MINISTRY OF EDUCATION AND SCIENCE,
YOUTH AND SPORTS OF UKRAINE

KHARKIV NATIONAL UNIVERSITY OF ECONOMICS

Misiura Ie.

PROBABILITY THEORY

Summary of Lectures

Харків. Вид. ХНЕУ, 2013

UDC 519.21(042.4)
BBC 22.17я73
M75

Reviewer – Doctor of Engineering, Professor, Head of Thin-Wall Structures Strength Department of A. M. Pidhornyi Institute of Machine-Building Problems *Shupikov O.*

Затверджено на засіданні кафедри вищої математики й економіко-математичних методів.

Протокол № 5 від 19.12.2012 р.

Misiura Ie.

M75 Probability Theory : summary of lectures / Ie. Misiura. – Kh. : Publishing House of KhNUE, 2013. – 96 p. (English)

Theoretical material on the chapter "Probability Theory" is given. It is the main part of the third module of the academic discipline "Higher and Applied Mathematics". The content of nine themes is revealed. Each of them consists of basic concepts, definitions, formulas, rules, theorems and solved problems for consolidating students' knowledge. Theoretical questions are given for students' self-test.

It is recommended for full-time students of the training direction 6.140103 "Tourism".

Наведено теоретичний матеріал з розділу "Теорія ймовірностей", що становить складову частину третього модуля навчальної дисципліни "Вища та прикладна математика". Розкрито зміст дев'яти тем, кожна з яких вміщує основні поняття, визначення, правила, теореми та значну кількість розв'язаних прикладів для закріплення знань студентом. Для самоперевірки подано перелік теоретичних запитань.

Рекомендовано для студентів напряму підготовки 6.140103 "Туризм" денної форми навчання.

UDC 519.21(042.4)
BBC 22.17я73

Introduction

Probability theory is the branch of mathematics which studies properties, laws and the analysis of mass random phenomena. The basic objects of probability theory are random variables, stochastic process and random events. In practice we often deal with random events, i.e. with events which can occur or can't occur under definite conditions which can't be analyzed by direct computations. Analysis of quantitative laws which can be described by mass random phenomena is the subject of probability theory.

Probability theory plays an important role in everyday life in economics, in business, in trade on financial markets, in risk assessment and many other areas where statistics is applied to the real world.

Owing to the study of probability theory a student is obliged to receive the basic knowledge of this part and use skills of applying the elements of probability theory in investigations where probability theory is applied as an instrument of investigation for forming economic mathematical models of economic processes and developments. This makes it possible for him to apply the acquired knowledge and skills for solving many practical problems of economics and business.

Module 3. Probability theory and mathematical statistics

Theme 24. Basic notions of probability theory

24.1. A stochastic experiment. A subject of probability theory. A mathematical model of stochastic experiments

Theory of probability is that part of mathematics that aims to provide insight into phenomena that depend on chance or on uncertainty. The most prevalent use of the theory comes through the frequentists' interpretation of probability in terms of the outcomes of repeated experiments, but probability is also used to provide a measure of subjective beliefs, especially as judged by one's willingness to place bets. If we want to predict the chance of something happening in the future, we use probability.

Let's consider the *fundamental concepts* of probability theory.

An ***experiment*** is a repeatable process that gives rise to a number of outcomes.

An ***outcome*** is something that follows as a result or consequence.

An ***event*** is a collection (or set) of *one or more* outcomes.

Events are sets and set notation is used to describe them. We use upper letters to denote events. They are denoted as A, B, C, …, A_1, A_2, ….

The simplest indivisible mutually exclusive outcomes of an experiment are called ***elementary events*** $\omega_1, \omega_2, \ldots$.

A ***sample space*** or a ***space of elementary events*** is called the set of all possible elementary outcomes of an experiment, which we denote by the symbol Ω.

Any subset of Ω is called a ***random event*** A (or simply an ***event*** A).

Elementary events that belong to A are said to ***favor*** A.

An event is ***certain*** (or ***sure***) if it always happens.

An event is **impossible** if it never happens.

Equally likely events are such events that have the equal chance to happen at an experiment.

Example 24.1. The experiment (tossing a coin once) has 2 outcomes: head (the first outcome) and tail (the second outcome). The event A is getting "head". For this experiment the sample space is $\Omega = \{head, tail\}$.

The ***probability of an event*** is the chance that the event will occur as a result of an experiment.

Where outcomes are *equally likely* the probability of an event is the number of outcomes in the event divided by the total number of possible outcomes in the sample space.

An *impossible* event has probability 0 and an event that is *certain* has probability 1.

When experiments or observations are made, various outcomes are possible even under the same conditions.

Probability theory deals with regularity of random outcomes of certain results with respect to given observations (in probability theory observations are also called experiments, since they have certain outcomes). Suppose, at least theoretically, that these experiments can be repeated arbitrarily many times under the same circumstances; namely, this discipline deals with the statistics of mass phenomena. The term ***stochastics*** is used for the mathematical handling of ***random phenomena***.

24.2. An algebra of random events. Probabilities in a discrete space of elementary events

The mathematics of probability is expressed most naturally in terms of sets, therefore, let's consider *basic operations with events*.

The ***intersection*** $C = A \cap B = A \cdot B$ of events A and B is the event that both *A* and *B* occur. The elementary outcomes of the intersection $A \cdot B$ are the elementary outcomes that simultaneously belong to A and B.

Example 24.2. If $A = \{1, 2, 3\}$ and $B = \{1, 3, 5\}$ are given, then $C = A \cap B = \{1, 3\}$.

When events A and B have no outcomes in common ($A \cap B = \varnothing$ (this symbol Ø is called the empty set)), they are ***mutually exclusive*** (or ***incompatible events***).

Example 24.3. If events $A = \{1, 2\}$ and $B = \{3, 5\}$ are given, then $C = A \cap B = \varnothing$, because events A and B have no outcomes in common.

When events A and B have common outcomes ($A \cap B \neq \varnothing$), they are ***not mutually exclusive*** (or ***compatible events***).

Example 24.4. In the experiment of throwing a dice the event A of getting an odd number ($A = \{1, 3, 5\}$) and the event B of getting a number greater than 3 ($B = \{4, 5, 6\}$) are not mutually exclusive, i.e. they are compatible, because $A \cap B = \{5\} \neq \varnothing$.

The ***union*** $C = A \cup B = A + B$ of events A and B is the event that at least one of the events A or B occurs. The elementary outcomes of the union $A + B$ are the elementary outcomes that belong to at least one of the events A and B.

Example 24.5. If events $A = \{1, 2, 3, 4, 5\}$ and $B = \{2, 4, 6\}$ are given, then $C = A \cup B = \{1, 2, 3, 4, 5, 6\}$.

Two events A and $\overline{A}$ are said to be ***opposite*** (***complementary***) if they simultaneously satisfy the following conditions: $A \cup \overline{A} = \Omega$ and $A \cap \overline{A} = \varnothing$.

The ***difference*** $C = A \setminus B = A - B$ of events A and B is the event that A occurs and B does not occur. The elementary outcomes of the difference $A \setminus B$ are the elementary outcomes of A that do not belong to B.

Example 24.6. If events $A = \{1, 2, 3, 4, 5\}$ and $B = \{1, 3, 5\}$ are given, then $C = A \setminus B = \{2, 4\}$.

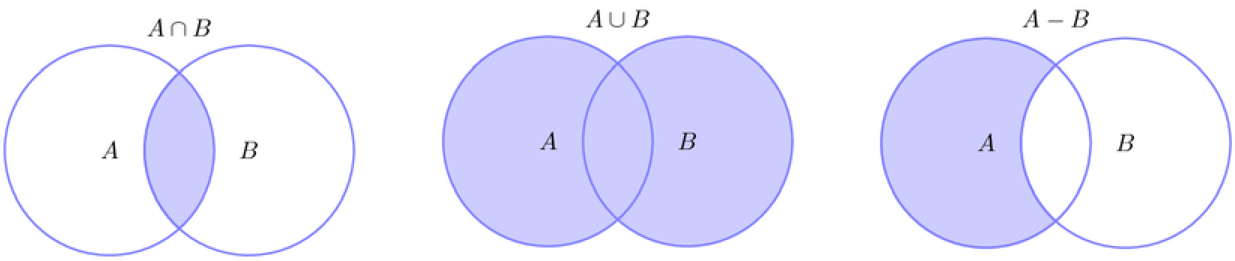
An event A ***implies*** an event B ($A \subset B$) if B occurs in each realization of an experiment for which A occurs.

Example 24.7. If events $A = \{1, 2, 3, 4, 5\}$ and $B = \{1, 3, 5\}$ are given, then the event A implies the event B or $A \subset B$.

Events A and B are said to be ***equivalent*** ($A = B$) if A implies B ($A \subset B$) and B implies A ($B \subset A$), i.e., if, for each realization of an experiment, both events A and B occur or do not occur simultaneously.

Example 24.8. If events $A = \{1, 2, 3\}$ and $B = \{3, 2, 1\}$ are given, then events A and B are equivalent or $A = B$.

Venn diagrams are useful for visualizing the relationships among sets or events (fig. 24.1).



(a) (b) (c)

Fig. 24.1. **The intersection $A \cap B$ (a), the union $A \cup B$ (b) and the difference $A \setminus B$ (c) of events A and B**

The ***axiomatic definition of probability***. Let a space of elementary events (a sample space) Ω be given and such single number $P(A)$ (the probability of an event A) corresponds to each event $A \subset \Omega$, that:

1) $P(A) \geq 0$;

2) for each pair of mutually exclusive events $A, B \subset \Omega$ the equality $P(A \cup B) = P(A) + P(B)$ takes place;

3) $P(\Omega) = 1$.

Then we say, that the probability is defined on events of Ω, and the number $P(A)$ is called the ***probability of an event*** A .

Let's suppose that $\Omega = \{\omega_1, \omega_2, ..., \omega_n\}$ is a finite space, where ω_1, $\omega_2, ..., \omega_n$ are the simplest indivisible mutually exclusive outcomes of an experiment or they are called elementary events. To each elementary event

$\omega_i \in \Omega\,(i=1, 2, ..., n)$ there is a corresponding number $p(\omega_i)$, called the probability of the elementary event ω_i. Thus a real function satisfying the following two conditions is defined on the space Ω:

1) nonnegativity condition: $p(\omega_i) \geq 0$ for any $\omega_i \in \Omega\,(i=1, 2, ..., n)$;

2) normalization condition: $\sum\limits_{i=1}^{n} p(\omega_i) = 1$.

The probability $P(A)$ of an event A for any set $A \subset \Omega$ is defined to be the sum of probabilities of the elementary events that form A, i.e.

$$P(A) = \sum_{\omega_i \in A} p(\omega_i).$$

This pair (a space of elementary events Ω and a real function P) thus defined is called a ***finite discrete probability space***.

24.3. Rules of a sum of incompatible events and a product of compatible events. Inclusion-exclusion principle

At solving problems of probability theory one can use the *following rules*:

1. The ***rule of sum*** is an intuitive principle stating that if there are a possible outcomes for an event A (or ways to do something) and b possible outcomes for another event B (or ways to do another thing) and two events can't both occur (or the two things can't be done) (A and B are mutually exclusive or incompatible events) then there are $a+b$ total possible outcomes for the events A and B (or total ways to do one of the things); formally, the sum of sizes of two incompatible sets is equal to the size of their union, i.e. $|A| + |B| = |A \cup B|$.

Example 24.9. A woman has decided to shop at one store today, either in the north part of town or the south part of town. If she visits the north part of town, she will either shop at a mall, a furniture store, or a jewelry store (3 ways). If she visits the south part of town then she will either shop at a clothing store or a shoe store (2 ways). Let A be the woman visiting the north part of town and B be the woman visiting the south part of town, i.e. $|A| = 3$ and $|B| = 2$.

Thus there are $|A \cup B| = |A| + |B| = 3 + 2 = 5$ possible shops the woman could end up shopping at today.

2. The ***rule of product*** is another intuitive principle stating that if there are a possible outcomes for an event A (or ways of doing something) and b possible outcomes for another event B (or ways of doing another thing) and two events can both occur (or the two things can be done) (A and B are not mutually exclusive or compatible events) then there are $a \cdot b$ total ways of performing both things, i.e. $|A \cap B| = |A| \cdot |B|$.

Example 24.10. When we decide to order pizza, we must first choose the type of crust: thin or deep dish (2 choices or $|A| = 2$). Next, we choose the topping: cheese, pepperoni, or sausage (3 choices or $|B| = 3$). Using the rule of product, you know that there are $|A \cap B| = |A| \cdot |B| = 2 \cdot 3 = 6$ possible combinations of ordering a pizza.

3. ***Inclusion-exclusion principle*** or the *rule of inclusion and exclusion*: the inclusion-exclusion principle relates to the size of the union of multiple sets, the size of each set and the size of each possible intersection of the sets. The smallest example is when there are two sets: the number of elements in the union of the events A and B is equal to the sum of the elements in the events A and B minus the number of elements in their intersection, i.e. $|A \cup B| = |A| + |B| - |A \cap B|$.

Example 24.11. 35 voters were queried about their opinions regarding two referendums. 14 supported referendum 1 and 26 supported referendum 2. How many voters supported both, assuming that every voter supported either referendum 1 or referendum 2 or both?

Solution. Let A be voters who supported referendum 1 and B be voters who supported referendum 2. Then we have $|A \cup B| = 35$, $|A| = 14$ and $|B| = 26$. Using the inclusion-exclusion principle we obtain:

$$|A \cup B| = |A| + |B| - |A \cap B| \qquad \text{or} \qquad |A \cap B| = |A| + |B| - |A \cup B|$$

$$|A \cap B| = 14 + 26 - 35 \qquad \text{or} \qquad |A \cap B| = 5.$$

Recommended bibliography: [1; 2; 4; 5; 7; 10; 11].

Theme 25. A classical definition of a probability and elements of a combinatory analysis. Statistical and geometrical definitions of a probability

25.1. A classical definition of a probability

Let a space of elementary events Ω be given and this space consists of n equally likely elementary outcomes (i.e. total number of outcomes) of the experiment, among which there are m outcomes, favorable for an event A (i.e. number of outcomes an event A can happen), and $\Omega \subset A$. Then the number:

$$P(A) = \frac{m}{n} \tag{25.1}$$

is called the ***probability of an event*** A.

As all events have probabilities between impossible (0) and certain (1), then probabilities are usually written as a fraction, a decimal or sometimes as a percentage. In this lecture probabilities will be written as fractions or decimals.

The probability is the non-dimensional quantity. It can be measured in percent from 0 to 100. For example, $P(A) = \frac{4}{10} = 0.4 = 40\,\%$.

Example 25.1. Suppose the event A we are going to consider is *rolling a die* once and obtaining a 3. The die could land in a total of six different ways. We say that the total number n of outcomes of rolling the die is six, which means there are six ways it could land. The number m of ways of obtaining the particular outcome of A is one.

We can apply the formula (25.1) and find: $P(A) = \frac{1}{6}$.

When we roll a die it has an equal chance of landing on any of the six numbers 1, 2, 3, 4, 5, or 6. These events are called equally likely events.

25.2. A basic notion of a combinatorial analysis

We often compose new sets, systems or sequences from the elements of a given set in a certain way. Depending on the way we do it, we get the notion of permutation, combination and arrangement. The basic problem of combinatorics is to determine how many different choices or arrangements are possible with the given elements (for instance, letters of an alphabet, books of a library, cars on a parking, etc.).

25.2.1. Collection of formulas of combinatorics without repetitions. A ***permutation without repetitions*** is called the number of different permutations of n different elements:

$$P_n = 1 \cdot 2 \cdot \ldots \cdot n = n!$$

Example 25.2. In a classroom 16 students are seated on 16 places. There are $P_{16} = 16!$ different possible permutations.

An ***arrangement without repetitions*** is called an ordering of k elements selected from n different ones, i.e. arrangements are combinations considering the order:

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

Example 25.3. How many different ways are there to choose a chairman, his deputy, and a first and a second assistant for them from 30 participants at an election meeting? This answer is $A_{30}^4 = \frac{30!}{(30-4)!} = 657720$.

A ***combination without repetitions*** is called a choice of k elements from n different elements without considering the order of them:

$$C_n^k = \frac{n!}{k! \cdot (n-k)!}$$

Example 25.4. There are $C_{30}^4 = \frac{30!}{4! \cdot (30-4)!} = 27405$ possibilities to choose an electoral board of 4 persons from 30 participants.

Example 25.5. 7 tickets were drawn among 17 students including 8 girls. What is the probability that there are 4 girls among ticket owners?

Solution. The number of possible ways of distributing 7 tickets among 17 students is equal to the number of combinations of 17 elements taken 7 at a time, i.e. C_{17}^{7}. The number of the selection of 4 girls from 8 is C_8^4. Each group of 4 can be connected with each group of 3 of 9 boys. The number of such groups of 3 is C_9^3. The number of results of distributing 7 tickets including 4 tickets for girls and 3 for boys, is $C_8^4 \cdot C_9^3$. Then the probability is:

$$P(A) = \frac{C_8^4 \cdot C_9^3}{C_{17}^{7}}.$$

25.2.2. Collection of the formulas of combinatorics with repetitions. If for different elements k out of n elements with replacement, no subsequent ordering is performed (i.e., each of the n elements can occur 1, 2, ..., or k times in any combination), then one speaks of ***combinations with repetitions***. The number $\overline{C}_n^k$ of all distinct combinations with repetitions of n elements taken k at a time is given by the formula:

$$\overline{C}_n^k = C_{n+k-1}^k.$$

Example 25.6. Consider the set of elements 1, 2, 3 ($n = 3$). Take $k = 2$ elements, there are $\overline{C}_3^2 = C_{3+2-1}^2 = C_4^2 = \frac{4!}{2! \cdot (4-2)!} = 6$ combinations with repetitions $[(1, 2), (1, 3), (2, 3), (1, 1), (2, 2), (3, 3)]$.

If for different elements k out of n elements with replacement, the chosen elements are ordered in some way, then one speaks of ***arrangements with repetitions***. The number $\overline{A}_n^k$ of distinct arrangements with repetitions of n elements taken k at a time is given by the formula:

$$\overline{A}_n^k = n^k.$$

Example 25.7. Consider the set of elements 1, 2, 3 ($n = 3$). Take $k = 2$ elements, there are $\overline{A}_3^2 = 3^2 = 9$ arrangements with repetitions $[(1, 2), (1, 3), (2, 3), (2, 1), (3, 1), (3, 2), (1, 1), (2, 2), (3, 3)]$.

Let's suppose that a set of n elements contains k distinct elements, of which the first occurs n_1 times, the second occurs n_2 times, ..., and the k-th occurs n_k times, $n_1 + n_2 + \ldots + n_k = n$. Permutations of n elements of this set are called ***permutations with repetitions on*** n ***elements***. The number $P_n(n_1, n_2, \ldots, n_k)$ of permutations with repetitions on n elements is given by the formula:

$$P_n(n_1, n_2, \ldots, n_k) = \frac{n!}{n_1! \cdot n_2! \cdot \ldots \cdot n_k!}.$$

Example 25.8. If there are two letters a and one letter b, the number of permutations with repetitions out of 3 elements and composition of letters 2, 1 equals $P_3(2, 1) = \frac{3!}{2! \cdot 1!} = 3$ $[(a, a, b), (a, b, a), (b, a, a)]$.

25.3. A geometric definition of a probability. A statistical definition of a probability and its properties

A ***statistical definition of a probability*** of an event A. Let A be an event belonging to the sample space Ω, i.e. $A \supset \Omega$, of an experiment. If the event A occurred n_A times while we repeated the experiment n times, then n_A is called the ***frequency***, and $\frac{n_A}{n}$ is called the relative frequency of the event A.

A ***geometric definition of a probability*** of an event A. Let Ω be a set of a positive finite measure $\mu(\Omega)$ and consist of all measurable (i.e. having a measure) subsets $A \supset \Omega$. The ***geometric probability*** of an event A is defined to be ratio of the measure of A to that of Ω, i.e.

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

The notion of geometric probability is not invariant under transformations of the sample space Ω and depends on how the measure $\mu(A)$ is introduced. As measures $\mu(A)$ and $\mu(\Omega)$ we can use different geometric measures, for example, lengths, areas or volumes.

Example 25.9. A point is randomly thrown into a disk of radius $R=1$. Find the probability of the event that the point lands in the disk of radius $r=\frac{1}{2}$ centered at the same point.

Solution. Let A be the event that the point lands in the smaller disk.

We find the probability $P(A)$ as the ratio of the area of the smaller disk to that of the larger disk:

$$P(A)=\frac{\pi r^2}{\pi R^2}=\frac{r^2}{R^2}=\frac{(1/2)^2}{1^2}=\frac{1}{4}.$$

25.4. Different types of events. Properties of probability

An event A is said to be ***impossible*** if it cannot occur for any realization of the experiment. Obviously, the impossible event does not contain any elementary outcome and hence should be denoted by the symbol Ø. Its probability is zero, i.e. $P(A)=0$.

Example 25.10. Let's roll a die and obtain a score of 7 (the event A). It's an impossible event, then $P(A)=0$.

Property 1. The probability of an impossible event is 0, i.e. $P(A)=0$.

An event A is said to be ***sure*** if it is equivalent to the space of elementary events Ω, i.e. $A=\Omega$, or it happens with probability 1.

Example 25.11. Let's roll dice and obtain a score less than 13 (the event A). It's a sure event or a space of elementary events Ω, because it consists of all possible outcomes of Ω*.* Then $P(A)=1$.

Property 2. The probability of a sure (certain) event is 1, i.e. $P(A)=1$.

Property 3. The probability of a space of elementary events Ω is 1, i.e. $P(\Omega)=1$.

Property 4. All probabilities that lie between zero and one are inclusive, i.e. $0\le P(A)\le 1$.

The event that A doesn't occur is called the ***complement*** of A, or the ***complementary event***, and is denoted by $\overline{A}$. The elementary outcomes of $\overline{A}$ are the elementary outcomes that don't belong to the event A.

Property 5. The probability of the event $\overline{A}$ opposite to the event A is equal to $P(\overline{A}) = 1 - P(A)$.

From this property we can obtain that $P(A) + P(\overline{A}) = 1$ for complementary events $\overline{A}$ and A and explain them in the next example.

Example 25.12. Helen rolls a die once. What is the probability she rolls an even number or an odd number?

Solution. The event of rolling an even number (A) and the event of rolling an odd number (B) are mutually exclusive events, because they both cannot happen at the same time, so we add the probabilities. In addition, these two events make up all the possible outcomes, so they are complementary events, i.e. B is $\overline{A}$. Let's write: $P(A) + P(B) = \frac{3}{6} + \frac{3}{6} = 1$

The events A and B are called ***equally likely events***, if $P(A) = P(B)$.

Property 6. Probabilities of equally likely events A and B are equal, i.e. $P(A) = P(B)$.

Example 25.13. When we roll a die it has an equal chance $\frac{1}{6}$ of landing on any of the six numbers 1, 2, 3, 4, 5, or 6. These events are called equally likely events.

Property 7. Nonnegativity: $P(A) \geq 0$ for any $A \subset \Omega$.

Property 8. For each $A \subset \Omega$ the inequality $P(A) \leq 1$ takes place.

Property 9. If an event A implies B, i.e. $A \subset B$, then $P(A) \leq P(B)$.

25.5. Probability addition theorems

The events are called ***compatible*** (***mutually exclusive***) if they can occur together in the same experiment.

The events are called ***incompatible*** (***not mutually exclusive***) if they cannot occur together in the same experiment.

Probability addition theorem for incompatible events. The probability of realization of at least one of two events A and B is given by the formula:

$$P(A+B) = P(A\,or\,B) = P(A) + P(B), \qquad (25.2)$$

where A and B are incompatible events.

The probability of such events are explained in the following example.

Example 25.14. Ann rolls a die once. a) What is the probability she rolls a 3 and a 6? b) What is the probability she rolls a 3 or a 6?

Solution. a) When one die is rolled, the event A of rolling a 3 and the event B of rolling a 6 are events that cannot both happen at the same time, and are called mutually exclusive events. So the probability of rolling a 3 and a 6 is impossible on one roll of a die, and equal to zero, i.e. $P(A and B) = 0$.

b) The probability of rolling a 3 (A) or a 6 (B) is also a mutually exclusive event and is calculated by the formula (25.2):

$$P(A+B) = P(A) + P(B) = \frac{1}{6} + \frac{1}{6} = \frac{2}{6} = \frac{1}{3}.$$

Probability addition theorem for compatible events. The probability of realization of at least one of two events A and B is given by the formula

$$P(A+B) = P(A \, or \, B) = P(A) + P(B) - P(A \cap B), \qquad (25.3)$$

where A and B are compatible events.

Example 25.15. Ann rolls a die once. What is the probability she rolls a prime number or an odd number?

Solution. When one die is rolled, the event A of rolling a prime or the event B of rolling an odd number are events that can both happen at the same time, and they are compatible events. Then the probability of A or B is calculated by the formula (25.3).

We have $A = \{2, 3, 5\}$, $B = \{1, 3, 5\}$ and obtain $P(A) = \frac{3}{6}$, $P(B) = \frac{3}{6}$.

We find $A \cap B = \{3, 5\}$, $P(A \cap B) = \frac{2}{6}$ and use the formula (25.3):

$$P(A+B) = P(A) + P(B) - P(A \cap B) = \frac{3}{6} + \frac{3}{6} - \frac{2}{6} = \frac{4}{6} = \frac{2}{3}.$$

Recommended bibliography: [2; 4; 6; 7; 10; 11].

Theme 26. Conditional probability and a notion of an event independence. Formulas of a total probability and Bayes

26.1. Conditional probability and theorem of a product for dependent events. Theorem of a product for independent events

The events are called ***independent*** if the occurrence of one of them does not change the probability of the occurrence of the other one.

The events are called ***dependent*** if the probability of each of them is changed in connection with the occurrence or nonoccurrence of the other one.

Multiplication theorem for independent events. When the outcome of one event has no effect on the outcome of another event, we say that the two events are independent events. To obtain the probability of independent events we multiply the probabilities of the separate events, i.e.

$$P(A\cdot B)=P(A\,and\;B)=P(A)\cdot P(B), \tag{26.1}$$

where A and B are independent events

Example 26.1. A coin is tossed and a die is rolled. What is the probability of obtaining a head and a prime number?

Solution. The result of tossing a coin cannot possibly affect the outcome of rolling a die. In other words, if the coin landed as a head, it would not affect the way the die would land. Then the outcomes are independent events.

The probability of A (tossing a head) is $\frac{1}{2}$, i.e. $P(A)=\frac{1}{2}$, and the probability of B (rolling a prime number with a die) is $\frac{3}{6}$, i.e. $P(B)=\frac{3}{6}$, because there are three numbers 2, 3, and 5 that are prime. Let's use the formula (3.1): $P(A\cdot B)=P(A)\cdot P(B)=\frac{1}{2}\cdot\frac{3}{6}=\frac{1}{4}$.

Multiplication theorem for dependent events. If A and B are dependent events, then:

$$P(A\cdot B)=P(A\,and\;B)=P(A)\cdot P(B|A)=P(B)\cdot P(A|B), \tag{26.2}$$

where $P(B|A)$ or $P_A(B)$ is called the ***conditional probability*** of the event B

given the event A (it means the probability that the event B will occur given that the event A has already occurred) and $P(A|B)$ is the conditional probability of the event A given the event B.

Example 26.2. There are 3 nonstandard electric bulbs among 50 electric ones. What is the probability that 2 electric bulbs taken at a time are nonstandard?

Solution. The probability of the event A that the first bulb is nonstandard equals $\frac{3}{50}$. The probability of the second bulb is nonstandard (the event B) on conditions that the first bulb is nonstandard (the event A) equals $\frac{2}{49}$, because the total number of bulbs and the number of nonstandard bulbs decreased by 1.

According to the formula (26.2) we have

$$P(A\cdot B)=P(A)\cdot P(B|A)=\frac{3}{50}\cdot\frac{2}{49}\approx 0.0024.$$

Two random events A and B are said to be ***independent*** if the conditional probability of A given B coincides with the unconditional probability of A, i.e. $P(A|B)=P(A)$.

A ***conditional probability*** from the formula (26.2) is expressed as:

$$P(B|A)=\frac{P(A\cdot B)}{P(A)}. \qquad (26.3)$$

Example 26.3. The probability that it is Friday and that a student is absent is 0.03. Since there are 5 school days in a week, the probability that it is Friday is 0.2. What is the probability that a student is absent given that today is Friday?

Solution. Let’s denote that it is Friday as the event A and a student is absent as the event B. Then the event that a student is absent given that today is Friday is denoted by $B|A$. Let’s find $P(B|A)$ using the formula (26.3):

$$P(B|A)=\frac{P(A\cdot B)}{P(A)}=\frac{0.03}{0.2}=0.15.$$

26.2. A complete group of events

Events $A_1, A_2, \ldots, A_n$ are called ***pairwise independent*** if every possible pair of these events is independent, i.e. $P(A_i \cap A_j) = P(A_i) \cdot P(A_j)$ for any i, j $(i \neq j)$.

One says that events $A_1, A_2, \ldots, A_n$ form ***a complete group*** of pairwise incompatible events (or mutually exclusive), if exactly one of them necessarily occurs for each realization of the experiment and no other event can occur.

If events $A_1, A_2, \ldots, A_n$ form a complete group of pairwise incompatible events, then $P(A_1) + P(A_2) + \ldots + P(A_n) = 1$.

For example, two opposite events A and $\overline{A}$ form a complete group of incompatible events.

Example 26.4. Let the probability that the shooter scores 10 points, when hitting the target, equals 0.4, 9 points – 0.2, 8 points – 0.2, 7 points – 0.1, 6 points and less – 0.1. What is the probability that the shooter scores no less then 9 points by one shot?

Solution. Let A_1 be the shooter scoring 10 points, A_2 be the shooter scoring 9 points, A_3 be the shooter scoring 8 points, A_4 be the shooter scoring 7 points, A_5 be the shooter scoring 6 points and less.

These events form the complete group of pairwise incompatible events, i.e. $P(A_1) + P(A_2) + P(A_3) + P(A_4) + P(A_5) = 1$.

Let C be the shooter scoring no less then 9 points by one short.

The required event will occur (mark it C) if the shooter scores either 9 (the event A_2) or 10 points (the event A_1). The events A_2 and A_1 are incompatible. Thus, $P(C) = P(A_1) + P(A_2) = 0.2 + 0.4 = 0.6$.

26.3. A notion of a pairwise independence of random events. An independence in a totality

A pairwise independent collection of events $A_1, A_2, \ldots, A_n$ is called a set of events any two of which are independent.

Any collection of mutually independent events is ***pairwise independent***.

Let events $A_1, A_2, \ldots, A_n$ be independent, A is at least one of n events occurs in the experiment. Then $\overline{A}$ is this event that no one of n events occurs in the experiment, i.e. $\overline{A} = \overline{A_1} \cdot \overline{A_2} \cdot \ldots \cdot \overline{A_n}$. The events A and $\overline{A}$ form a complete group of incompatible events, therefore,

$$P(A) = 1 - P(\overline{A}) = 1 - P(\overline{A_1}) \cdot P(\overline{A_2}) \cdot \ldots P(\overline{A_n}). \qquad (26.4)$$

This formula (26.4) is called the ***probability that at least one of*** n ***events occurs***.

Let's denote $P(A_1) = p_1$, $P(\overline{A_1}) = 1 - p_1 = q_1$, $\ldots$, $P(A_n) = p_n$, $P(\overline{A_n}) = 1 - p_n = q_n$ and transform the formula (26.4):

$$P(A) = 1 - P(\overline{A_1}) \cdot P(\overline{A_2}) \cdot \ldots P(\overline{A_n}) = 1 - q_1 \cdot q_2 \cdot \ldots \cdot q_n. \qquad (26.5)$$

Example 26.5. Three students are going to take the exam. The probability that the first student passes it equals 0.9, the second one is 0.75, the third one is 0.6. What is the probability that at least one of three students passes the exam?

Solution. Let A_1 be the event that the first student passes the exam, A_2 be the event that the second student passes the exam, A_3 be the event that the third one does it.

Each student can pass the exam or not. Then $P(A_1) = p_1 = 0.9$, $P(\overline{A_1}) = 1 - p_1 = q_1 = 0.1$, $P(A_2) = p_2 = 0.75$, $P(\overline{A_2}) = 1 - p_2 = q_2 = 0.25$, $P(A_3) = p_3 = 0.6$, $P(\overline{A_3}) = 1 - p_3 = q_3 = 0.4$.

Events A_1, A_2, A_3 are independent. If the event A is at least one of three students passes the exam, then the complementary event $\overline{A}$ (not A) is no student passes the exam (it means $\overline{A_1} \cdot \overline{A_2} \cdot \overline{A_3}$).

Let's use the formula (26.5) and obtain:

$$P(A) = 1 - q_1 \cdot q_2 \cdot q_3 = 1 - 0.1 \cdot 0.25 \cdot 0.4 = 1 - 0.01 = 0.99.$$

If all events $A_1, A_2, \ldots, A_n$ have equal probability, i.e. $P(A_1) = P(A_2) = \ldots = P(A_n) = p$, then $P(\overline{A_1}) = P(\overline{A_2}) = \ldots = P(\overline{A_n}) = 1 - p = q$ and from the formula (26.5) we have:

$$P(A) = 1 - q^n. \tag{26.6}$$

Let's define the necessary number of trials (n) with the given reliability P no less than $P(A)$, i.e. $P(A) \geq P$, using the formula (26.6):

$$P(A) = 1 - q^n \geq P \quad \text{or} \quad 1 - (1-p)^n \geq P$$

or

$$(1-p)^n \leq 1 - P.$$

Let's take a natural logarithm of both parts of this inequality:

$$n \cdot \ln(1-p) \leq \ln(1-P).$$

Hence

$$n \leq \frac{\ln(1-P)}{\ln(1-p)}.$$

26.4. The formula of a total probability

Let's suppose that a complete group of pairwise incompatible events $H_1, H_2, \ldots, H_n$ is given and the unconditional probabilities $P(H_1)$, $P(H_2)$, $\ldots$, $P(H_n)$, as well as the conditional probabilities $P(A|H_1)$, $P(A|H_2)$, $\ldots$, $P(A|H_n)$ of an event A, are known. Then the probability of A can be determined by the ***total probability formula***

$$P(A) = \sum_{k=1}^{n} P(H_i) \cdot P(A|H_i). \tag{26.7}$$

Each of the events $H_1, H_2, \ldots, H_n$ is called ***hypothesis***.

$P(H_i)$ is called ***a priori probability*** (***premature probability***).

Example 26.6. Three machines produce the same type of product in a factory. The first one gives 200 articles, the second one does 300 articles and the third one does 500 articles. It is known that the first machine produces 1 % of defective articles, the second one does 2 %, the third one does 4 %. What is the probability that an article selected randomly from the total production will be defective?

Solution. Let A be the event that the chosen article is defective.

Let's consider the following complete group of events (hypotheses): H_1 denotes the event that the randomly selected article is made by the first machine, H_2 denotes the event that the randomly selected article is made by the second machine, H_3 denotes the event that the randomly selected article is made by the third machine.

Let's find their probabilities: $P(H_1)=\frac{200}{200+300+500}=\frac{200}{1000}=0.2$, $P(H_2)=\frac{300}{1000}=0.3$, $P(H_3)=\frac{500}{1000}=0.5$.

Since events H_1, H_2 and H_3 form the complete group, then $P(H_1)+P(H_2)+P(H_3)=0.2+0.3+0.5=1$.

Let's define conditional probabilities $P(A|H_1)$, $P(A|H_2)$, $P(A|H_3)$:

$$P(A|H_1)=0.01,\ P(A|H_2)=0.02,\ P(A|H_3)=0.04.$$

Here $A|H_1$ is a defective article produced by the first machine, $A|H_2$ is a defective article produced by the second machine, $A|H_3$ is a defective article produced by the third machine.

Let's use the total probability formula (26.7) and find:

$$P(A)=P(H_1)\cdot P(A|H_1)+P(H_2)\cdot P(A|H_2)+P(H_3)\cdot P(A|H_3)=$$

$$=0.2\cdot 0.01+0.3\cdot 0.02+0.5\cdot 0.04=0.028.$$

We have 2.8 % of defective articles from the total production.

26.5. Bayes' formula

If it is known that the event A has occurred but it is unknown which of the events $H_1, H_2, \ldots, H_n$ has occurred, then ***Bayes' formula*** is used:

$$P(H_k|A)=\frac{P(H_k)\cdot P(A|H_k)}{P(A)},\ k=1,2,\ldots,n \qquad (26.8)$$

and $P(H_1|A)+P(H_2|A)+\ldots+P(H_n|A)=1.$

where $P(H_i|A)$ is called ***a posteriori probability*** *(final probability)*.

Example 26.7. Let's use the condition of example 26.6 and solve the following problem. It is known that a selected article is defective. What is the probability that this article was made by the second machine?

Solution. The desired probability of the event $H_2|A$ (the selected article was made by the second machine under condition that it is known that it is defective) is determined by Bayes' formula (26.8):

$$P(H_2|A)=\frac{P(H_2)\cdot P(A|H_2)}{P(A)}=\frac{0.3\cdot 0.02}{0.028}=\frac{6}{28}=\frac{3}{14}.$$

Recommended bibliography: [2; 5; 6; 7; 10; 11].

Theme 27. A model of repeated trials of Bernoulli's scheme. Theorems of de Moivre–Laplace and Poisson as investigations of an asymptotic behavior of binomial distribution

27.1. Repeated independent trials. Bernoulli's scheme. A distribution of a number of successes in a set of independent stochastic experiments. A binomial distribution

Trials in which events occurring in distinct trials are independent are said to be ***independent***. Here the probability of each event A of the form $A=A_1\cdot A_2\cdot\ldots\cdot A_n$ is defined as $P(A)=P(A_1)\cdot P(A_2)\cdot\ldots\cdot P(A_n)$.

Let independent events occur in n independent trials. In each trial the event A can occur on can't occur.

A sequence of n independent trials is also called a ***Bernoulli scheme****.*

In this case, some event A occurs with probability $p=P(A)$ (the probability of "success") and does not occur with probability $q=P(\overline{A})=1-$

$-P(A)=1-p$ (the probability of "failure") in each trial.

If k is the number of occurrences of the event A (the number of "successes") in n independent Bernoulli trials, then the probability that A occurs exactly k times is given by the formula:

$$P(k)=P_n(k)=C_n^k p^k q^{n-k}, \qquad (27.1)$$

where $C_n^k=\dfrac{n!}{k!\cdot(n-k)!}$ is a combination of n things taken k at a time.

This relation is called the ***Bernoulli formula (binomial distribution)****.*

The probability that the event occurs at least m times in n independent trials is calculated by the formula:

$$P_n(k\geq m)=\sum_{k=m}^{n}C_n^k p^k q^{n-k}=1-\sum_{k=0}^{m-1}C_n^k p^k q^{n-k}.$$

The probability that the event occurs at least once in n independent trials is calculated by the formula:

$$P_n(k\geq 1)=1-q^n.$$

The probability that the event A occurs no less than k_1 and no more than k_2 times ($k_1<k_2$) satisfies the relation:

$$P_n(k_1\leq k\leq k_2)=P_n(k_1)+P_n(k_1+1)+\ldots+P_n(k_2)=$$

$$=C_n^{k_1}p^{k_1}q^{n-k_1}+\ldots+C_n^{k_2}p^{k_2}q^{n-k_2}.$$

Example 27.1. The probability of a train's arrival at a station on time is equal to 0.8. What is the probability that out of 4 expecting trains 2 trains will arrive on time?

Solution. Let A be a train arriving at a station on time, $P(A)=p=0.8$. Then $\overline{A}$ is a train that doesn't arrive at a station on time and $q=P(\overline{A})=1-P(A)=1-0.8=0.2$. Here $n=4<30$, $k=2$.

According to Bernoulli formula (27.1) we have

$$P_4(2)=C_4^2\cdot 0.8^2\cdot 0.2^{4-2}=\frac{4!}{2!\cdot(4-2)!}\cdot 0.8^2\cdot 0.2^2=6\cdot 0.64\cdot 0.04=0.1536.$$

27.2. The most probable number of successes and its probability

The number k_0 of occurrences of the event A in the independent trials is called the ***most probable number*** if the probability of the event occuring such number k_0 times is maximum (the largest value).

Let the event A occur with probability $p=P(A)$ and do not occur with probability $q=P(\overline{A})=1-P(A)=1-p$ in the trial. Then the ***most probable number*** k_0 is defined by inequality:

$$np-q\le k_0\le np+p, \tag{27.2}$$

where k_0 is a whole number.

Example 27.2. The probability of finding a mistake on a book page is equal to 0.002. 500 pages are checked. Find the most probable number of pages with mistakes.

Solution. Let A be finding a mistake on a book page, $P(A)=p=0.002$. Then $\overline{A}$ is lack of a mistake on a book page and $q=P(\overline{A})=1-P(A)=1-0.002=0.998$.

According to (27.2) we have

$$500\cdot 0.002-0.998\le k_0\le 500\cdot 0.002+0.002$$

or

$$1-0.998\le k_0\le 1+0.002$$

or

$$0.002\le k_0\le 1.002.$$

Then $k_0=1$.

27.3. Approximate methods of calculating binomial probabilities and their accuracy. Limit theorems for Bernoulli process

It is very difficult to use Bernoulli's formula for large n and k. In this case, one has to use approximate formulas for calculating $P_n(k)$ with desired accuracy.

27.3.1. Local theorem of de Moivre–Laplace. Suppose that the number of independent trials increases unboundedly ($n \to \infty$ or n approaches infinity) and the probability $p = const$, $0 < p < 1$, then the probability $P_n(k)$ that A occurs exactly k times out of n satisfies the limit relation

$$P_n(k) \approx \frac{1}{\sqrt{npq}} \varphi\left(\frac{k - np}{\sqrt{npq}}\right), \tag{27.3}$$

where the limit expression $\varphi(x)$ is Laplace differential function or the probability density of the standard normal distribution, i.e. $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$.

This function is even, i.e. $\varphi(-x) = \varphi(x)$.

The function value ($x < 4$) is defined by Laplace differential function table (table 1, appendix A). For the values $x > 4$ $\varphi(x) \approx 0$.

Example 27.3. The probability of the birth of a boy is equal to 0.51. Find the probability that among 200 newborns there will be the same number of boys and girls.

Solution. Let A be the birth of a boy, $P(A) = p = 0.51$. Then $\overline{A}$ is the birth of a girl and $P(\overline{A}) = 1 - P(A) = 1 - 0.51 = 0.49$. Here $n = 200$, $k = 100$.

According to the local theorem of de Moivre–Laplace (27.3) we have

$$P_{200}(100) \approx \frac{1}{\sqrt{200 \cdot 0.51 \cdot 0.49}} \varphi\left(\frac{100 - 200 \cdot 0.51}{\sqrt{200 \cdot 0.51 \cdot 0.49}}\right) = \frac{\varphi(-0.28)}{7.07}.$$

The function $\varphi(x)$ is even, then we obtain $\varphi(-0.28) = \varphi(0.28)$. Let's apply Laplace differential function table (appendix A) and have $\varphi(0.28) = 0.3836$. Let's substitute this value into the previous formula and obtain:

$$P_{200}(100) \approx \frac{\varphi(-0.28)}{7.07} = \frac{\varphi(0.28)}{7.07} = \frac{0.3836}{7.07} \approx 0.0543.$$

27.3.2. Integral theorem of de Moivre-Laplace. Let's suppose that $n \to \infty$ and the probability $p = const$, $0 < p < 1$, then the probability $P_n(k)$ that A occurs no less than k_1 and no more than k_2 times ($k_1 < k_2$) satisfies the limit relation:

$$P_n(k_1 \le k \le k_2) \approx \Phi\left(\frac{k_2 - np}{\sqrt{npq}}\right) - \Phi\left(\frac{k_1 - np}{\sqrt{npq}}\right), \tag{27.4}$$

where the limit expression $\Phi(x)$ is Laplace integral function or the cumulative distribution function of the standard normal distribution, i.e.:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{x^2}{2}} dx.$$

This function is odd, i.e. $\Phi(-x) = -\Phi(x)$.

The function value ($x < 4$) is defined by Laplace integral function table (table 2, appendix B). For values $x > 4$ $\Phi(x) \approx 0.5$.

Example 27.4. The probability of the birth of a girl is equal to 0.49. Find the probability that among 200 newborns there will be from 95 to 110 girls.

Solution. Let A be the birth of a girl, $P(A) = p = 0.49$. Then $\overline{A}$ is the birth of a boy and $q = P(\overline{A}) = 1 - P(A) = 1 - 0.49 = 0.51$.

Here $n = 200$, $k_1 = 95$, $k_2 = 110$.

According to integral theorem of de Moivre–Laplace (27.4) we have

$$P_{200}(95 \le k \le 110) \approx \Phi\left(\frac{110 - 200 \cdot 0.49}{\sqrt{200 \cdot 0.49 \cdot 0.51}}\right) - \Phi\left(\frac{95 - 200 \cdot 0.49}{\sqrt{200 \cdot 0.49 \cdot 0.51}}\right) =$$

$$= \Phi\left(\frac{12}{7.07}\right) - \Phi\left(\frac{-3}{7.07}\right) = \Phi(1.70) - \Phi(-0.42).$$

The function $\Phi(x)$ is odd, then $\Phi(-0.42) = -\Phi(0.42)$. Let's apply Laplace integral function table (appendix B) and have $\Phi(1.70) = 0.4554$ and $\Phi(0.42) = 0.1628$. Let's substitute these values into the previous formula and obtain:

$$P_{200}(95 \le k \le 110) \approx 0.4554 + 0.1628 = 0.6182.$$

27.3.3. Poisson theorem. If the number of independent trials increases unboundedly ($n \to \infty$) and the probability p simultaneously decays ($p \to 0$) so that their product np is a constant ($np = \lambda = const$), then the probability $P_n(k)$ satisfies the limit relation:

$$P_n(k) \approx \frac{\lambda^k}{k!} e^{-\lambda}. \qquad (27.5)$$

The probability that the event A occurs no less than k_1 and no more than k_2 times ($k_1 < k_2$) satisfies the relation:

$$P_n(k_1 \le k \le k_2) = P_n(k_1) + P_n(k_1 + 1) + ... + P_n(k_2) = \frac{\lambda^{k_1}}{k_1!} e^{-\lambda} + ... + \frac{\lambda^{k_2}}{k_2!} e^{-\lambda}.$$

Example 27.5. The probability of finding a mistake on a book page is equal to 0.002. 1000 pages are checked. Find the probability that there is a mistake on 3 pages.

Solution. Let A be finding a mistake on a book page, $P(A) = p = 0.002$. Then $\overline{A}$ is lack of a mistake on a book page and $q = P(\overline{A}) = 1 - P(A) = 1 - 0.002 = 0.998$. Here $n = 1000$, $k = 3$. Then $\lambda = 1000 \cdot 0.002 = 2$.

According to Poisson formula (27.5) we have:

$$P_{1000}(3) \approx \frac{\lambda^3}{3!} e^{-\lambda} = \frac{2^3}{3!} e^{-2} = \frac{8}{6} \cdot (2.72)^{-2} \approx 0.1802.$$

27.4. Probability of deviation of relative frequency from the probability

Let some event A occur with probability $p = P(A)$, $0 < p < 1$ and don't occur with probability $q = P(\overline{A}) = 1 - P(A) = 1 - p$ in each of n independent trials.

It is necessary to define the probability of deviation of relative frequency from the constant probability, i.e. find the probability of inequality $\left|\frac{m}{n}-p\right|\le\varepsilon$.

Then the probability of an absolute value of deviation of relative frequency from its constant probability less than or equal to ε equals $2\Phi\left(\varepsilon\sqrt{\frac{n}{pq}}\right)$, i.e.:

$$P\left(\left|\frac{m}{n}-p\right|\le\varepsilon\right)\approx 2\Phi\left(\varepsilon\sqrt{\frac{n}{pq}}\right),$$

where $\frac{m}{n}$ is a relative frequency, p is the constant probability of A, n is the number of trials, ε is an accuracy; $\Phi(x)$ is Laplace integral function (appendix B).

Example 27.6. For defining the level of students' knowledge in the given subject 100 students are given tests. The probability of carrying out a test excellently is 0.1. Find

a) the probability P that the relative frequency deviates from the probability p by the value $\varepsilon=0.01$;

b) the accuracy ε, which probability of deviation of relative frequency from the probability p is $P=0.95$;

c) how many students it is necessary to take that with the accuracy $\varepsilon=0.02$ the probability of deviation of relative frequency from the probability p will be $P=0.9$.

Solution. a) Let's find $P\left(\left|\frac{m}{n}-p\right|\le\varepsilon\right)$. If $p=0.1$, then $q=1-p=$ $=1-0.1=0.9$. Let's substitute:

$$P\left(\left|\frac{m}{n}-0.1\right|\le 0.01\right)\approx 2\Phi\left(0.01\sqrt{\frac{100}{0.1\cdot 0.9}}\right)=2\Phi(0.33)=2\cdot 0.1293=0.2586.$$

b) According to the condition $P\left(\left|\frac{m}{n}-p\right|\le\varepsilon\right)=0.95$. Let's find ε.

Then $2\Phi\left(\varepsilon\sqrt{\frac{n}{pq}}\right)=0.95$ or $\Phi\left(\varepsilon\sqrt{\frac{n}{pq}}\right)=0.95/2$ or $\Phi\left(\varepsilon\sqrt{\frac{n}{pq}}\right)=$

$=0.475$ (appendix B) or $\Phi\left(\varepsilon\sqrt{\frac{n}{pq}}\right)=\Phi(1.96)$ or $\varepsilon\sqrt{\frac{n}{pq}}=1.96$.

Let's substitute: $\varepsilon=1.96\sqrt{\frac{pq}{n}}=1.96\sqrt{\frac{0.1\cdot 0.9}{100}}=0.0588\approx 0.06$.

c) According to the condition $P\left(\left|\frac{m}{n}-p\right|\le\varepsilon\right)=0.9$ and $\varepsilon=0.02$. Let's find n. Then $P\left(\left|\frac{m}{n}-p\right|\le\varepsilon\right)=0.9=2\Phi\left(\varepsilon\sqrt{\frac{n}{pq}}\right)$. Thus $\varepsilon\sqrt{\frac{n}{pq}}=1.65$.

Let's substitute:

$$0.02\sqrt{\frac{n}{0.1\cdot 0.9}}=1.65 \text{ or } n=\frac{1.65^2\cdot 0.1\cdot 0.9}{0.02^2}=612.1\approx 613.$$

Thus, it is necessary to take 613 students.

Recommended bibliography: [5; 6; 7; 10; 12].

Theme 28. Discrete random variables, their distribution laws and numerical characteristics

28.1. A definition of random variables and their classification

A variable is called ***random,*** if it can receive real values with definite probabilities as a result of experiment.

In general, random variables can be discrete or continuous.

The random variable X is called ***discrete***, if such non-negative function exists $P(X=x_i)=p_i$, $i=\overline{1,n}$, $\sum_{i=1}^{n}p_i=1$, which determines the correspondence between the value x_i of the variable X and the probability p_i, that X receives this value.

A random variable is denoted by X, Y, Z and so on and its possible values are denoted by x_i, y_i, z_i For example, if X is a random variable, then its values are $x_1,x_2,\ldots,x_n$ (these values form a complete group of events, therefore $\sum_{i=1}^{n}p_i=1$).

Discrete random variables X and Y are called ***independent random variables***, if the events $X = x_i$ and $Y = y_i$ are independent for arbitrary i and j.

28.2. The distribution law of a discrete random variable

The ***distribution law (row)*** of a discrete random variable is called a set of all its possible values and probabilities which these values possess. It's often written in the form of a table

x_i	x_1	x_2	$\cdots$	x_n
p_i	p_1	p_2	$\cdots$	p_n

Example 28.1. Two balls are drawn in succession without replacement from an urn containing 4 red balls and 3 black balls. The possible outcomes and the values x_i of the random variable X, where X is the number of red balls, are:

Sample space	x_i
RR	2
RB	1
BR	1
BB	0

Let's find the probability of each value x_i: $P(X=0) = P(BB) = \frac{1}{7}$,

$$P(X=1) = P(RB+BR) = \frac{2}{7} + \frac{2}{7} = \frac{4}{7}, \qquad P(X=2) = P(RR) = \frac{2}{7}.$$

Let's write the distribution law of this discrete random variable:

x_i	0	1	2
p_i	$\frac{1}{7}$	$\frac{4}{7}$	$\frac{2}{7}$

Distribution law (row) can be graphically plotted (fig. 5.1). Values of a variable x_i are marked on X-axis, the corresponding probabilities p_i are marked on Y-axis. The obtained points are connected with the help of segments. It results in *a **distribution polygon***.

Example 28.2. Distribution law of a discrete random variable X

x_i	-2	2	6	10	14
p_i	0.05	0.16	0.35	0.31	0.13

is given. Draw a distribution polygon.

Solution. Let’s plot the distribution polygon for the given distribution law (fig. 28.2).

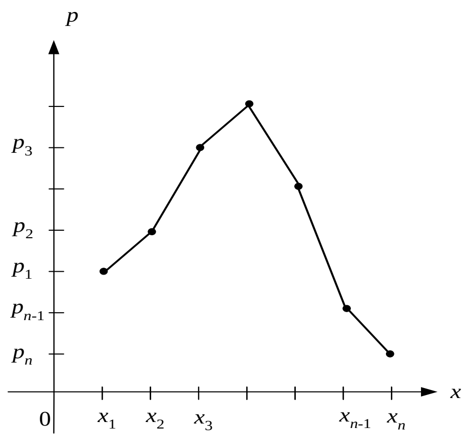


Fig. 28.1. **A distribution polygon**

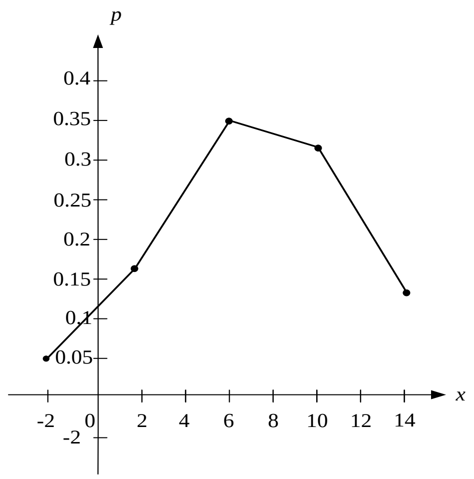


Fig. 28.2. **The distribution polygon**

28.3. The numerical characteristics of distribution

The ***mathematical expectation*** of a discrete random variable X is called a sum of products of possible values x_i, which a variable X is taken, and their corresponding probabilities p_i.

$$M(X) = \sum_{i=1}^{n} x_i \cdot p_i = x_1 \cdot p_1 + x_2 \cdot p_2 + ... + x_n \cdot p_n. \tag{28.1}$$

For existence of the expectation (28.1), it is necessary that the corresponding series converge absolutely.

The expectation is the main characteristic defining the "position" of a random, i.e. the number near which its possible values are concentrated.

General properties of a mathematical expectation:

1. $M(cX) = c \cdot M(X)$, $c \in R$;

2. $M(X+Y) = M(X) + M(Y)$, where X, Y are the discrete random variables;

3. $M(X \cdot Y) = M(X) \cdot M(Y)$ for independent random variables X and Y.

4. $M(\alpha X) = \alpha M(X)$ for any real α.

5. $M(\alpha X + \beta Y) = \alpha M(X) + \beta M(Y)$ for any real α and β.

6. $M(X - Y) = M(X) - M(Y)$.

7. $M(\alpha X - \beta Y) = \alpha M(X) - \beta M(Y)$ for any real α and β.

The ***variance*** of a random variable X is called a mathematical expectation of deviation square of a random variable from its mathematical expectation, i.e.:

$$D(X) = M\left[(X - M(X))^2\right] = \sum_{i=1}^{n} (x_i - M(X))^2 \cdot p_i \qquad (28.2)$$

or *variance* of a random variable X equals a mathematical expectation of its square minus a square of its mathematical expectation, i.e.:

$$D(X) = M(X^2) - [M(X)]^2, \qquad (28.3)$$

where
$$M(X^2) = \sum_{i=1}^{n} x_i^2 \cdot p_i. \qquad (28.4)$$

Properties of a variance:

1. $D(C) = 0$ for any real C.

2. The variance is nonnegative: $D(X) \geq 0$.

3. $D(\alpha X + \beta) = \alpha^2 \cdot D(X)$ for any real α and β.

4. $D(\alpha X) = \alpha^2 \cdot D(X)$ for any real α.

5. $D(X+Y) = D(X) + D(Y)$ for independent random variables X and Y.

6. $D(X-Y)=D(X)+D(Y)$.

7. $D(X \cdot Y)=D(X) \cdot D(Y)+D(X) \cdot M^2(Y)+D(Y) \cdot M^2(X)$.

Root-mean-square deviation *(*or ***standard deviation****)* of a random variable X is the square root of its variance, i.e.:

$$\sigma(X)=\sqrt{D(X)}. \tag{28.5}$$

*A **mode*** of a discrete random variable M_o is a value preceded and followed by values associated with probabilities smaller $P(M_o)$.

*A **median*** of a discrete random variable M_e is the "middle" value. It is the value of X for which $P(X \le x)$ is greater than or equal to 0.5 and $P(X \ge x)$ is greater than or equal to 0.5.

The expectation $M\left((X-a)^k\right)$ is called the k***-th moment*** of a discrete random variable X about a. The moments about zero are usually referred to simply as the moments of a random variable and sometimes they are called ***initial moments***. The k***-th moment*** satisfies the relation:

$$\nu_k=\sum_{i=1}^{n} x_i^k \cdot p_i .$$

If $a=M(X)$ then k*-th* moment of the random variable X about a is called the k***-th central moment****.* It satisfies the relation:

$$\mu_k=\sum_{i=1}^{n}(x_i-M(X))^k \cdot p_i .$$

Example 28.3. Distribution law of a discrete random variable X:

x_i	-2	2	6	10	14
p_i	0.05	0.16	0.35	0.31	0.13

is given. Find numerical characteristics of a discrete random variable X.

Solution. A. Let's calculate $M(X)$ by the formula (28.1):

$$M(X)=\sum_{i=1}^{5} x_i \cdot p_i = x_1 \cdot p_1 + x_2 \cdot p_2 + x_3 \cdot p_3 + x_4 \cdot p_4 + x_5 \cdot p_5 =$$

$$=-2\cdot 0.05+2\cdot 0.16+6\cdot 0.35+10\cdot 0.31+14\cdot 0.13=$$

$$=-0.1+0.32+2.1+3.1+1.82=7.24.$$

B. Let's calculate $D(X)$ by the formulas (28.2) and (28.3), using (28.4):

1) $D(X)=\sum_{i=1}^{5}(x_i - M(X))^2 \cdot p_i = (-2-7.24)^2 \cdot 0.05 + (2-7.24)^2 \cdot 0.16 +$

$$+(6-7.24)^2 \cdot 0.35+(10-7.24)^2 \cdot 0.31+(14-7.24)^2 \cdot 0.13=17.5024;$$

2) $M(X^2) = x_1{}^2 \cdot p_1 + x_2{}^2 \cdot p_2 + x_3{}^2 \cdot p_3 + x_4{}^2 \cdot p_4 + x_5{}^2 \cdot p_5 =$

$$=(-2)^2 \cdot 0.05+2^2 \cdot 0.16+6^2 \cdot 0.3+10^2 \cdot 0.31++14^2 \cdot 0.13=69.92,$$

$$D(X)=69.92-7.24^2=17.5024.$$

C. Let's calculate $\sigma(X)$ by the formula (28.5):

$$\sigma(X)=\sqrt{D(X)}=\sqrt{17.5024}=4.1836.$$

28.4. The distribution function

The probability of the fact that a random variable X receives a value less than x, is called a ***distribution function*** of a random variable X and is marked as $F(x)$:

$$F(x)=P(X<x).$$

General properties of the cumulative distribution function:

1. $F(x)$ is a bounded function, i.e. $0 \le F(x) \le 1$.
2. $F(x)$ is a non-decreasing function for $x \in (-\infty, \infty)$, i.e. if $x_2 > x_1$,

then $F(x_2) \geq F(x_1)$.

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

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

5. $F(x)$ is left continuous; i.e.

$$\lim_{x \to x_0 - 0} F(x) = F(x_0).$$

6. The probability that a random variable X lies in the interval (x_1, x_2) is equal to the increment of its cumulative distribution function on this interval, i.e.

$$P(x_1 < X < x_2) = F(x_2) - F(x_1).$$

Example 28.4. Find the distribution function of the random variable X, which is defined by the distribution law:

x_i	1	2	3	4
p_i	0.7	0.21	0.063	0.027

Find the probability that the random variable X possesses a value less than 1 and more than 4.

Solution. A random variable X doesn't possess the values less than 1, thus for $x \leq 1$ events $X < x$ are impossible and $F(x) = 0$.

If $1 < x \leq 2$, then $F(x) = 0.7$, because X can possess only the value $x = 1$ with the probability $p = 0.7$.

If $2 < x \leq 3$, then $F(x) = 0.7 + 0.21 = 0.91$, because X can possess only the values $x = 1$or $x = 2$ with the probability $p = 0.7$ and $p = 0.21$ (addition theorem for independent events).

If $3 < x \leq 4$, then $F(x) = 0.7 + 0.21 + 0.063 = 0.973$, because X can possess only the values $x = 1$, $x = 2$ or $x = 3$ with the probability $p = 0.7$, $p = 0.21$ and $p = 0.063$ (addition theorem for independent events).

If $x > 4$, then $F(x) = 1$, because the event $X \leq 4$ is reliable and its probability equals 1.

The required integral function is defined by the formula:

$$F(x)=\begin{cases}0, & if\ x\le 1\\ 0.7, & if\ 1<x\le 2\\ 0.91, & if\ 2<x\le 3\\ 0.973, & if\ 3<x\le 4\\ 1, & if\ x>4\end{cases}.$$

A graph of integral function will be of the form:

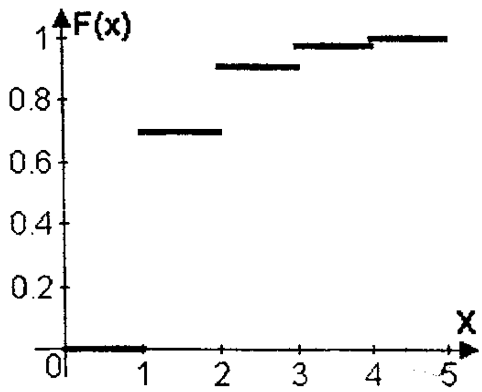


Fig. 28.3. **The graph of integral function**

Let's find the probability that a random variable X possesses a value less than 1 and more than 4, i.e. $P(1<X<4)$:

$$P(1<X<4)=F(4)-F(1)=0.973-0=0.973.$$

28.5. Numerical characteristics of an arithmetic average, a totality of random variables. Properties of numerical characteristics

Let $X_1, X_2, \ldots, X_n$ and $M(X_1), M(X_2), \ldots, M(X_n)$ be random variables and their mathematical expectations, respectively.

Let X be a random variable which equals $X=\dfrac{X_1+X_2+\ldots+X_n}{n}$ (it is the ***arithmetic average***).

According to properties of a mathematical expectation we obtain:

$$M(X)=\frac{M(X_1)+M(X_2)+\ldots+M(X_n)}{n},$$

i.e. the ***mathematical expectation of the arithmetic average*** of n random variables equals the arithmetic average of their mathematical expectations.

Let $D(X_1), D(X_2), \ldots, D(X_n)$ be variances of these random variables and $\max(D(X_1), D(X_2), \ldots, D(X_n)) = D$.

According to the condition $D(X_1) \leq D$, $D(X_2) \leq D$, …, $D(X_n) \leq D$ we obtain:

$$D(X) = \frac{D(X_1) + D(X_2) + \ldots + D(X_n)}{n^2} \leq \frac{n \cdot D}{n^2} = \frac{D}{n},$$

i.e. the variance of the arithmetic average of n random variables whose variances are bounded is n times less than the maximum of variances.

If random variables $X_1, X_2, \ldots, X_n$ are identical distributed, i.e. $M(X_1) = M(X_2) = \ldots = M(X_n) = a$, $D(X_1) = D(X_2) = \ldots = D(X_n) = D$, then:

$$M(X) = \frac{n \cdot a}{n} = a, \qquad M(X) = \frac{n \cdot D}{n^2} = \frac{D}{n},$$

i.e. the mathematical expectation of n identical distributed random variables equals their common mathematical expectation and the variance is n times less than the common variance.

Hence we have:

$$\sigma(X) = \frac{\sigma}{\sqrt{n}},$$

i.e. the root-mean-square deviation of the arithmetic average of n identical distributed random variables equals $\frac{\sigma}{\sqrt{n}}$, where ($\sigma = \sqrt{D}$).

28.6. Basic laws of discrete random distributions and their characteristics

28.6.1. Binomial distribution law. A random variable X has the *binomial distribution* with parameters (n, p) (fig. 28.5) if

$$P_n(x = k) = C_n^k p^k q^{n-k}, \ k = 0, 1, \ldots, n, \qquad (28.6)$$

where $0 < p < 1$, $q = 1 - p$, $n \geq 1$.

The numerical characteristics are given by the formulas:

$$M(X) = np, \tag{28.7}$$

$$D(X) = npq, \tag{28.8}$$

$$\sigma(X) = \sqrt{npq}. \tag{28.9}$$

Binomial distribution law of a discrete random variable is often written in the form of a table:

x_i	0	1	$\ldots$	$n-1$	n
p_i	q^n	$C_n^1 pq^{n-1}$	$\ldots$	$C_n^{n-1} p^{n-1} q$	p^n

The binomial distribution is a model of random experiments consisting of n independent identical Bernoulli trials.

Example 28.5. The probability of passing an exam excellently for each of three students equals 0.4. Make up a distribution law of a number of excellent marks which are got by the students at the exam. Find a mathematical expectation, a variance and a root-mean square deviation of a discrete random variable.

Solution. Let a discrete random variable X be a number of students with the mark "5" (a 5-point system). It has such possible values:

$x_1 = 0$ (no student passed the exam with the mark "5");

$x_2 = 1$ (one student passed the exam with the mark "5");

$x_3 = 2$ (two students passed the exam with the mark "5");

$x_4 = 3$ (three students passed the exam with the mark "5").

Students passing an exam with the mark "5" are independent events. The probabilities of passing an exam of each student are equal, then we use Bernoulli's formula (28.6). According to the condition we have: $n = 3$, $p = 0.4$, $q = 1 - p = 0.6$.

Let's find:

$$x_1 = 0,\ P_3(0) = C_3^0 p^0 q^{3-0} = 1 \cdot 1 \cdot q^3 = 0.6^3 = 0.216;$$

$$x_2 = 1,\ P_3(1) = C_3^1 p^1 q^{3-1} = 3 \cdot p \cdot q^2 = 3 \cdot 0.4 \cdot 0.6^2 = 0.432;$$

$$x_3 = 2,\ P_3(2) = C_3^2 p^2 q^{3-2} = 3 \cdot p^2 \cdot q = 3 \cdot 0.4^2 \cdot 0.6 = 0.288;$$

$$x_4 = 3,\ P_3(3) = C_3^3 p^3 q^{3-3} = 1 \cdot p^3 \cdot q^0 = 1 \cdot 0.4^3 \cdot 1 = 0.064.$$

The distribution law of the discrete random variable X is defined by the table:

x_i	0	1	2	3
p_i	0.216	0.432	0.288	0.064

According to the formulas (28.7) – (28.9) for numerical characteristics we obtain:

$$M(X) = np = 3 \cdot 0.4 = 1.2; \qquad D(X) = npq = 3 \cdot 0.4 \cdot 0.6 = 0.72;$$

$$\sigma(X) = \sqrt{npq} = \sqrt{0.72} \approx 0.85.$$

28.6.2. Geometric distribution law. A random variable X has a *geometric distribution* with parameters p (fig. 28.6) if:

$$P_n(x = k) = pq^k,\ k = 0, 1, 2, \ldots$$

where $0 < p < 1,\ q = 1 - p,\ n \geq 1.$

The numerical characteristics can be calculated by the formulas:

$$M(X) = \frac{1-p}{p} = \frac{q}{p}, \qquad D(X) = \frac{1-p}{p^2} = \frac{q}{p^2}, \qquad \sigma(X) = \sqrt{\frac{q}{p^2}}.$$

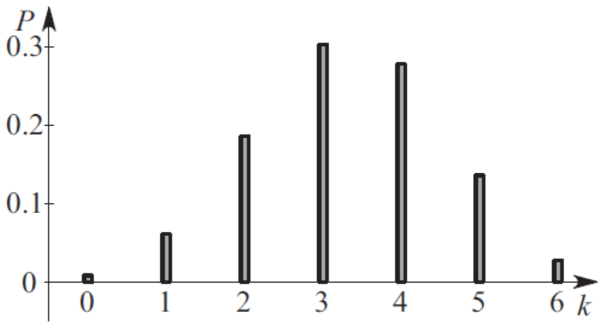


Fig. 28.5. **Binomial distribution law for $p = 0.55$, $n = 6$**

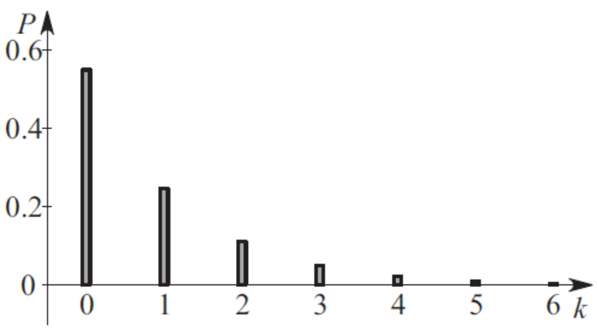


Fig. 28.6. **Geometric distribution law for $p = 0.55$, $n = 6$**

The geometric distribution describes a random variable X equal to the number of failures before the first success in a sequence of Bernoulli trials with probability p of success in each trial.

28.6.3. Hypergeometric distribution law. A random variable X has the *hypergeometric distribution* with parameters (N, p, n) (fig. 28.7) if:

$$P_n(x=k)=\frac{C_{Np}^k\,C_{Nq}^{n-k}}{C_N^n},\ k=0,1,...,n,$$

where $0<p<1$, $q=1-p$, $0\le n\le N$, $N>0$.

If $n<<N$ (in practice, $n<0{,}1N$), then: $\dfrac{C_{Np}^k\,C_{Nq}^{n-k}}{C_N^n}\approx C_n^k p^k q^{n-k}$,

i.e., the hypergeometric distribution tends to the binomial distribution.

The numerical characteristics are given by the formulas:

$$M(X)=np,\qquad D(X)=\frac{N-n}{N-1}npq,\qquad \sigma(X)=\sqrt{\frac{N-n}{N-1}npq}.$$

A typical scheme in which the hypergeometric distribution arises is as follows: n elements are randomly drawn without replacement from a population of N elements containing exactly Np elements of type *I* and Nq elements of type *II*. The number of elements of type *I* in the sample is described by the hypergeometric distribution.

28.6.4. Poisson distribution law. A random variable X has the *Poisson distribution* with parameters λ $(\lambda>0)$ (fig. 28.8) if:

$$P_n(x=k)=\frac{\lambda^k e^{-\lambda}}{k!},\ k=0,1,2,..., \tag{28.10}$$

where $\lambda=np$, $k!=1\cdot 2\cdot ...\cdot k$.

Poisson distribution law with parameter λ of a discrete random variable is often written in the form of a table:

x_i	0	1	2	…	n
p_i	$e^{-\lambda}$	$\lambda e^{-\lambda}$	$\dfrac{\lambda^2 e^{-\lambda}}{2!}$	…	$\dfrac{\lambda^n e^{-\lambda}}{n!}$

Poisson distribution can be obtained as a limit of a binomial distribution if n goes to ∞ $(n \to \infty)$ and p goes to 0 $(p \to 0)$ (in this case $q = 1 - p \to 1$).

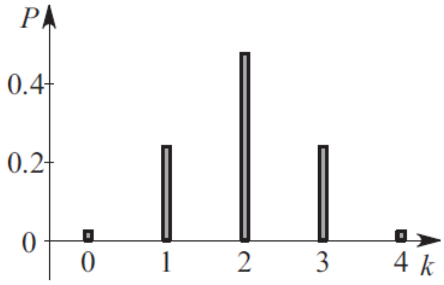


Fig. 28.7. **Hypergeometric distribution law for $\boldsymbol{p} = 0.5$, $\boldsymbol{N} = 10$, $\boldsymbol{n} = 4$**

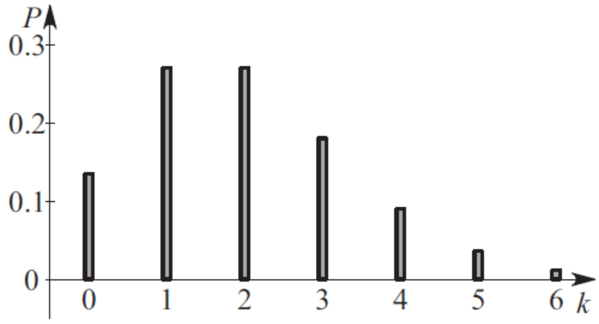


Fig. 28.8. **Poisson distribution law for $\lambda = 2$**

The numerical characteristics can be calculated by the formulas:

$$M(X) = np = \lambda, \tag{28.11}$$

$$D(X) = npq \approx \lambda, \tag{28.12}$$

$$\sigma(X) = \sqrt{npq} = \sqrt{\lambda}. \tag{28.13}$$

The Poisson distribution is the limit distribution for many discrete distributions such as the hypergeometric distribution, the binomial distribution, distributions arising in problems of arrangement of particles in cells, etc. The Poisson distribution is an acceptable model for describing the random number of occurrences of certain events on a given time interval in a given domain in space.

Example 28.6. The probability of finding a mistake on a book page is equal to 0.004. 500 pages are checked. Make up a distribution law of a number of finding a mistake on a book page. Find a mathematical expectation, a variance and a root-mean square deviation of a discrete random variable.

Solution. Let X be a number of finding a mistake on a book page, then the possible values of X are 0, 1, 2, 3, …, 500. Here $p = 0.004, n = 500$, then $\lambda = 500 \cdot 0.004 = 2$. Let's make up the distribution law of X according to Poisson formula (28.10):

$$P_{500}(x=0)=\frac{2^0\cdot e^{-2}}{0!}\approx 0.13534,\qquad P_{500}(x=1)=\frac{2^1\cdot e^{-2}}{1!}\approx 0.27067,$$

$$P_{500}(x=2)=\frac{2^2\cdot e^{-2}}{2!}\approx 0.27067,\qquad P_{500}(x=3)=\frac{2^3\cdot e^{-2}}{3!}\approx 0.18045,$$

$$P_{500}(x=4)=\frac{2^4\cdot e^{-2}}{4!}\approx 0.09022,\qquad P_{500}(x=5)=\frac{2^5\cdot e^{-2}}{5!}\approx 0.03609,$$

$$P_{500}(x=6)=\frac{2^6\cdot e^{-2}}{6!}\approx 0.01203,\qquad P_{500}(x=7)=\frac{2^7\cdot e^{-2}}{7!}\approx 0.00344,$$

$$P_{500}(x=8)=\frac{2^8\cdot e^{-2}}{8!}\approx 0.00086,\qquad P_{500}(x=9)=\frac{2^9\cdot e^{-2}}{9!}\approx 0.00019,$$

$P_{500}(x=10)=\dfrac{2^{10}\cdot e^{-2}}{10!}\approx 0.00004$ and so on.

At $k\geq 11$ we have that $P_{500}(x\geq 11)\approx 0$.

Let's calculate the total sum of probabilities: $\sum\limits_{i=1}^{500} p_i\approx 0.99999\approx 1$.

Let's find the numerical characteristics by the formulas (28.11) – (28.13):

$$M(X)=\lambda=2;\qquad D(X)\approx\lambda=2;\qquad \sigma(X)=\sqrt{\lambda}=\sqrt{2}\approx 0.41421.$$

So, the distribution law has the form:

x_i	0	1	2	3	4	5
p_i	0.13534	0.27067	0.27067	0.18045	0.09022	0.03609
x_i	6	7	8	9	10	…
p_i	0.01203	0.00344	0.00086	0.00019	0.00004	…

28.6.5. Negative binomial distribution law. A random variable X has the *negative binomial distribution* $(r,\ p)$ (fig. 28.9) if:

$$P_n(x=k)=C_{r+k-1}^{r-1}\ p^r(1-p)^k,\ k=0,\ 1,\ldots,\ r,$$

where $0<p<1,\ r>0$.

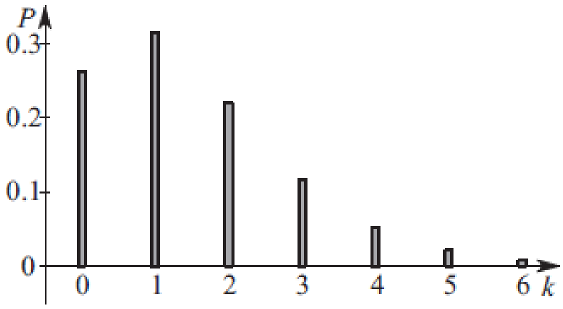


Fig. 28.9. **Negative binomial distribution law for $\boldsymbol{p}=0.8$, $\boldsymbol{n}=6$**

The numerical characteristics can be calculated by the formulas:

$$M(X)=\frac{r(1-p)}{p};\qquad D(X)=\frac{r(1-p)}{p^2};\qquad \sigma(X)=\frac{\sqrt{r(1-p)}}{p}.$$

The negative binomial distribution describes the number X of failures before the r-th success in a Bernoulli with probability p of success on each trail.

For $r=1$, the negative binomial distribution coincides with the geometric distribution.

Recommended bibliography: [2; 6; 8; 11].

Theme 29. Continuous and absolutely continuous random variables. Function and density of distribution of probabilities. Numerical characteristics

29.1. A definition of continuous random variables. A distribution function of probabilities of random variables and its properties

A ***continuous random variable*** is a random variable where the data can take infinitely many values on some numerical interval or a random variable which takes an infinite number of possible values. Continuous random variables are usually measurements.

Examples include height, weight, the amount of sugar in an orange, the time required to run a mile.

A continuous random variable is characterized by two functions:

1) a distribution function (the integral distribution function) $F(x)$;

2) a density function (the differential distribution function) $f(x)$.

The probability of the fact that a random variable X receives a value less than x, is called a ***cumulative distribution function*** of a random variable X and is marked as $F(x)$:

$$F(x) = P(X < x).$$

General properties of the integral distribution function:

1. $F(x)$ is a bounded function, i.e. $0 \le F(x) \le 1$.

2. $F(x)$ is a non-decreasing function for $x \in (-\infty, \infty)$, i.e. if $x_2 > x_1$, then $F(x_2) \ge F(x_1)$.

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

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

5. The probability that a random variable X lies in the interval (x_1, x_2) is equal to the increment of its cumulative distribution function on this interval; i.e. $P(x_1 < X < x_2) = F(x_2) - F(x_1)$.

6. For continuous random variables:

$$P(x_1 \le X < x_2) = P(x_1 < X < x_2) = P(x_1 < X \le x_2) = P(x_1 \le X \le x_2).$$

29.2. Absolutely continuous random variables. A distribution density of function of absolutely continuous random variables

The random variable X is called a ***continuous random variable***, if for any numbers $a < b$ such non-negative function $f(x)$ exists, that:

$$P(a < X < b) = \int_a^b f(x)dx.$$

The random variable X is called an ***absolutely continuous random variable*****,** if there is a non-negative function $f(x)$ on R that:

$$P(X \le x) = \int_{-\infty}^{x} f(t)dt, \text{ for every } x \in (-\infty, \infty).$$

The term a *continuous random variable* is a synonym of an *absolutely continuous random variable.*

An absolutely continuous random variable is a random variable whose cumulative distribution function is a continuous function.

The function $f(x)$ is called a ***density function*** of a continuous random variable.

General properties of the density function:

1. $f(x)$ is a non-negative function, i.e. $f(x) \ge 0$ for all *x*.
2. $f(x)$ is a non-decreasing function for $x \in (-\infty, \infty)$, then:

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

(the condition of normalization of the function $f(x)$).

3. The relationship between the functions $f(x)$ and $F(x)$:

$$F(x) = \int_{-\infty}^{x} f(x)dx \qquad \text{and} \qquad f(x) = F'(x).$$

4. The probability that a random variable X lies in the interval (x_1, x_2) is equal to the increment of its density distribution function on this interval; i.e.: $P(x_1 < X < x_2) = \int_{x_1}^{x_2} f(x)dx$.

Example 29.1. The density function of a continuous variable X is given by $f(x) = \begin{cases} 0, & x \le 0 \\ c\left(4x - 2x^2\right), & 0 < x \le 2 \\ 0, & x > 2 \end{cases}$. A. What is the value of c?

B. Find $P(X > 1)$.

Solution. Since $f(x)$ is a probability density function, we must have the condition of normalization of this function that $\int\limits_{-\infty}^{\infty} f(x)dx = 1$, implying that:

$$\int\limits_{-\infty}^{\infty} f(x)dx = \int\limits_{-\infty}^{0} 0dx + \int\limits_{0}^{2} c\left(4x - 2x^2\right)dx + \int\limits_{2}^{+\infty} 0dx = 1$$

or

$$\int\limits_{0}^{2} c\left(4x - 2x^2\right)dx = 1 \quad \text{or} \quad c\left(4\frac{x^2}{2} - 2\frac{x^3}{3}\right)\Bigg|_0^2 = c\left(2x^2 - \frac{2x^3}{3}\right)\Bigg|_0^2 = 1$$

or

$$c\left(8 - \frac{16}{3} - 0\right) = c \cdot \frac{8}{3} = 1 \quad \text{or} \quad c = \frac{3}{8}.$$

So, the probability density function is:

$$f(x) = \begin{cases} 0, & x \le 0 \\ \frac{3}{8}\left(4x - 2x^2\right), & 0 < x \le 2 \\ 0, & x > 2 \end{cases}.$$

Let's find $P(X > 1)$:

$$P(X > 1) = P(1 < X < +\infty) = \int\limits_{1}^{+\infty} f(x)dx = \int\limits_{1}^{2} \frac{3}{8}\left(4x - 2x^2\right)dx + \int\limits_{2}^{+\infty} 0dx =$$

$$= \frac{3}{8}\left(2x^2 - \frac{2x^3}{3}\right)\Bigg|_1^2 = \frac{3}{8}\left(\left(2 \cdot 2^2 - \frac{2 \cdot 2^3}{3}\right) - \left(2 \cdot 1^2 - \frac{2 \cdot 1^3}{3}\right)\right) = \frac{3}{8}\left(\frac{8}{3} - \frac{4}{3}\right) = \frac{1}{2}.$$

Example 29.2. The differential distribution function of a continuous variable X is given by $f(x) = \begin{cases} 0, & x \le 1 \\ x - \frac{1}{2}, & 1 < x \le 2 \\ 0, & x > 2 \end{cases}$. Find the integral distribution function $F(x)$. Plot the graphs of the functions $f(x)$ and $F(x)$.

Solution. The integral distribution function $F(x)$ according to the formula is $F(x)=\int\limits_{-\infty}^{x} f(x)dx$.

If $x \le 1$, then $f(x)=0$ and $F(x)=\int\limits_{-\infty}^{x} f(x)dx=\int\limits_{-\infty}^{x} 0dx=0$.

If $1<x\le 2$, then $F(x)=\int\limits_{-\infty}^{x} f(x)dx=\int\limits_{-\infty}^{1} f(x)dx+\int\limits_{1}^{x} f(x)dx=$

$$=\int\limits_{-\infty}^{1} 0dx+\int\limits_{1}^{x}\left(x-\frac{1}{2}\right)dx=\left(\frac{x^2}{2}-\frac{1}{2}x\right)\Bigg|_1^x=\frac{x^2}{2}-\frac{1}{2}x-\left(\frac{1}{2}-\frac{1}{2}\right)=\frac{x^2}{2}-\frac{x}{2}.$$

If $x>2$, then $F(x)=\int\limits_{-\infty}^{x} f(x)dx=\int\limits_{-\infty}^{1} f(x)dx+\int\limits_{1}^{2} f(x)dx+\int\limits_{2}^{x} f(x)dx=$

$$=\int\limits_{-\infty}^{1} 0dx+\int\limits_{1}^{2}\left(x-\frac{1}{2}\right)dx+\int\limits_{2}^{x} 0dx=\left(\frac{x^2}{2}-\frac{1}{2}x\right)\Bigg|_1^2=\frac{2^2}{2}-\frac{1}{2}\cdot 2-\left(\frac{1^2}{2}-\frac{1}{2}\cdot 1\right)=$$

$$=2-1-0=1.$$

Let's write the formula for the integral distribution function $F(x)$:

$$F(x)=\begin{cases} 0, & x\le 1 \\ \dfrac{x^2}{2}-\dfrac{x}{2}, & 1<x\le 2. \\ 1, & x>2 \end{cases}$$

Example 29.3. The integral distribution function of a continuous variable X is given by $F(x)=\begin{cases} 0, & x\le 0{,}5 \\ \dfrac{2x-1}{2}, & 0{,}5<x\le 1{,}5 \\ 1, & x>1{,}5 \end{cases}$. Find the density function $f(x)$.

Solution. It is known that $f(x) = F'(x)$. Thus:

$$f(x) = F'(X) = \begin{cases} (0)', & x \le 0{,}5 \\ \left(\dfrac{2x-1}{2}\right)', & 0{,}5 < x \le 1{,}5 \\ (1)', & x > 1{,}5 \end{cases} = \begin{cases} 0, & x \le 0{,}5 \\ \dfrac{2-0}{2}, & 0{,}5 < x \le 1{,}5 \\ 0, & x > 1{,}5 \end{cases} =$$

$$= \begin{cases} 0, & x \le 0{,}5 \\ 1, & 0{,}5 < x \le 1{,}5. \\ 0, & x > 1{,}5 \end{cases}$$

29.3. Numerical characteristics of absolutely continuous random variables and their properties

Let's consider basic numerical characteristics of an absolutely continuous random variable.

The ***mathematical expectation*** $M(X)$ of an absolutely continuous random variable is calculated by the formula:

$$M(X) = \int_{-\infty}^{\infty} x \cdot f(x)dx. \qquad (29.1)$$

For existence of the expectation (29.1), it is necessary that the corresponding integral converge absolutely.

General properties of a mathematical expectation:

1. $M(C) = C$ for any real C.
2. $M(\alpha X) = \alpha M(X)$ for any real α.
3. $M(X+Y) = M(X) + M(Y)$.
4. $M(\alpha X + \beta Y) = \alpha M(X) + \beta M(Y)$ for any real α and β.
5. $M(X-Y) = M(X) - M(Y)$.
6. $M(\alpha X - \beta Y) = \alpha M(X) - \beta M(Y)$ for any real α and β.
7. $M(X \cdot Y) = M(X) \cdot M(Y)$.

Here X, Y are mutually independent random variables.

The ***variance*** $D(X)$ of an absolutely continuous random variable is defined by the formula:

$$D(X)=\int_{-\infty}^{\infty} x^2\cdot f(x)dx-[M(X)]^2$$

or

$$D(X)=\int_{-\infty}^{\infty} (x-M(X))^2\cdot f(x)dx.$$

General properties of a variance:

1. $D(C)=0$ for any real C.
2. The variance is nonnegative: $D(X)\geq 0$.
3. $D(\alpha X+\beta)=\alpha^2\cdot D(X)$ for any real α and β.
4. $D(\alpha X)=\alpha^2\cdot D(X)$ for any real α.
5. $D(X+Y)=D(X)+D(Y)$ and $D(X-Y)=D(X)+D(Y)$.
6. $D(X\cdot Y)=D(X)\cdot D(Y)+D(X)\cdot M^2(Y)+D(Y)\cdot M^2(X)$.

The ***root-mean-square deviation*** *(or **standard deviation**)* $\sigma(X)$ of an absolutely continuous random variable is the square root of its variance:

$$\sigma(X)=\sqrt{D(X)}.$$

A root-mean-square deviation has the same dimension as the random variable itself.

*A **mode*** of an absolutely continuous random variable M_o is a point of maximum of the probability density function $f(x)$.

The expectation $M\left((X-a)^k\right)$ is called the k***-th moment*** of an absolutely continuous random variable X about a. The moments about zero are usually referred to simply as the moments of a random variable and sometimes they are called *initial moments*. The k*-th moment* satisfies the relation:

$$\nu_k=\int_{-\infty}^{+\infty} x^k\cdot f(x)dx.$$

If $a = M(X)$ then k-*th* moment of the random variable X about a is called the k***-th central moment.*** The k-th central moment satisfies the relation:

$$\mu_k = \int_{-\infty}^{+\infty} (x - M(X))^k \cdot f(x) dx.$$

Example 29.4. The density function of a continuous variable X is given by $f(x) = \begin{cases} 0, & x \le 0{,}5 \\ 1, & 0{,}5 < x \le 1{,}5 \\ 0, & x > 1{,}5 \end{cases}$. Calculate: a) the mathematical expec-tation $M(X)$ and the variance $D(X)$; b) the probability that a random variable X lies in the interval $(1.0, 2.3)$; c) plot graphs of the functions $F(X)$ and $f(x)$.

Solution. A. Let's find the mathematical expectation:

$$M(X) = \int_{-\infty}^{\infty} x \cdot f(x) dx = \int_{-\infty}^{0.5} x \cdot 0 dx + \int_{0.5}^{1.5} x \cdot 1 dx + \int_{1.5}^{\infty} x \cdot 0 dx =$$

$$= 0 + \int_{0.5}^{1.5} x dx + 0 = \left. \frac{x^2}{2} \right|_{0.5}^{1.5} = \frac{1}{2} \cdot \left[1.5^2 - 0.5^2\right] = \frac{1}{2} \cdot [2.25 - 0.25] = \frac{1}{2} \cdot 2 = 1.$$

Let's calculate the variance: $D(X) = \int_{-\infty}^{\infty} x^2 \cdot f(x) dx - [M(X)]^2 =$

$$= \int_{-\infty}^{0.5} x^2 \cdot 0 dx + \int_{0.5}^{1.5} x^2 \cdot 1 dx + \int_{1.5}^{\infty} x^2 \cdot 0 dx - 1^2 = 0 + \int_{0.5}^{1.5} x^2 dx + 0 - 1 =$$

$$= \left. \frac{x^3}{3} \right|_{0.5}^{1.5} - 1 = \frac{1}{3} \cdot \left[1.5^3 - 0.5^3\right] - 1 = \frac{1}{3} \cdot [3.375 - 0.125] - 1 =$$

$$= \frac{1}{3} \cdot 3.25 - 1 \approx 0.0833.$$

B. Let's find the probability that a random variable X lies in the interval $(1.0, 2.3)$ by the formula $P(x_1 < X < x_2) = F(x_2) - F(x_1)$.

Thus, $P(1 < X < 2.3) = F(2.3) - F(1) = 1 - \frac{2 \cdot 1 - 1}{2} = 1 - \frac{1}{2} = 0.5$.

C. Let's plot graphs of the functions $F(X)$ and $f(x)$ (fig. 29.1, 29.2).

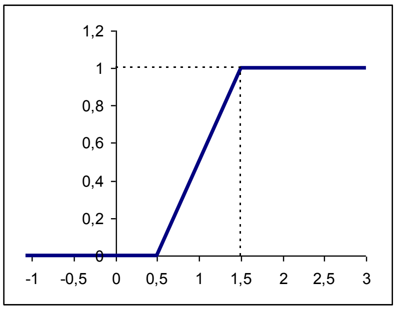


Fig. 29.1. **The graph of $F(x)$**

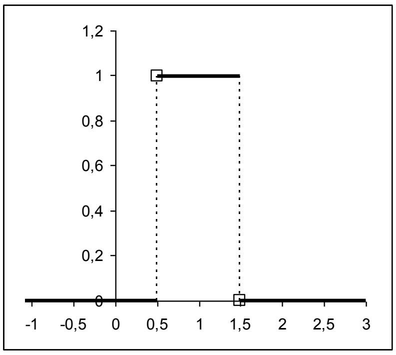


Fig. 29.2. **The graph of $f(x)$**

Recommended bibliography: [5; 6; 8; 10; 11].

Theme 30. Uniform, exponential and normal laws of probabilities distribution. Transformation of sequences of normal distributed random variable

30.1. A uniform law of probabilities distribution and its numerical characteristics

The uniform law of distribution is characterized by a probability density function $f(x)$ (the differential distribution function) (fig. 30.1) and a cumulative distribution function $F(x)$ (the integral distribution function) (fig. 30.2).

$$f(x) = \begin{cases} 0, & x \le a \\ \frac{1}{b-a}, & a < x \le b, \\ 0, & x > b \end{cases} \qquad F(x) = \begin{cases} 0, & x \le a \\ \frac{x-a}{b-a}, & a < x \le b. \\ 1, & x > b \end{cases}$$

The probability that a random variable X lies in the interval (α, β) is equal to the increment of its integral distribution function on this interval; i.e.:

$$P(\alpha < X < \beta) = F(\beta) - F(\alpha) = \frac{\beta - \alpha}{b - a}.$$

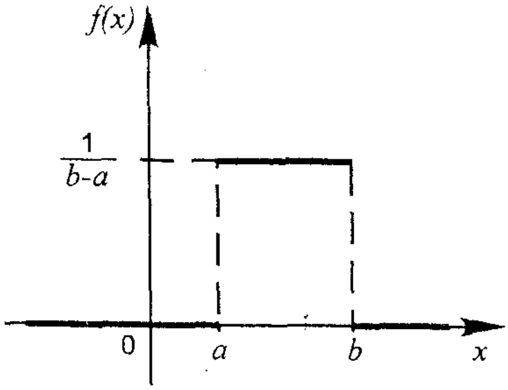


Fig. 30.1. **The graph of $f(x)$**

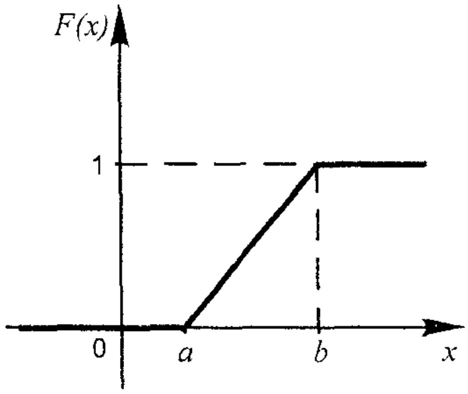


Fig. 30.2. **The graph of $F(x)$**

Numerical characteristics of uniform random variables are

1) *mathematical expectation*: $M(X) = \dfrac{a+b}{2}$;

2) *variance*: $D(X) = \dfrac{(b-a)^2}{12}$;

3) *root-mean-square deviation (or standard deviation)*: $\sigma(X) = \dfrac{b-a}{2\sqrt{3}}$.

Example 30.1. The parameters a, b of the uniform law of distribution are given: $a = 2$ and $b = 6$. Find: a) functions $f(x)$ and $F(x)$; b) the mathematical expectation $M(X)$, the variance $D(X)$ and the root-mean-square deviation $\sigma(X)$; c) $P(0 < X < 3)$.

Solution. Let's find functions $f(x)$ and $F(x)$ substituting $a = 2$ and $b = 6$ into formulas for functions:

$$f(x) = \begin{cases} 0, & x \le 2 \\ \dfrac{1}{6-2}, & 2 < x \le 6 \\ 0, & x > 6 \end{cases} = \begin{cases} 0, & x \le 2 \\ \dfrac{1}{4}, & 2 < x \le 6, \\ 0, & x > 6 \end{cases}$$

$$F(x)=\begin{cases}0, & x\le 2\\ \dfrac{x-2}{6-2}, & 2<x\le 6\\ 1, & x>6\end{cases}=\begin{cases}0, & x\le 2\\ \dfrac{x-2}{4}, & 2<x\le 6.\\ 1, & x>6\end{cases}$$

Let's calculate the numerical characteristics:

$$M(X)=\frac{a+b}{2}=\frac{2+6}{2}=4, \qquad D(X)=\frac{(6-2)^2}{12}=\frac{4^2}{12}=\frac{4}{3},$$

$$\sigma(X)=\frac{b-a}{2\sqrt{3}}=\frac{6-2}{2\sqrt{3}}=\frac{4}{2\sqrt{3}}=\frac{2}{\sqrt{3}}.$$

Let's calculate the probability that the uniform random variable X lies in the interval $(0,3)$:

$$P(\alpha<X<\beta)=P(0<X<3)=F(3)-F(0)=\frac{3-0}{6-2}=\frac{3}{4}=0.75.$$

30.2. An exponential law of distribution

The exponential distribution law is characterized by a probability density function $f(x)$ (the differential distribution function) (fig. 30.3) and a cumulative distribution function $F(x)$ (the integral distribution function) (fig. 30.4).

$$f(x)=\begin{cases}0, & x<0\\ \lambda\cdot e^{-\lambda x}, & x\ge 0\end{cases}, \qquad F(x)=\begin{cases}0, & x<0\\ 1-e^{-\lambda x}, & x\ge 0\end{cases}.$$

The probability that a random variable X lies in the interval (α,β) is equal to the increment of its integral distribution function on this interval; i.e.

$$P(\alpha<X<\beta)=F(\beta)-F(\alpha)=e^{-\lambda\alpha}-e^{-\lambda\beta}.$$

Numerical characteristics of exponential random variables are

1) *mathematical expectation*: $M(X)=\dfrac{1}{\lambda}$;

2) *variance*: $D(X)=\dfrac{1}{\lambda^2}$;

3) *root-mean-square deviation (or standard deviation)*: $\sigma(X)=\dfrac{1}{\lambda}$.

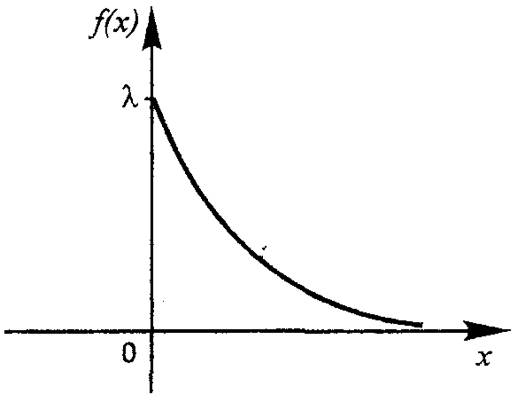


Fig. 30.3. **The graph of $f(x)$**

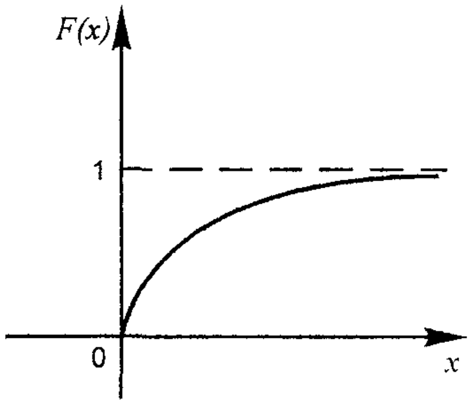


Fig. 30.4. **The graph of $F(x)$**

Example 30.2. The probability density function of the exponential law of distribution $f(x)=\begin{cases}0, & x<0\\ 0.05\cdot e^{-0.05x}, & x\ge 0\end{cases}$ is given. Find: a) the function $F(x)$; b) the mathematical expectation $M(X)$, the variance $D(X)$ and the root-mean-square deviation $\sigma(X)$; c) $P(2<X<10)$.

Solution. We have that $\lambda=0.05$ from the formula of $f(x)$. Let's substitute this parameter into the formula for $F(x)$ and find the numerical characteristics:

$$F(x)=\begin{cases}0, & x<0\\ 1-e^{-\lambda x}, & x\ge 0\end{cases}=\begin{cases}0, & x<0\\ 1-e^{-0.05x}, & x\ge 0\end{cases},$$

$$M(X)=\frac{1}{\lambda}=\frac{1}{0.05}=20,\ D(X)=\frac{1}{\lambda^2}=\frac{1}{0.05^2}=400,\ \sigma(X)=\frac{1}{\lambda}=\frac{1}{0.05}=20.$$

Let's calculate the probability $P(2<X<10)$ that a random variable X lies in the interval $(2,10)$:

$$P(\alpha < X < \beta) = P(2 < X < 10) = e^{-\lambda\alpha} - e^{-\lambda\beta} = e^{-0.05\cdot 2} - e^{-0.05\cdot 10} =$$

$$= 0.90484 - 0.60653 = 0.29831.$$

30.3. A normal law of probabilities distribution and its standard representation

A random variable X has the ***normal distribution*** with parameters (a, σ^2) if its probability density function $f(x)$ and the cumulative distribution function $F(x)$ have the forms:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x-a)^2}{2\sigma^2}}, \qquad F(x) = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x-a)^2}{2\sigma^2}} dx.$$

The normal distribution law is characterized by two functions: the probability density function $f(x)$ (the differential distribution function) and the cumulative distribution function $F(x)$ (the integral distribution function).

Let's check whether that satisfies the normalization property of the differential distribution function. Indeed, $f(x) > 0$.

Let's calculate:

$$\int_{-\infty}^{+\infty} f(x)dx = \int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x-a)^2}{2\sigma^2}} dx = \left| \begin{array}{l} t = \dfrac{x-a}{\sigma},\ x = \sigma t + a, \\ dx = \sigma dt,\ t_1 = -\infty,\ t_2 = +\infty \end{array} \right| =$$

$$= \int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{t^2}{2}} \cdot \sigma dt = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{t^2}{2}} dt = \frac{1}{\sqrt{2\pi}} \cdot \sqrt{2\pi} = 1,$$

where

$$\int_{-\infty}^{+\infty} e^{-\frac{t^2}{2}} dt = \sqrt{2\pi} \qquad (30.1)$$

is called Poisson's integral.

Let's define the integral distribution function $F(x)$ for the normal distribution law:

$$F(x) = \int_{-\infty}^{x} f(x)dx = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x-a)^2}{2\sigma^2}} dx .$$

Let's calculate the obtained integral:

$$F(x) = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x-a)^2}{2\sigma^2}} dx = \left| \begin{array}{l} t = \dfrac{x-a}{\sigma},\ x = \sigma t + a \\ dx = \sigma dt,\ t_1 = -\infty,\ t_2 = +\infty \end{array} \right| =$$

$$= \int_{-\infty}^{\frac{x-a}{\sigma}} \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{t^2}{2}} \cdot \sigma dt = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{x-a}{\sigma}} e^{-\frac{t^2}{2}} dt .$$

Let's use the property of additivity of an integral, i.e.

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0} e^{-\frac{t^2}{2}} dt + \frac{1}{\sqrt{2\pi}} \int_{0}^{\frac{x-a}{\sigma}} e^{-\frac{t^2}{2}} dt .$$

Let's apply Poisson's integral $\int_{-\infty}^{0} e^{-\frac{t^2}{2}} dt = \frac{1}{2}\sqrt{2\pi}$ and transform this integral $\frac{1}{\sqrt{2\pi}} \int_{0}^{\frac{x-a}{\sigma}} e^{-\frac{t^2}{2}} dt$ using Laplace integral function $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{0}^{x} e^{-\frac{t^2}{2}} dt$, i.e. $\Phi\left(\frac{x-a}{\sigma}\right) = \frac{1}{\sqrt{2\pi}} \int_{0}^{\frac{x-a}{\sigma}} e^{-\frac{t^2}{2}} dt$, and obtain:

$$F(x) = \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{2}\sqrt{2\pi} + \Phi\left(\frac{x-a}{\sigma}\right) = \frac{1}{2} + \Phi\left(\frac{x-a}{\sigma}\right).$$

Thus, the linear transformation $t = \frac{x-a}{\sigma}$ reduces the normal distribution with parameters $\left(a,\sigma^2\right)$ to the standard normal distribution with parameters $(0,1)$ and the cumulative distribution function

$$f(x)=\frac{\varphi(t)}{\sigma},\qquad F(x)=\frac{1}{2}+\Phi(t),$$

where $\varphi(t)=\frac{1}{\sqrt{2\pi}}\cdot e^{-\frac{t^2}{2}}$ and $\Phi(t)=\frac{1}{\sqrt{2\pi}}\int\limits_0^t \cdot e^{-\frac{t^2}{2}}dt$ are Laplace differential and integral functions.

The values of the probability density function $f(x)$ and the cumulative distribution function $F(x)$ are computed by (see appendix A and appendix B).

Graphs of the differential function $f(x)$ and the integral function $F(x)$ are shown in fig. 30.5 and fig. 30.6. For the normal distribution the curve of $f(x)$ reaches the maximum at $x=a$ and it is symmetric relative to the line $x=a$.

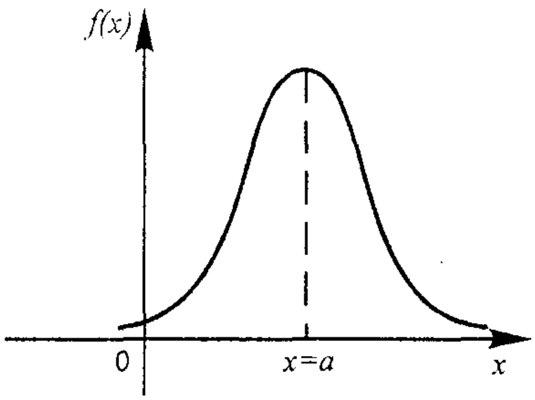


Fig. 30.5. **A density curve $f(x)$ graph**

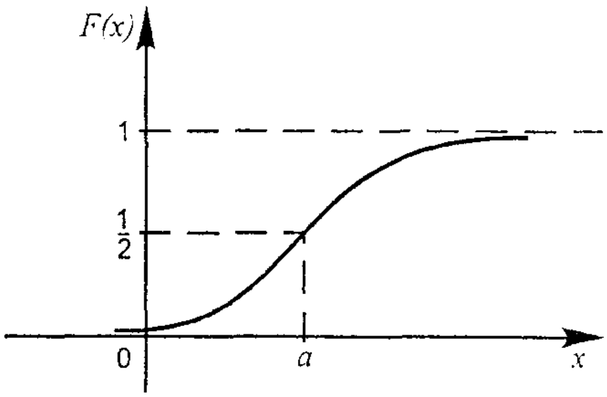


Fig. 30.6. **The graph of $F(x)$**

Numerical characteristics of normal distribution law are:

1) *mathematical expectation*: $M(X)=a \quad (M(t)=0)$,

2) *variance*: $D(X)=\sigma^2 \quad (D(t)=1)$,

3) *root-mean-square deviation*: $\sigma(X)=\sigma \quad (\sigma(t)=1)$.

Let's check that $M(X)=a$ using the definition of the mathematical expectation of the absolutely continuous variable:

$$M(X)=\int_{-\infty}^{+\infty} x\cdot f(x)dx=\int_{-\infty}^{+\infty} x\cdot\frac{1}{\sigma\sqrt{2\pi}}\cdot e^{-\frac{(x-a)^2}{2\sigma^2}}dx=\left|\begin{array}{l} t=\dfrac{x-a}{\sigma},\ x=\sigma t+a \\ dx=\sigma dt, \\ t_1=-\infty,\ t_2=+\infty \end{array}\right|=$$

$$=\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{+\infty}(\sigma t+a)e^{-\frac{t^2}{2}}\sigma dt=\frac{\sigma}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} te^{-\frac{t^2}{2}}dt+\frac{a}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} e^{-\frac{t^2}{2}}dt.$$

The first integral equals zero, because it is the integral of the odd function on the symmetric interval relative to the origin. The second one is the Poisson's integral (30.1).

Thus, $M(X)=\dfrac{a}{\sqrt{2\pi}}\cdot\sqrt{2\pi}=a$.

Let's check that $D(X)=\sigma^2$ using the definition of the variance of the absolutely continuous variable:

$$D(X)=\int_{-\infty}^{+\infty}(x-M(X))^2\cdot f(x)dx=\int_{-\infty}^{+\infty}(x-a)^2\cdot\frac{1}{\sigma\sqrt{2\pi}}\cdot e^{-\frac{(x-a)^2}{2\sigma^2}}dx=$$

$$=\left|\begin{array}{l} t=\dfrac{x-a}{\sigma},\ x=\sigma t+a \\ dx=\sigma dt,\ t_1=-\infty,\ t_2=+\infty \end{array}\right|=\frac{\sigma^2}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} t^2e^{-\frac{t^2}{2}}dt=$$

$$=\left|\begin{array}{l} u=t,\quad du=dt \\ dv=te^{-\frac{t^2}{2}}dt,\quad v=\int e^{-\frac{t^2}{2}}d\left(\dfrac{t^2}{2}\right)=-e^{-\frac{t^2}{2}} \end{array}\right|=-\frac{\sigma^2}{\sqrt{2\pi}}\underbrace{te^{-\frac{t^2}{2}}\Big|_{-\infty}^{+\infty}}_{=0}+$$

$$+\frac{\sigma^2}{\sqrt{2\pi}}\underbrace{\int_{-\infty}^{+\infty} e^{-\frac{t^2}{2}}dt}_{=\sqrt{2\pi}}=0+\frac{\sigma^2}{\sqrt{2\pi}}\cdot\sqrt{2\pi}=\sigma^2.$$

Thus, $D(X) = \sigma^2$.

Let's find probability that a random variable X, distributed by the normal law with parameters (a, σ^2), lies in the interval (α, β):

$$P(\alpha < X < \beta) = F(\beta) - F(\alpha) = \frac{1}{2} + \Phi\left(\frac{\beta - a}{\sigma}\right) - \left(\frac{1}{2} + \Phi\left(\frac{\alpha - a}{\sigma}\right)\right) =$$

$$= \Phi\left(\frac{\beta - a}{\sigma}\right) - \Phi\left(\frac{\alpha - a}{\sigma}\right).$$

Thus, the probability that a random variable X lies in the interval (α, β) is equal to the increment of its integral distribution function on this interval; i.e.

$$P(\alpha < X < \beta) = F(\beta) - F(\alpha) = \Phi\left(\frac{\beta - a}{\sigma}\right) - \Phi\left(\frac{\alpha - a}{\sigma}\right).$$

Example 30.7. The probability density function of the normal law of distribution $f(x) = \frac{1}{2\sqrt{2\pi}} \cdot e^{-\frac{(x-3)^2}{8}}$ is given. Find the integral function, calculate the numerical characteristics and $P(1 < X < 7)$.

Solution. We have that $a = 3$ and $\sigma = 2$ from the formula of $f(x)$. Let's substitute these parameters into the formula for $F(x)$ and find the numerical characteristics:

$$F(x) = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x-a)^2}{2\sigma^2}} dx = \int_{-\infty}^{x} \frac{1}{2\sqrt{2\pi}} \cdot e^{-\frac{(x-3)^2}{2 \cdot 2^2}} dx =$$

$$= \int_{-\infty}^{x} \frac{1}{2\sqrt{2\pi}} \cdot e^{-\frac{(x-3)^2}{8}} dx,$$

$$M(X) = a = 3, \qquad D(X) = \sigma^2 = 2^2 = 4, \qquad \sigma(X) = \sigma = 2.$$

Let's calculate the probability $P(1 < X < 7)$ that a random variable X

lies in the interval $(1,7)$:

$$P(\alpha < X < \beta) = P(1 < X < 7) = \Phi\left(\frac{\beta - a}{\sigma}\right) - \Phi\left(\frac{\alpha - a}{\sigma}\right) =$$

$$= \Phi\left(\frac{7-3}{2}\right) - \Phi\left(\frac{1-3}{2}\right) = \Phi(2) - \Phi(-1) = \left\| \begin{array}{l} \text{use the property} \\ \Phi(-x) = \Phi(x) \end{array} \right\| = \Phi(2) + \Phi(1) =$$

$$= \left\| \begin{array}{l} \text{use appendix B} \\ \Phi(2) = 0.4772 \\ \Phi(1) = 0.3413 \end{array} \right\| = 0.4772 - 0.3413 = 0.1359.$$

Let's find probability that a module of the deviation of the normal distributed random variable from its mathematical expectation is less than any nonnegative ε, i.e. $P(|X - a| < \varepsilon)$:

$$P(|X - a| < \varepsilon) = P(-\varepsilon < X - a < \varepsilon) = P(a - \varepsilon < X < \varepsilon + a) =$$

$$= \Phi\left(\frac{a + \varepsilon - a}{\sigma}\right) - \Phi\left(\frac{a - \varepsilon - a}{\sigma}\right) = \Phi\left(\frac{\varepsilon}{\sigma}\right) - \Phi\left(-\frac{\varepsilon}{\sigma}\right) = 2\Phi\left(\frac{\varepsilon}{\sigma}\right).$$

Thus,

$$P(|X - a| < \varepsilon) = 2\Phi\left(\frac{\varepsilon}{\sigma}\right). \qquad (30.2)$$

Three sigma rule. Let's transform the formula (30.2). Let $\varepsilon = \sigma \cdot t$, then $P(|X - a| < \sigma t) = 2\Phi(t)$.

If $t = 1$, i.e. $\varepsilon = \sigma$, then $P(|X - a| < \sigma) = 2\Phi(1) = 0.6826$. It means that 68 % of values of a random variable X is located on the interval $(a \pm \sigma)$.

If $t = 2$, i.e. $\varepsilon = 2\sigma$, then $P(|X - a| < 2\sigma) = 2\Phi(2) = 0.9544$. It means that 95 % of values of a random variable X is located on the interval $(a \pm 2\sigma)$.

If $t = 3$, i.e. $\varepsilon = 3\sigma$, then $P(|X - a| < 3\sigma) = 2\Phi(3) = 0.9973$. Hence *three sigma rule* means the normal distributed random variable X possesses all its values on the interval $(a \pm 3\sigma)$ with the probability 100 %.

30.4. Gamma-distribution

A ***random variable*** X that is ***gamma-distributed*** with shape k and scale θ is denoted by $\Gamma(k, \theta)$.

The probability density function and the cumulative distribution function of the gamma distribution can be expressed in terms of the gamma function parameterized in terms of a shape parameter k and scale parameter θ and the lower incomplete gamma function, i.e.

$$f(x) = \begin{cases} \dfrac{1}{\theta^k \cdot \Gamma(k)} x^{k-1} e^{-\frac{x}{\theta}}, \text{ if } x \geq 0 \\ 0, \qquad\qquad\qquad\quad , \text{if } x < 0 \end{cases}, \qquad F(x) = \int_0^x f(x)dx = \frac{\gamma\left(k, \frac{x}{\theta}\right)}{\Gamma(k)},$$

where $\Gamma(k) = \int_0^{+\infty} t^{k-1} e^{-t} dt$ is the gamma function, both k and θ are positive values.

Numerical characteristics of gamma distribution law are

1) *mathematical expectation:* $M(X) = k\theta$;

2) *variance:* $D(X) = k\theta^2$;

3) *root-mean-square deviation (or standard deviation)*: $\sigma(X) = \theta\sqrt{k}$.

30.5. χ^2-distributions (chi-square) of Student and Fisher, their relationship with a standard normal law

A random variable $X = \chi^2(n)$ has the ***chi-square distribution*** with n degrees of freedom if its probability density function and the cumulative distribution function have the forms:

$$f(x) = \begin{cases} \dfrac{1}{2^{\frac{n}{2}} \cdot \Gamma\left(\frac{\alpha}{2}\right)} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}, \text{if } x > 0 \\ 0, \qquad\qquad\qquad\qquad , \text{if } x \leq 0 \end{cases}, \tag{30.3}$$

$$F(x)=\frac{1}{2^{\frac{n}{2}}\cdot\Gamma\left(\frac{\alpha}{2}\right)}\int_0^x t^{\frac{n}{2}-1}e^{-\frac{t}{2}}dt,$$

where $\Gamma\left(\frac{\alpha}{2}\right)$ is the gamma function.

Numerical characteristics of gamma distribution law are:

1) *mathematical expectation:* $M(X)=M\left(\chi^2(n)\right)=n$;

2) *variance:* $D(X)=D\left(\chi^2(n)\right)=2n$;

3) *root-mean-square deviation*: $\sigma(X)=\sigma\left(\chi^2(n)\right)=\sqrt{2n}$.

Main property of chi-square distribution. For an arbitrary n the sum:

$$X=\sum_{k=1}^{n}X_k^2$$

of squares of independent random variables obeying the standard normal distribution has the chi-square distribution with n degrees of freedom.

The values $\chi^2(n)$ are tabulated.

Relationship with other distributions:

1. For $n=1$, the formula (30.3) gives the probability density function of the square X^2 of a random variable with the standard normal distribution.

2. For $n=2$, the formula (30.3) gives the exponential distribution with parameter $\lambda=\frac{1}{2}$.

3. As $n\to\infty$ the random variable $X=\chi^2(n)$ has an asymptotically normal distribution with parameters $(n, 2n)$.

A random variable $X=t(n)$ has ***Student's distribution** (t**-distribution**) **with** n **degrees of freedom*** ($n>0$) if its probability density function and the cumulative distribution function have the forms (fig. 30.7, 30.8):

$$f(x)=\frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\cdot\Gamma\left(\frac{n}{2}\right)}\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}}, \tag{30.4}$$

$$F(x)=\frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\cdot\Gamma\left(\frac{n}{2}\right)}\int_{-\infty}^{x}\left(1+\frac{t^2}{n}\right)^{-\frac{n+1}{2}}dt,$$

where $\Gamma\left(\frac{n}{2}\right)$ is the gamma function and $x\in(-\infty,+\infty)$.

Numerical characteristics of gamma distribution law are

1) *mathematical expectation:* $M(X)=M(t(n))=0$ if $n>1$;

2) *variance:* $D(X)=D(t(n))=\begin{cases}\frac{n}{n-2}, \text{ for } n>2\\ 0, \quad \text{for } n\le 2\end{cases}$;

3) *root-mean-square deviation*: $\sigma(X)=\sqrt{D(t(n))}$.

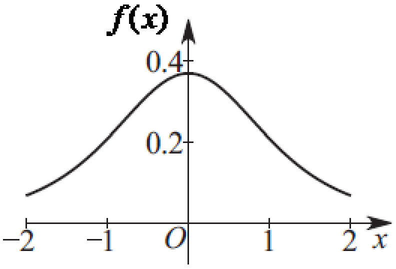


Fig. 30.7. **Probability density function of Student's t-distribution for $n=3$**

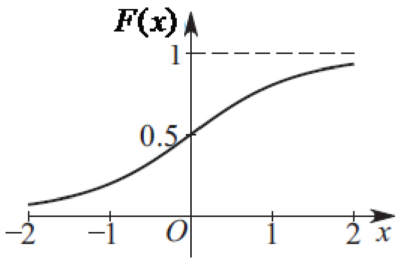


Fig. 30.8. **Cumulative distribution function of Student's t-distribution for $n=3$**

Main property of Student's distribution. If η and $\chi^2(n)$ are independent random variables and η has the standard normal distribution, then the random variable

$$t(n)=\eta\sqrt{\frac{n}{\chi^2(n)}}$$

has Student's distribution with n degrees of freedom.

The values $t(n)$ are tabulated.

Relationship with other distributions: as $n \to \infty$ Student's distribution is an asymptotically normal distribution with parameters $(0, 1)$.

Student's distribution is used when testing the hypothesis about the mean of a normally distributed population with an unknown variance.

A random variable X has F*-distribution* (or *Fisher–Snedecor distribution*) *with parameters* k_1 *and* k_2 if its probability density function has the form:

$$f(x) = \frac{\Gamma\left(\frac{k_1+k_2}{2}\right) k_1^{\frac{k_1}{2}} k_2^{\frac{k_2}{2}}}{\Gamma\left(\frac{k_1}{2}\right)\Gamma\left(\frac{k_2}{2}\right)} x^{\frac{k_1}{2}-1} (k_1 x + k_2)^{-\frac{k_1+k_2}{2}},$$

where $\Gamma(n)$ is the gamma function and $x \geq 0$.

F***-distribution*** (or ***Fisher–Snedecor distribution***) is called the distribution of a random variable $X = F = \dfrac{\frac{1}{k_1}\chi^2(k_1)}{\frac{1}{k_2}\chi^2(k_2)}$, where $\chi^2(k_1)$ and $\chi^2(k_2)$ are random variables which have χ^2-distribution with k_1 and k_2 degrees of freedom, respectively.

Main property of F*-distribution*. F-distribution (or Fisher–Snedecor distribution) is the distribution of a random variable $X = F = \dfrac{\frac{1}{k_1}\chi^2(k_1)}{\frac{1}{k_2}\chi^2(k_2)}$, where $\chi^2(k_1)$ and $\chi^2(k_2)$ are random variables which have χ^2-distribution with k_1 and k_2 degrees of freedom, respectively.

The values F are tabulated.

Relationship with other distributions: as $n \to \infty$ F-distribution is an asymptotically normal distribution.

Student's distribution is used when testing the hypothesis about the mean of a normally distributed population with an unknown variance.

Recommended bibliography: [2; 5; 7; 9; 11].

Theme 31. Random vectors and laws of their distributions: joint (consistent, marginal and conditional. Systems of independent random variables. Conditional and marginal numerical characteristics

31.1. Random vectors and joint law of probabilities distribution, its components

The concept of a random vector is a multidimensional generalization of a random variable.

Let's suppose that random variables $X_1, X_2, \dots, X_n$ are defined on a sample space Ω or, in other words, each outcome of a random experiment on a sample space Ω may need to be described by a set of $n > 1$ random variables $X_1, X_2, \dots, X_n$. Then one says that an ***n-dimensional random vector*** $\overline{X} = (X_1, X_2, \dots, X_n)$ or a ***system of random variables*** is given. The random variables $X_1, X_2, \dots, X_n$ can be viewed as the coordinates of points in an n-dimensional space.

For multidimensional random variables we can use basic concepts of one-dimensional random variables.

The ***distribution function*** $F_{\overline{X}}(x_1, x_2, \dots, x_n)$ of a random vector $\overline{X} = (X_1, X_2, \dots, X_n)$ is defined by the formula:

$$F_{\overline{X}}(x_1, x_2, \dots, x_n) = P\{\omega : X_1(\omega) < x_1, X_2(\omega) < x_2, \dots, X_n(\omega) < x_n\}.$$

This distribution function $F_{\overline{X}}(x_1, x_2, \dots, x_n)$ is nondecreasing one of its arguments $x_1, x_2, \dots, x_n$ and it defines a law of probabilities distribution of a random vector $\overline{X} = (X_1, X_2, \dots, X_n)$.

A ***random vector*** $\overline{X} = (X_1, X_2, \dots, X_n)$ is called ***discrete*** if there exists a finite set or a infinite set of n-dimensional random vectors $\overline{x}_1, \overline{x}_2, \dots,$

$\bar{x}_n$ such that:

$$\sum_{i=1}^{\infty} P\left(\overline{X} = \bar{x}_i\right) = 1.$$

A ***distribution law*** of a discrete random vector is defined completely by definition of vectors $\bar{x}_1, \bar{x}_2, \ldots$ and their probabilities $p_1 = P\left(\overline{X} = \bar{x}_1\right)$, $p_2 = P\left(\overline{X} = \bar{x}_2\right)$, … such that $p_1 + p_2 + ... = 1$.

A ***random vector*** $\overline{X}$ is called ***absolutely continuous*** (or, simply, ***continuous***) if there exists nonnegative function $f_{\overline{X}}\left(x_1, x_2, \ldots, x_n\right)$ such that for any $\bar{x} = \left(x_1, x_2, \ldots, x_n\right)$ a distribution function $F_{\overline{X}}\left(x_1, x_2, \ldots, x_n\right) = F_{\overline{X}}\left(\bar{x}\right)$ can be presented in a form of n-dimensional integral, i.e.:

$$F_{\overline{X}}\left(\bar{x}\right) = \int_{-\infty}^{x_1} dt_1 \int_{-\infty}^{x_2} dt_2 \ldots \int_{-\infty}^{x_n} f_{\overline{X}}\left(t_1, t_2, \ldots, t_n\right) dt_n .$$

This function $f_{\overline{X}}\left(x_1, x_2, \ldots, x_n\right)$ is called a ***density of probabilities distribution*** of a random vector $\overline{X}$.

A density of distribution also determines distribution law *of* a random vector as:

$$f_{\overline{X}}\left(x_1, x_2, \ldots, x_n\right) = \frac{\partial^n F_{\overline{X}}\left(x_1, x_2, \ldots, x_n\right)}{\partial x_1 \, \partial x_2 \ldots \, \partial x_n} .$$

For ***independent random variables*** $X_1, X_2, \ldots, X_n$ a distribution function $F_{\overline{X}}\left(x_1, x_2, \ldots, x_n\right)$ of *distribution* n-dimensional random vector $\overline{X} = \left(X_1, X_2, \ldots, X_n\right)$ is equal to a product of distribution functions of random variables $X_1, X_2, \ldots, X_n$, i.e.

$$F_{\overline{X}}\left(x_1, x_2, \ldots, x_n\right) = F_{X_1}\left(x_1\right) \cdot F_{X_2}\left(x_2\right) \cdot \ldots \cdot F_{X_n}\left(x_n\right).$$

This condition presents a base of a definition of random variables' independence.

If $\overline{X} = \left(X_1, X_2, \ldots, X_n\right)$ is an absolutely continuous random vector, then for independent random variables we have

$$f_{\overline{X}}(x_1, x_2, \ldots, x_n) = f_{X_1}(x_1) \cdot f_{X_2}(x_2) \cdot \ldots \cdot f_{X_n}(x_n).$$

If $\overline{X} = (X_1, X_2, \ldots, X_n)$ is a discrete random vector, then for independent random variables we have:

$$P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) = P(X_1 = x_1) \cdot P(X_2 = x_2) \cdot \ldots \cdot P(X_n = x_n).$$

Let's consider the following properties of random vectors by the example of systems of two-dimensional random variables.

31.2. A system of two-dimensional random variables. Probabilities of a function of joint distribution, a component of two-dimensional vector. Marginal functions of distribution of a component of a random vector

Let's denote $Z(X, Y)$ as a two-dimensional random vector and call each random variable X and Y as a component.

The ***distribution function*** $F(x, y) = F_{X,Y}(x, y)$ of a two-dimensional discrete or absolutely continuous random vector $Z(X, Y)$ or the ***joint distribution function of the random variables*** X and Y is defined as the probability of the simultaneous occurrence (intersection) of the events $(X < x)$ and $(Y < y)$, i.e. the probability that X possesses the value less than x at Y less than y:

$$F(x, y) = F_{X,Y}(x, y) = P(X < x, Y < y). \tag{31.1}$$

Geometrically, $F(x, y)$ can be interpreted as the probability that the random point (X, Y) lies in the lower left infinite quadrant with vertex (x, y) (fig. 31.1).

Given the joint distribution of random variables X and Y, one can find the distributions of each of the random variable X and Y, known as the ***marginal distributions***:

$$F_X(x) = P(X < x) = P(X < x, Y < +\infty) = F_{X,Y}(x, +\infty),$$

$$F_Y(y) = P(Y < y) = P(X < +\infty, Y < y) = F_{X,Y}(+\infty, y).$$

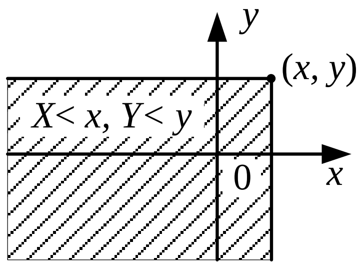


Fig. 31.1. **Geometrical interpretation of the distribution function $F(x, y)$**

The ***marginal distributions*** don't completely characterize the two-dimensional random variable (X, Y), i.e. the joint distribution of the random variables X and Y can't in general be reconstructed from the marginal distributions.

Let's enumerate *properties* of the distribution function $F(x, y)$ (which are similar to properties of one-dimensional random variable).

1. $0 \le F(x, y) \le 1$, because $0 \le P(X < x, Y < y) \le 1$.

2. $F(x, y)$ is a nondecreasing function of each of the arguments x and y.

3. If one of the arguments of $F(x, y)$ approaches $+\infty$, then this function tends to the distribution function of another argument which doesn't approach $+\infty$, i.e.:

$$\lim_{y \to \infty} F(x, y) = F(x, \infty) = F(x), \tag{31.2}$$

$$\lim_{x \to \infty} F(x, y) = F(\infty, y) = F(y). \tag{31.3}$$

4. If both arguments of $F(x, y)$ tend to $+\infty$, then this function approaches 1, i.e.:

$$\lim_{\substack{x \to \infty \\ y \to \infty}} F(x, y) = F(\infty, \infty) = P(x < \infty, y < \infty) = 1.$$

5. If both or one of the arguments of $F(x, y)$ tends to $-\infty$, then this function approaches 0, i.e.:

$$\lim_{\substack{x\to-\infty \\ y\to-\infty}} F(x,y)=\lim_{x\to-\infty} F(x,y)=\lim_{y\to-\infty} F(x,y)=0$$

6. If components X and Y are independent $F(x,y)=F_1(x)\cdot F_2(y)$.

7. The probability that the random vector (X,Y) lies into an arbitrarily rectangle $(a<X<b, c<Y<d)$ with sides parallel to the coordinate axes is calculated by the following formula:

$$P(a<X<b, c<Y<d)=F(b,d)+F(a,c)-F(a,d)-F(b,c). \quad (31.4)$$

8. The function $F(x, y)$ is left continuous in each of the arguments.

A ***two-dimensional random vector*** is said to be ***discrete*** if each of the random variables X and Y is discrete.

If the random variable X takes the values $x_1, x_2, \ldots, x_n$ and the random variable Y takes the values $y_1, y_2, \ldots, y_m$, then the random vector (X,Y) can take only the pairs of values.

A ***distribution law of a discrete two-dimensional random vector*** is called a set of possible values $X=x_i$, $Y=y_j$ $(i=\overline{1,n};\ j=\overline{1,m})$ and their corresponding probabilities of joint occurrence $p_{ij}=P(X=x_i, Y=y_j)=$ $=p(x_i, y_j)$ under condition $\sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i, y_j)=1$.

It is convenient to describe the distribution of a two-dimensional discrete random variable using the distribution law shown in table 31.1. Here each cell (i, j) contains the probability $p(x_i, y_j)$ of a distribution of events $(X=x_i, Y=y_j)$ $(i=\overline{1,n};\ j=\overline{1,m})$.

Since events $(X=x_i, Y=y_j)$ $(i=\overline{1,n};\ j=\overline{1,m})$ which mean that the random variable X possesses the value x_i and the random variable Y possesses the value y_j are disjoint and only possible, i.e. form a complete group of events, then a sum of their probabilities equals 1, i.e.

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

Table 31.1

The distribution law of a discrete two-dimensional random vector (X,Y)

$X \backslash Y$	y_1	y_2	$\dots$	y_m	$\sum_{i=1}^{n} p_{x_i}$
x_1	p_{11}	p_{12}	$\dots$	p_{1m}	p_{x_1}
x_2	p_{21}	p_{22}	$\dots$	p_{2m}	p_{x_2}
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
x_n	p_{n1}	p_{n2}	$\dots$	p_{nm}	p_{x_n}
$\sum_{j=1}^{m} p_{y_j}$	p_{y_1}	p_{y_2}	$\dots$	p_{y_m}	1

A balance column or raw of this table of distribution (X,Y) gives probabilities respectively for distribution laws of one-dimensional components $(x_i,\, p_i)$ or $(y_j,\, p_j)$.

In order to find the probability that one-dimensional random variable possesses a definite value using the table of distribution (see table 31.1) it is necessary to summarize probabilities p_{ij} of the corresponding raw (column) for this value of the given table:

$$p_{x_i} = \sum_{j=1}^{m} p_{ij}\,, \qquad i = \overline{1,\, n}, \tag{31.6}$$

or

$$p_{y_j} = \sum_{i=1}^{n} p_{ij}\,, \qquad j = \overline{1,\, m}. \tag{31.7}$$

If some numbers p_{ij} are equal to zero then in this case an occurrence of this random variable is impossible.

On the base of the distribution law of a random vector (X,Y) (tabl. 31.1) we can define the distribution law of each one-dimensional random variable X and Y. Since the event $(X = x_i)$ is a sum of events

$(X = x_i, Y = y_1)$, $(X = x_i, Y = y_2)$, …, $(X = x_i, Y = y_m)$, then its probability is calculated by the formula (31.6).

The distribution law of one-dimensional random variable X at calculated values p_{x_i} can be given in a form of a table 31.2 and similarly to that a distribution law of one-dimensional random variable Y is given in a form of a table 31.3.

Table 31.2

A distribution law of one-dimensional random variable X

Values of X	x_1	x_2	…	x_n
Probabilities	p_{x_1}	p_{x_2}	…	p_{x_n}

Table 31.3

A distribution law of one-dimensional random variable Y

Values of Y	y_1	y_2	…	y_m
Probabilities	p_{y_1}	p_{y_2}	…	p_{y_m}

31.3. Absolute continuous distributions. A density of joint distribution and its properties. Marginal densities of distribution of components of a random vector

In addition to a distribution function $F(x, y)$ a characteristic of a system of two absolutely continuous random variables is a density of a distribution of probabilities $f(x, y)$.

A ***density of a distribution of probabilities*** $f(x, y)$ for a system of two absolutely continuous random variables (X, Y) is the second mixed derivative of its distribution function $F(x, y)$:

$$f(x, y) = \frac{\partial^2 F(x, y)}{\partial x \, \partial y}. \qquad (31.8)$$

This function $f(x, y)$ can exist under condition that $F(x, y)$ is continuous relative to its arguments x and y and twice differentiable.

Geometrically the function $f(x, y)$ in three-dimensional space can be compared with a definite surface which is called *a distribution surface of* probabilities of a system of two absolutely continuous random variables (X, Y) (fig. 31.3).

1. The function $f(x, y) \geq 0$, because $F(x, y)$ is nondecreasing relative to arguments x and y.

2. For this function $f(x, y)$ the normalization condition is fulfilled, i.e.:

$$\iint_{\Omega} f(x, y)dxdy = 1, \tag{31.9}$$

where Ω is a domain of a definition of an absolutely continuous random variable; $f(x, y)dxdy$ is the probability of a location of a system of two absolutely continuous random variables (X, Y) in the rectangle with sides $dx,\ dy$ (see fig. 31.3).

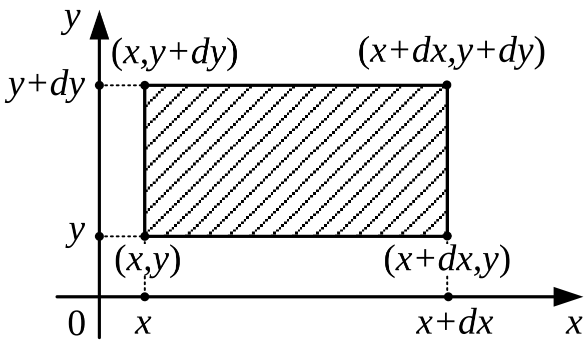


Fig. 31.3. **A distribution surface of probabilities of a system of two absolutely continuous random variables (X, Y)**

A density of a distribution of probabilities has the following properties $f(x, y)$:

If $\Omega = \{-\infty < x < +\infty, -\infty < y < +\infty\}$, then (31.9) has the following form:

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x, y)dxdy = 1. \tag{31.10}$$

3. The relationship between a distribution function of probabilities of a system of two random variables and a density of probabilities is defined by:

$$F(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(x, y) dxdy.$$

4. If $\Omega = \{a < x < b, c < y < d\}$, then:

$$F(x, y) = \int_{a}^{x} \int_{c}^{y} f(x, y) dxdy. \qquad (31.11)$$

5. The probability of a location of a system of two random variables (x, y) into the domain $D \subset \Omega$ is calculated by the following formula:

$$P((x, y) \in D) = \iint_{D} f(x, y) dxdy.$$

The probability of a location of a system of two random variables (x, y) into the rectangular domain $D = \{a < x < b, c < y < d\}$:

$$P(a < x < b, c < y < d) = \int_{a}^{b} \int_{c}^{d} f(x, y) dxdy. \qquad (31.12)$$

6. If $f_1(x)$, $f_2(y)$ are distribution densities of each component, then

$$f_1(x) = \int_{-\infty}^{+\infty} f(x, y) dy, \qquad f_2(y) = \int_{-\infty}^{+\infty} f(x, y) dx$$

$$\int_{-\infty}^{+\infty} f_1(x) dx = 1, \qquad \int_{-\infty}^{+\infty} f_2(y) dy = 1. \qquad (31.13)$$

7. If components X and Y are independent, then:

$$f(x, y) = f_1(x) \cdot f_2(y).$$

31.4. Conditional laws of probabilities distribution of a random vector. A characteristic of a set of independent random variables

31.4.1. Conditional laws of probabilities distribution of a discrete random vector. Let's consider a discrete two-dimensional random variable $Z(X, Y)$. Possible values of its components are $x_1, x_2, \ldots, x_n$ and $y_1, y_2, \ldots, y_m$.

Let's suppose as the result of a trial the random variable Y has possessed the value y_1 $(Y = y_1)$ then the random variable X can take one of possible values: $x_1, x_2, \ldots, x_n$. Let's denote the conditional probability that $X = x_i$ under condition $Y = y_1$ as $p_{y_1}(x_i)$ or $p(x_i / y_1)$ ($i = \overline{1, n}$).

In the general case we denote conditional probabilities of the component X under condition $Y = y_j$ as $p(x_i / y_j)$ ($i = \overline{1, n}$; $j = \overline{1, m}$) and conditional probabilities of the component Y under condition $X = x_i$ as $p(y_j / x_i)$.

A conditional distribution of the component X under condition $Y = y_j$ is called a set of conditional probabilities $p(x_i / y_j)$ under condition the case $Y = y_j$ has occurred. Similarly to that we define a conditional distribution of the component Y under condition $X = x_i$.

According to a distribution of a two-dimensional variable $Z(X,Y)$ we can get conditional laws of a distribution of components X and Y:

for X:
$$p(x_i / y_j) = \frac{p(x_i, y_j)}{p(y_j)},$$

for Y:
$$p(y_j / x_i) = \frac{p(x_i, y_j)}{p(x_i)}.$$

It is necessary to mark that a sum of probabilities of a conditional distribution of each component equals 1.

31.4.2. Conditional laws of probabilities distribution of a continuous random vector. In the case of ***continuous distribution*** of a variable

$Z(X,Y)$ we have conditional densities of a distribution of the component X under condition $Y = y_j$ and the component Y under condition $X = x_i$.

A conditional density $f_1(x/y)$ of a distribution of the component X under condition $Y = y_j$ is called a ratio of a density of joint distribution $f(x,y)$ of a system (X,Y) to a density of a distribution $f_2(y)$ of the component Y:

$$f_1(x/y) = \frac{f(x,y)}{f_2(y)}.$$

Similarly to that a conditional density $f_2(y/x)$ of a distribution of the component Y under condition $X = x_i$ is defined by the formula:

$$f_2(y/x) = \frac{f(x,y)}{f_1(x)}.$$

If a density of joint distribution $f(x,y)$ is known then we can find $f_1(x)$ and $f_2(y)$ using the formulas (31.13).

Let's write down the following properties for $f_1(x/y)$ and $f_2(y/x)$:

$$f_1(x/y) \geq 0, \qquad \int_{-\infty}^{+\infty} f_1(x/y)dx = 1,$$

$$f_2(y/x) \geq 0, \qquad \int_{-\infty}^{+\infty} f_2(y/x)dy = 1.$$

31.5. Numerical characteristics of joint (consistent) distributions of systems of random variables: marginal and conditional

A conditional mathematical expectation of a discrete random variable Y under condition $X = x_i$ is called a product of possible values of the component Y and their conditional probabilities:

$$M(Y/X = x_i) = \sum_{j=1}^{m} y_j p(y_j/x_i).$$

A conditional mathematical expectation $M(Y/X=x_i)$ is a function of x, i.e. $M(Y/X=x_i)=f(x)$, which is called the regression function of Y on X.

Similarly to this we have formulas for a conditional mathematical expectation $M(X/Y=y_j)$ of a discrete random variable X under condition $Y=y_j$:

$$M(X/Y=y_j)=\sum_{i=1}^{n} x_i p(x_i/y_j), \qquad M(X/Y=y_j)=g(y),$$

where $g(y)$, which is called the regression function of X on Y.

For continuous random variables we have:

$$M(Y/X=x)=\int_{-\infty}^{+\infty} y\cdot f_2(y/x)dy, \qquad M(X/Y=y)=\int_{-\infty}^{+\infty} x\cdot f_1(x/y)dx.$$

31.6. Numerical characteristics of a system of two random variables. The covariance and the correlation coefficient of a two-dimensional random vector

For two-dimensional random variable $Z(X,Y)$ we can find a mathematical expectation and a variance of each component:

$$M(X)=m_x, \quad M(Y)=m_y, \quad D(X)=\sigma_x^2, \quad D(Y)=\sigma_y^2.$$

However, these characteristics don't completely characterize the variable $Z(X,Y)$, therefore they don't indicate the degree of the dependence between components. This role is fulfilled by the covariance (or the correlation moment) μ_{xy} and the correlation coefficient r_{xy}.

The ***covariance*** (or the ***correlation moment***) μ_{xy} of random variables X and Y is called the mathematical expectation of a product of derivations of these variables from their mathematical expectations:

$$\mu_{xy}=M\big((X-m_x)(Y-m_y)\big).$$

For calculation of the covariance (or the correlation moment) of discrete random variables the following formula is used:

$$\mu_{xy} = \sum_{i=1}^{n}\sum_{j=1}^{m}(x_i - m_x)(y_j - m_y)p(x_i, y_j),$$

for continuous random variables this formula is used:

$$\mu_{xy} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(x - m_x)(y - m_y)f(x, y)dxdy.$$

The covariance (or the correlation moment) μ_{xy} has the following *properties*:

1. $\mu_{xy} = \mu_{yx}$ is the symmetric property.
2. $\mu_{xy} = 0$ if X and Y are independent variables.
3. $|\mu_{xy}| \le \sigma_x \sigma_y$.
4. $\mu_{xx} = D(X)$.

The correlation moment of random variables X and Y has the dimension equal to the product of X and Y dimensions.

Along with the correlation moment of random variables X and Y one often uses the correlation coefficient r_{xy} which is a dimensionless normalized variable.

The correlation coefficient r_{xy} of random variables X and Y is the ratio of the correlation moment of X and Y to the product of their root-mean-square deviations (or standard deviations), i.e.

$$r_{xy} = \frac{\mu_{xy}}{\sigma_x \sigma_y}.$$

Properties of the correlation coefficient r_{xy}:

1. $r_{xy} = r_{yx}$.
2. $r_{xx} = 1$ and $r_{yy} = 1$.
3. $|r_{xy}| \le 1$.

4. $r_{xy} = 0$ if X and Y are independent variables, i.e. there is no linear relation between the random variables.

5. $r_{xy} = \pm 1$ if there exists the linear correlation dependence between X and Y.

Recommended bibliography: [5; 7; 11; 12].

Theme 32. Laws of large numbers and central limiting theorem

Let's consider fundamental theorems of probability theory. We can find an intuitive way to view the probability of a certain outcome as the frequency with which the outcome occurs in long run, when the experiment is repeated a large number of times. We can also define probability mathematically as a value of a distribution function for the random variable representing the experiment.

32.1. A convergence of sequences of random variables in a probability and almost surely

Let's consider basic concepts.

A sequence of random variables $X_1, X_2, \ldots$ is said to be ***converge in probability*** to a random variable X if

$$\lim_{n \to \infty} P(|X_n - X| \geq \varepsilon) = 0$$

for each $\varepsilon > 0$, i.e. if for any $\varepsilon > 0$ and $\delta > 0$ there exists a number N, depending on ε and δ, such that the inequality

$$P(|X_n - X| > \varepsilon) < \delta$$

holds for $n > N$.

A sequence of random variables $X_1, X_2, \ldots$ is said to be ***converge almost surely*** (or ***with probability 1***) to a random variable X if

$$P\left[\omega \in \Omega : \lim_{n \to \infty} X_n(\omega) = X(\omega)\right] = 1.$$

Convergence almost surely implies convergence in probability.

32.2. Inequalities of Markov and Chebyshev. Laws of large numbers and conditions of their fulfillment

The law of large numbers consists of several theorems establishing average results and revealing conditions for this stability to occur.

The notion of convergence in probability is most often used for the case in which the limit random variable X has the degenerate distribution concentrated at a point a $(P(\xi = a) = 1)$ and

$$X_n = \frac{1}{n}\sum_{k=1}^{n} Y_k ,$$

where $Y_1, Y_2, \dots$ are arbitrary random variables.

A sequence $Y_1, Y_2, \dots$ satisfies the ***weak law of large numbers*** if the limit relation

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{k=1}^{n} Y_k - a\right| \geq \varepsilon\right) \equiv \lim_{n\to\infty} P(|X_n - a| \geq \varepsilon) = 0 \qquad (32.1)$$

holds for any $\varepsilon > 0$. Equivalently, $\lim\limits_{n\to\infty} P(|X_n - a| < \varepsilon) = 1$.

If the relation

$$P\left(\omega \in \Omega : \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} Y_k = a\right) = P\left(\omega \in \Omega : \lim_{n\to\infty} X_n = a\right) = 1$$

is satisfied instead of (32.1), i.e. the sequence X_n converges to the number a with probability 1, then the sequence $Y_1, Y_2, \dots$ satisfies the ***strong law of large numbers***.

Let's consider important inequalities.

Markov inequality. For any nonnegative random variable X that has an expectation $M(X)$, the inequality

$$P(X \geq \varepsilon) \leq \frac{M(X)}{\varepsilon^2}$$

holds for each $\varepsilon > 0$.

Chebyshev inequality. For any random variable X with finite variance $D(X)$, the inequality

$$P(|X - M(X)| \le \varepsilon) \le 1 - \frac{D(X)}{\varepsilon^2}$$

holds for each $\varepsilon > 0$.

This inequality gives the possibility to estimate an error if we suppose that the mathematical expectation is replaced by the average value of a bounded sample.

Proof. Events $|X - M(X)| \le \varepsilon$ and $|X - M(X)| > \varepsilon$ form a complete group of events, i.e.

$$P(|X - M(X)| \le \varepsilon) + P(|X - M(X)| > \varepsilon) = 1. \qquad (32.2)$$

Let's remember the definition of the variance $D(X)$:

$$D(X) = M(X - M(X))^2 = \sum_{i=1}^{n} (x_i - M(X))^2 \cdot p_i .$$

If we truncate the summands in which $|x_i - M(X)| \le \varepsilon$, then

$$D(X) \ge \sum_{i=1}^{k} (x_i - M(X))^2 \cdot p_i \ge \varepsilon^2 \sum_{i=1}^{k} p_i = \varepsilon^2 \cdot P(|X - M(X)| > \varepsilon).$$

Thus, we obtain:

$$D(X) \ge \varepsilon^2 \cdot P(|X - M(X)| > \varepsilon) \text{ or } P(|X - M(X)| > \varepsilon) \le \frac{D(X)}{\varepsilon^2}.$$

Using the formula (32.2) we have:

$$P(|X - M(X)| \le \varepsilon) = 1 - P(|X - M(X)| > \varepsilon) \ge 1 - \frac{D(X)}{\varepsilon^2}.$$

The law of large numbers gives a relation between the probability $P(A)$

of a random event A and its relative frequency $\frac{m}{n}$ with a large number of repeated experiments.

Chebyshev theorem. If $X_1, X_2, \ldots, X_n$ is a sequence of independent random variables with uniformly bounded finite variances ($D_1(X) \le D$, $D_2(X) \le D$, …, $D_n(X) \le D$) then the limit relation

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} M(X_i)\right| \le \varepsilon\right) = 1$$

holds for each $\varepsilon > 0$.

Bernoulli theorem. Let m_n be the number of occurrences of an event A (the number of successes) in n independent trials and let $p = P(A)$ be the probability of the occurrence of the event A (the probability of success) in each of the trials. Then the sequence of relative frequencies m_n / n of the occurrences of the event A in n independent trials converges in probability to $p = P(A)$ as $n \to \infty$, i.e. the limit relation

$$\lim_{n\to\infty} P\left(\left|\frac{m_n}{n} - p\right| < \varepsilon\right) = 1$$

holds for each $\varepsilon > 0$.

32.3. A convergence in distribution and a weak convergence

Let's suppose that a sequence $F_1(x), F_2(x), \ldots, F_n(x)$ of cumulative distribution functions converges to a distribution function $F(x)$, i.e.

$$\lim_{n\to\infty} F_n(x) = F(x)$$

for every point x at which $F(x)$. In this case, we say that the sequence $X_1, X_2, \ldots, X_n$ of the corresponding random variables ***converges to the random variable X in a distribution***.

A sequence $F_1(x), F_2(x), \ldots, F_n(x)$ of cumulative distribution functions ***weakly converges*** to a distribution function $F(x)$ if:

$$\lim_{n\to\infty} M(g(X_n)) = M(g(X))$$

for any bounded continuous function g as $n \to \infty$.

Convergence in distribution and weak convergence of cumulative distribution functions are equivalent.

32.4. Central limit theorem. Lyapunov theorem for sequence of an independent identically distributed random variable

A sequence X_n of random with distribution function F_{X_n} is called ***asymptotically normally distributed*** if there exists a sequence of pairs of real numbers m_n, σ_n^2 such that the random variables:

$$\left(\frac{X_n - m_n}{\sigma_n}\right)$$

converge in probability to a standard normal variable. This occurs if and only if the limit relation:

$$\lim_{n\to\infty} P\left(\alpha \le \frac{X_n - a}{\sigma_n} \le \beta\right) = \Phi(\beta) - \Phi(\alpha)$$

where $\Phi(x)$ is Laplace cumulative distribution function (appendix B), holds for any α and β ($\alpha \le \beta$).

Lyapunov theorem. If $X_1, X_2, \ldots, X_n$ is a sequence of independent random variables satisfying the Lyapunov condition:

$$\lim_{n\to\infty} \frac{\sum_{i=1}^{n} \nu_3(X_i)}{\sqrt{\sum_{i=1}^{n} D(X_i)}} = 0,$$

where $\nu_3(X_i)$ is the third initial moment of the random variable X_i, then the sequence of random variables $Y_n = \frac{\sum_{i=1}^{n}(X_i - M(X_i))}{\sqrt{\sum_{i=1}^{n} D(X_i)}}$ converges in distribution to the normal law, i.e. the following limit exists:

$$\lim_{n\to\infty} P\left(\frac{\sum_{i=1}^{n}(X_i - M(X_i))}{\sqrt{\sum_{i=1}^{n} D(X_i)}} < x\right) = \Phi(x).$$

***Central limit theorem*.** Let m_n be the number of occurrences of an event A (the number of successes) in n independent trials and let $p = P(A)$ be the probability of the occurrence of the event A (the probability of success) in each of the trials. Then the sequence of relative frequencies m_n / n of the occurrences of the event A in n independent trials has an asymptotically normal probability distribution with parameters $(p,\ p(1-p)/n)$.

Let $X_1, X_2, \ldots, X_n$ be a sequence of independent identically distributed variables with finite mathematical expectation $M(X_i) = a$ and finite variance σ^2. Then as $n \to \infty$ the random variable $\frac{1}{n}\sum_{i=1}^{n} X_i$ has an asymptotically normal probability distribution with parameters $(a, \sigma^2 / n)$.

This theorem can be interpreted as stating for large n, i.e. the sequence X_n of random variables approximately has a normal distribution with mean a and standard deviation $\sigma / \sqrt{n}$.

Recommended bibliography: [5; 6; 7; 9; 11].

Theoretical questions

1. A stochastic experiment. A random event. A probabilistic space.
2. An outcome. An impossible event. A sure event.
3. Equally likely events. Elementary events.
4. An intersection, a union, a difference of events.
5. Theorem of a sum of compatible events.
6. Theorem of a sum of incompatible events.
7. A classical definition of a probability.
8. A geometrical definition of a probability.
9. A statistical definition of a probability.
10. Permutations, arrangements, combinations with repetitions.
11. Permutations, arrangements, combinations without repetitions.
12. The rule of a sum.
13. The rule of a product.
14. Inclusion-exclusion principle.
15. A conditional probability.
16. Theorem of a product for dependent events.
17. Theorem of a product for independent events.
18. A notion of a pairwise independence of random events.
19. A complete group of events.
20. Formulas of a total probability and Bayes.
21. Repeated independent trials.
22. Bernoulli’s scheme.
23. A binomial distribution.
24. The most probable number of successes and its probability.
25. Local theorem of Moivre–Laplace.
26. Integral theorem of Moivre–Laplace.
27. Poisson’s theorem.
28. Probability of deviation of relative frequency from probability.
29. A definition of random variables and their classification.
30. A distribution law of a discrete random variable.
31. The numerical characteristics of a distribution: a mathematical expectation, a variance, a root-mean-square deviation, initial and central moments, a mode, a median.
32. Binomial distribution law and its characteristics.

33. Geometric distribution law and its characteristics.
34. Hypergeometric distribution law and its characteristics.
35. Poisson distribution law and its characteristics.
36. Negative binomial distribution law and its characteristics.
37. A definition of continuous random variables.
38. A definition of absolutely continuous random variables.
39. Distribution function of probabilities of random variables and its properties.
40. A density of distribution and its properties.
41. Distribution density functions of absolutely continuous random variables.
42. Numerical characteristics of absolutely continuous random variables and their properties.
43. A uniform law of probabilities distribution and its numerical characteristics.
44. An exponential law of distribution and its numerical characteristics.
45. A normal law of probabilities distribution and its standard representation.
46. Gamma-distribution.
47. Chi-square distribution of Student and its relationship with a standard normal law.
48. Chi-square distribution of Fisher and its relationship with a standard normal law.
49. A random vector. A system of two random variables.
50. A discrete random vector.
51. A continuous random vector.
52. Numerical characteristics of consistent distributions of systems of random variables: marginal and conditional.
53. A covariance of a two-dimensional random vector.
54. A correlation coefficient of a two-dimensional random vector.
55. A convergence in probability.
56. An almost surely convergence.
57. Inequality of Markov. Inequality of Chebishev.
58. Laws of large numbers and conditions of their fulfillment.
59. Chebyshev theorem. Lyapunov theorem.
60. Bernoulli theorem.
61. A convergence in distribution and a weak convergence
62. Central limit theorem.

Recommended bibliography

1. Англо-русский словарь математических терминов / под ред. П. С. Александрова. – М. : Мир, 1994. – 416 с.
2. Бузько Я. П. Теорія ймовірностей та математична статистика курсу "Математика для економістів". Тексти лекцій / Я. П. Бузько, О. О. Єгоршин, Н. В. Панова. – Х. : РВВ ХНЕУ, 1999. – 96 с.
3. Вправи з курсу "Математика для економістів" для студентів усіх спеціальностей усіх форм навчання. Ч. 3 / укл. Е. Ю. Железнякова, А. В. Ігначкова, З. Г. Попова. – Х. : РВВ ХНЕУ, 2003. – 108 с.
4. Гмурман В. Е. Руководство по решению задач по теории вероятностей и математической статистике / В. Е. Гмурман. – М. : Высшая школа, 2001. – 576 с.
5. Гмурман В. Е. Теория вероятностей и математическая статистика : учебн. пособие для вузов / В. Е. Гмурман. – 6-е изд. – М. : Высшая школа, 1998. – 480 с.
6. Кремер Н. Ш. Теория вероятностей и математическая статистика / Н. Ш. Кремер. – М. : ЮНИТИ-ДАНА, 2000. – 544 с.
7. Теорія ймовірностей. Конспект лекцій / О. К. Шевченко, Г. К. Снурнікова, О. Д. Анохіна. – Х. : Вид. ХНЕУ, 2009. – 88 с.
8. Guidelines for practical tasks of the educational discipline "Higher and applied mathematics" for foreign and English-learning full-time students of the direction 6.140103 "Tourism" (Random variables and their characteristics) / compiled by Ie. Misiura. – Kh. : Publishing House of KhNUE, 2012. – 36 p. (English, Ukrainian)
9. Kelbert M. Probability and statistics. Volume 1. Basic probability and statistics / Yuri Suhov, Mark Kelbert. – New York; Madrid : Cambridge University Press, 2005. – 360 p.
10. Ross S. Introduction to probability and mathematical statistics / Sheldon Ross. – San Diego : Elsevier Academic Press, 2004. – 641 p.
11. Tavare S. Lectures on probability theory and statistics / Simor Tavare, Ofer Zeitouni. – Berlin; Heidelberg : Springer-Verlag, 2004. – 322 p.

12. Probability and statistics for engineers and scientists / R. E. Walpole, R. H. Myers, S. L. Myers, et al. – London : Pearson education LTD, 2007. – 823 p.

Appendices

Appendix A

(to be continued)

Values of Laplace differential function $\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$

x	0	1	2	3	4	5	6	7	8	9
0,0	0,3989	0,3989	0,3989	0,3988	0,3986	0,3984	0,3982	0,3980	0,3977	0,3973
0,1	0,3970	0,3965	0,3961	0,3956	0,3951	0,3945	0,3939	0,3932	0,3925	0,3918
0,2	0,3910	0,3902	0,3894	0,3885	0,3876	0,3867	0,3857	0,3847	0,3836	0,3825
0,3	0,3814	0,3802	0,3790	0,3778	0,3765	0,3752	0,3739	0,3726	0,3712	0,3697
0,4	0,3683	0,3668	0,3652	0,3637	0,3621	0,3605	0,3589	0,3572	0,3555	0,3538
0,5	0,3521	0,3503	0,3485	0,3467	0,3448	0,3429	0,3410	0,3391	0,3372	0,3352
0,6	0,3332	0,3312	0,3292	0,3271	0,3251	0,3230	0,3209	0,3187	0,3166	0,3144
0,7	0,3123	0,3101	0,3079	0,3056	0,3034	0,3011	0,2989	0,2966	0,2943	0,2920
0,8	0,2897	0,2874	0,2850	0,2827	0,2803	0,2780	0,2756	0,2732	0,2709	0,2685
0,9	0,2661	0,2637	0,2613	0,2589	0,2565	0,2541	0,2516	0,2492	0,2468	0,2444
1,0	0,2420	0,2396	0,2371	0,2347	0,2323	0,2299	0,2275	0,2251	0,2227	0,2203
1,1	0,2179	0,2155	0,2131	0,2107	0,2083	0,2059	0,2036	0,2012	0,1989	0,1965
1,2	0,1942	0,1919	0,1895	0,1872	0,1849	0,1826	0,1804	0,1781	0,1758	0,1736
1,3	0,1714	0,1691	0,1669	0,1647	0,1626	0,1604	0,1582	0,1561	0,1539	0,1518
1,4	0,1497	0,1476	0,1456	0,1435	0,1415	0,1394	0,1374	0,1354	0,1334	0,1315
1,5	0,1295	0,1276	0,1257	0,1238	0,1219	0,1200	0,1182	0,1163	0,1145	0,1127
1,6	0,1109	0,1092	0,1074	0,1057	0,1040	0,1023	0,1006	0,0989	0,0973	0,0957
1,7	0,0940	0,0925	0,0909	0,0893	0,0878	0,0863	0,0848	0,0833	0,0818	0,0804
1,8	0,0790	0,0775	0,0761	0,0748	0,0734	0,0721	0,0707	0,0694	0,0681	0,0669
1,9	0,0656	0,0644	0,0632	0,0620	0,0608	0,0596	0,0584	0,0573	0,0562	0,0551
2,0	0,0540	0,0529	0,0519	0,0508	0,0498	0,0488	0,0478	0,0468	0,0459	0,0449
2,1	0,0440	0,0431	0,0422	0,0413	0,0404	0,0396	0,0387	0,0379	0,0371	0,0363
2,2	0,0355	0,0347	0,0339	0,0332	0,0325	0,0317	0,0310	0,0303	0,0297	0,0290
2,3	0,0283	0,0277	0,0270	0,0264	0,0258	0,0252	0,0246	0,0241	0,0235	0,0229
2,4	0,0224	0,0219	0,0213	0,0208	0,0203	0,0198	0,0194	0,0189	0,0184	0,0180
2,5	0,0175	0,0171	0,0167	0,0163	0,0158	0,0154	0,0151	0,0147	0,0143	0,0139
2,6	0,0136	0,0132	0,0129	0,0126	0,0122	0,0119	0,0116	0,0113	0,0110	0,0107
2,7	0,0104	0,0101	0,0099	0,0096	0,0093	0,0091	0,0088	0,0086	0,0084	0,0081
2,8	0,0079	0,0077	0,0075	0,0073	0,0071	0,0069	0,0067	0,0065	0,0063	0,0061
2,9	0,0060	0,0058	0,0056	0,0055	0,0053	0,0051	0,0050	0,0048	0,0047	0,0046
3,0	0,0044	0,0043	0,0042	0,0040	0,0039	0,0038	0,0037	0,0036	0,0035	0,0034

x	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
3,1	0,0033	0,0032	0,0031	0,0030	0,0029	0,0028	0,0027	0,0026	0,0025	0,0025
3,2	0,0024	0,0023	0,0022	0,0022	0,0021	0,0020	0,0020	0,0019	0,0018	0,0018
3,3	0,0017	0,0017	0,0016	0,0016	0,0015	0,0015	0,0014	0,0014	0,0013	0,0013
3,4	0,0012	0,0012	0,0012	0,0011	0,0011	0,0010	0,0010	0,0010	0,0009	0,0009
3,5	0,0009	0,0008	0,0008	0,0008	0,0008	0,0007	0,0007	0,0007	0,0007	0,0006
3,6	0,0006	0,0006	0,0006	0,0005	0,0005	0,0005	0,0005	0,0005	0,0005	0,0004
3,7	0,0004	0,0004	0,0004	0,0004	0,0004	0,0004	0,0003	0,0003	0,0003	0,0003
3,8	0,0003	0,0003	0,0003	0,0003	0,0003	0,0002	0,0002	0,0002	0,0002	0,0002
3,9	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0001	0,0001

Appendix B
(to be continued)

Values of Laplace cumulative distribution function $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt$

x	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
0,0	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0754
0,2	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0,3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
0,4	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1808	0,1844	0,1879
0,5	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224
0,6	0,2258	0,2291	0,2324	0,2356	0,2389	0,2422	0,2454	0,2486	0,2518	0,2549
0,7	0,2580	0,2612	0,2642	0,2673	0,2704	0,2734	0,2764	0,2794	0,2823	0,2852
0,8	0,2881	0,2910	0,2939	0,2967	0,2996	0,3023	0,3051	0,3078	0,3106	0,3133
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389
1,0	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621
1,1	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830
1,2	0,3849	0,3869	0,3888	0,3906	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015
1,3	0,4032	0,1049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177
1,4	0,4192	0,4207	0,4222	0,4236	0,4251	0,4265	0,4274	0,4292	0,4306	0,4319
1,5	0,4332	0,4345	0,4357	0,4370	0,4382	0,4394	0,4406	0,4418	0,4430	0,4441
1,6	0,4452	0,4463	0,4474	0,4484	0,4495	0,4505	0,4515	0,4525	0,4535	0,4545
1,7	0,4554	0,4564	0,4573	0,4582	0,4591	0,4599	0,4608	0,4616	0,4625	0,4633
1,8	0,4641	0,4648	0,4656	0,4664	0,4671	0,4678	0,4686	0,4693	0,4700	0,4706
1,9	0,4713	0,4719	0,4726	0,4732	0,4738	0,4744	0,4750	0,4756	0,4762	0,4757
2,0	0,4772	0,4778	0,4783	0,4788	0,4796	0,4798	0,4803	0,4808	0,48Ï2	0,4817
2,1	0,4821	0,4826	0,4830	0,4834	0,4838	0,4842	0,4846	0,4850	0,4854	0,4857
2,2	0,4861	0,4864	0,4868	0,4871	0,4874	0,4878	0,4881	0,4884	0,4887	0,4890
2,3	0,4893	0,4896	0,4898	0,4901	0,4903	0,4906	0,4909	0,4911	0,4913	0,4916
2,4	0,4918	0,4920	0,4922	0,4924	0,4927	0,4929	0,4930	0,1932	0,4934	0,4936
2,5	0,4938	0,4940	0,4941	0,4943	0,4945	0,4946	0,4948	0,4949	0,4951	0,4952
2,6	0,4953	0,4955	0,4956	0,4957	0,4958	0,4960	0,4961	0,4962	0,4963	0,4964
2,7	0,4965	0,4966	0,4967	0,4968	0,4969	0,4970	0,4971	0,4972	0,4973	0,4973
2,8	0,4974	0,4975	0,4976	0,4977	0,4977	0,4978	0,4979	0,4980	0,4980	0,4981
2,9	0,4981	0,4982	0,4982	0,4983	0,4984	0,4984	0,4985	0,4985	0,4986	0,4986
3,0	0,4986	0,4986	0,4987	0,4987	0,4988	0,4988	0,4988	0,4989	0,4989	0,4990

Appendix B (the ending)

x	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
3,1	0,4990	0,4990	0,4991	0,4991	0,4991	0,4992	0,4992	0,4992	0,4992	0,4993
3,2	0,4993	0,4993	0,4993	0,4994	0,4994	0,4994	0,4994	0,4994	0,4995	0,4995
3,3	0,4995	0,4995	0,4995	0,4996	0,4996	0,4996	0,4996	0,4996	0,4997	0,4997
3,4	0,4997	0,4997	0,4997	0,4997	0,4997	0,4998	0,4998	0,4998	0,4998	0,4998
3,5	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998
3,6	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4999	0,4999	0,4999	0,4999
3,7	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,8	0,4999	0,4999	0,4999	0,4999	0,4999	0,5000	0,5000	0,5000	0,5000	0,5000
3,9	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000

EDUCATIONAL EDITION

Misiura Ie.

PROBABILITY THEORY

Summary of lectures
for full-time students of the training direction
6.140103 "Tourism"

Editorial director **L. Malyarets**
Editor-in-chief **L. Syedova**

Editor **O. Alyoshina**
Proof-reader **O. Novytska**

НАВЧАЛЬНЕ ВИДАННЯ

Місюра Євгенія Юріївна

ТЕОРІЯ ЙМОВІРНОСТЕЙ

Конспект лекцій
для студентів напряму підготовки
6.140103 "Туризм"
денної форми навчання

(англ. мовою)

Відповідальний за випуск **Малярець Л. М.**
Відповідальний редактор **Сєдова Л. М.**

Редактор **Альошина О. М.**
Коректор **Новицька О. С.**

План 2013 р. Поз. № 89-К.
Підп. до друку Формат 60 x 90 1/16. Папір MultiCopy. Друк Riso.
Ум.-друк. арк. 6,0. Обл.-вид. арк. 7,5. Тираж прим. Зам. №

Видавець і виготівник — видавництво ХНЕУ, 61166, м. Харків, пр. Леніна, 9а

Свідоцтво про внесення до Державного реєстру суб'єктів видавничої справи
Дк № 481 від 13.06.2001 р.