule applies. According to it,

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

There's no need to remember the specific value, but it is important to understand that, when the random variable has normal distribution N_{a,σ^2}, almost the entire density is located at distance 3σ from the average a.

6 Joint Probability Distribution of Two Random Variables

Real-world problems often require to consider not one but several random variables. The distribution of one of them usually depends on the distribution of the others. In this section, we will study issues related to the distribution law of vector random variables, discuss the nature of the possible relationship between coordinates of a random vector and the ways to determine the measure of this relationship. In this course, we will consider only two-dimensional discrete variables.

6.1 Joint Probability Distribution of Two Discrete Random Variables

Definition 6.1.1 *The joint probability distribution of random variables ξ_1 and ξ_2 is set of probabilities $\mathsf{P}(\xi_1 = a, \xi_2 = b)$, where numbers a include all possible values $a_1, ..., a_n$ of random variable ξ_1, and numbers b include all possible values $b_1, ..., b_k$ of random variable ξ_2, and*

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

Let random variable ξ_1 take values $a_1, ..., a_n$, and random variable ξ_2 take values $b_1, ..., b_k$. The joint probability distribution of random variables is often written in the form of a table:

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

The joint probability distribution allows recovering the so-called marginal (one-dimensional) distributions of random variables ξ_1 and ξ_2 according to the rules

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

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

Example 6.1.1 *Assume that we have the table of the joint probability distribution of two random variables ξ_1 and 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

Let's fill out the tables given the marginal distributions.

To find the marginal distributions, we sum up the probabilities by the columns for random variable ξ_1, and by the rows for random variable ξ_2. For example, for $\xi_1 = -1$, we sum up the probabilities in the column corresponding to value -1 of random variable ξ_1 and obtain

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

Next, we do the same for $\xi_2 = 7$ but use the row:

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

Following the operations with rows and columns, we obtain the marginal distributions given by the following tables:

ξ_1	-1	-2	5
P	0.35	0.35	0.3

,

ξ_2	3	5	7
P	0.4	0.2	0.4

.

The joint probability distribution helps to understand whether the random variables are related or not, because the distribution allows us to conclude if the equality is satisfied or not.

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

or not satisfied.

Example 6.1.2 *Let's check whether random variables ξ_1 and ξ_2 from the previous example are independent. They are independent if the equality*

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

is satisfied for all a and b. Otherwise, the random variables are dependent.

The tables of joint probability and marginal distributions take the following form:

$\xi_2\backslash\xi_1$	-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

Then,

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

On the other hand,

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

Since $0.1 \neq 0.14$, *we can conclude that the random variables are dependent.*

We would like to draw attention to another important aspect of the joint probability distribution. The known joint probability distribution of random variables allows us to write the distribution of various functions of these random variables, in particular, the distribution of the sum, difference, or product. The known marginal distributions do not allow us to do the same. The next example will prove it.

Example 6.1.3 *Let the joint probability distribution of random variables* ξ_1 *and* ξ_2 *be given by the following table* ($r \in [0, 0.5]$)

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

The marginal distributions of random variables ξ_1 *and* ξ_2 *are the same and do not depend on* r*:*

ξ_1	0	1
P	$\frac{1}{2}$	$\frac{1}{2}$

ξ_2	0	1
P	$\frac{1}{2}$	$\frac{1}{2}$

Let's find the distribution of random variable $\xi_1 + \xi_2$. *The sum can take values* $0, 1, 2$, *and*

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

and

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

Thus,

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

Hence, the distribution depends on r, *given the constant marginal distributions.*

6.2 Covariance. Correlation Coefficient

Usually, the event of interest has more than one feature. For example, person features include height, weight, age, income level, place of residence, etc. Car features are maximum speed, mileage, engine displacement, etc. So, one may ask about the relationship between the variables if there's any. Can we say that a change in one feature is likely to result in a change in another? For example, if we eat five croissants a day, our weight most likely will increase. When the dollar rate is rising, the Euro exchange rate will change as well. Mortgage rates are becoming lower when the demand for real estate increases.

When we consider mathematical functions, we observe a direct dependency of the function value and the argument. For example, $y = 2x + 3$. Here, each value x corresponds to one strictly defined value y. When dealing with random variables, such a direct functional relationship cannot be constructed. We can consider the correlation. It is a probabilistic relationship between the values, which occurs when one of the values depends not only on the given second value but, perhaps, on other random conditions.

Let's pose the question about the dependence or independence of two random variables. Let ξ_1 and ξ_2 be discrete random variables with values $a_1, ..., a_n$ and $b_1, ..., b_k$, respectively. Their joint distribution is known. Thus,

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

If we assume that ξ_1 and ξ_2 are independent,

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

therefore,

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

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

Hence, if random variables ξ_1 and ξ_2 are independent, this is reflected in the algebraic property:

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

if the equality is not satisfied, the variables are dependent.

Definition 6.2.1 $\mathsf{E}(\xi_1\xi_2) - \mathsf{E}\xi_1\mathsf{E}\xi_2$ *is called the covariance of random variables* ξ_1 *and* ξ_2 *is denoted by* $\mathsf{cov}(\xi_1, \xi_2)$.

There is a problem with covariance. It is measured in squared units of random variables ξ_1 and ξ_2. It is not good because, for example, 100-times increase in one random variable leads to the same increase in covariance. However, the strength of the relationship does not change. Here's the correlation coefficient comes to aid.

Definition 6.2.2 *The correlation coefficient between two variables* ξ_1 *and* ξ_2 *with non-zero variances is called the variable*

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

The correlation coefficient has the following properties:

1. Its absolute value does not exceed unity, that is, $|\rho(\xi_1, \xi_2)| \leq 1$;
2. If ξ_1, ξ_2 are independent, $\rho(\xi_1, \xi_2) = 0$;
3. $|\rho(\xi_1, \xi_2)| = 1$ if and only if $\xi_1 = a\xi_2 + b$, given that $a \cdot \rho(\xi_1, \xi_2) > 0$.

Note that the correlation coefficient shows the degree of a linear relationship between random variables. The closer it is to unity in absolute value, the closer the relationship to linear.

So, if it is equal to 1, an increase in the values of one random variable leads to an increase in another. If it is equal to -1, an increase in one leads to a decrease in another.

At the same time, when the correlation is not equal to unity, the relationship between two random variables in statistics should be interpreted with great caution. Suppose that we have data on the flood severity and the number of rescuers involved. These indicators correlate, and the correlation coefficient is positive. However, an increase in the number of rescuers does not mean that a terrible disaster has occurred. It also does not mean that the collective dismissal of rescuers will prevent floods and disasters.

Moreover, zero correlation does not mean the absence of the relationship.

Example 6.2.1 *Let's show that the condition*

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

is not sufficient for the independence of the random variables. To do this, let's assume that sample space Ω *consists of* $\Omega = \{-\frac{\pi}{2}, 0, \frac{\pi}{2}\}$, *and two deliberately dependent random variables* $\xi_1 = \sin\omega$ *and* $\xi_2 = \cos\omega$ *are defined on it. We can write down the laws of distribution of random variables:*

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

ξ_2	0	1
P	$\frac{2}{3}$	$\frac{1}{3}$

It's clear that $\mathsf{E}\xi_1 = 0$, $\mathsf{E}\xi_2 = \frac{1}{3}$, *thus,* $\mathsf{E}\xi_1\mathsf{E}\xi_2 = 0$. *Moreover,* $\xi_1\xi_2 = 0$ *is always true on sample space* Ω, *therefore,* $\mathsf{E}(\xi_1\xi_2) = 0$ *is always true too. Hence, the condition*

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

is satisfied, but the variables are dependent.

Example 6.2.2 *Random variables* ξ_1 *and* ξ_2 *are given by the joint distribution table. The task is to calculate the covariance and correlation coefficient. We need to enter the solution rounded to the hundredth.*

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

The covariance of two random variables is calculated as:

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

Let's fill out the distribution table for random variable $\xi_1\xi_2$. *To do so, we need to consider all possible products of the random variable values. The product probability for a particular case will be at the intersection of the corresponding column and row. If the product values are the same (though, it is not our case), there's no need to add the corresponding probabilities. Values with zero probability can be omitted from the table.*

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

.

Let's fill out the tables given the marginal distributions ξ_1 *and* 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

Let's find the corresponding expected values:

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

Then,

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

We also need correlation coefficient ρ for our random variables. Recall that

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

Let's find the respective variances ξ_1 and ξ_2 and plug them in the expression for ρ:

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

As you can see, the correlation coefficient is not 0, so there is a relationship (another example has already shown it), although it is far from linear.

Remark 6.2.1 *In addition to the discrete joint distribution of random variables, continuous joint distributions of random variables are also very common in practice. Working with the latter requires a profound knowledge of mathematics; therefore, this course will not consider it. We would only like to note that, the provided definitions, as well as the properties of covariance and correlation coefficient apply to two-dimensional continuous random variables.*