A.K. Iskakova, A.B. Kokazhaeva, M.Zh. Baisalova

PROBABILITY THEORY LECTURES

Almaty, 2020

УДК 378
ББК 74.58
Р 19

Published by the Editorial Board of the Kazakh National Women's Pedagogical University on 14.11.2019 Presented by the decision of the protocol of №3

Reviewed by:

Baychemirov Zh.D. – PhD, Associate Professor of department «Mathematics and Mathematical modeling» Kazakh National teacher training university after Abayr.

Atakhan N. - PhD, senior teacher of department «Mathematics» Kazakh National Pedagogical University.

Iskakova A.K., Kokazhaeva A.B., Baisalova M.Zh.

P19 Probability theory lectures. Textbook. /- Almaty: /Kazakh National Women's Pedagogical University. 2020 - 88 pp.

ISBN 978-601-346-051-2

This textbook is intended for students of degree programs as mathematics, physics and computer science. It contains a theoretical course, many examples and tasks. The textbook compiled in accordance with the program of the course.

Fig. 6, Appl. 2, Bibl. 12.

Approved at a meeting of scientific and methodical council, recommended for printing in ________ of 2020.

УДК 378

ISBN 978-601-346-051-2

RANDOM EVENTS AND THEIR PROBABILITIES

Random events. Many phenomena in nature, the equipment, economy and in other areas have random character, meaning it is pretty hard to predict how things will occur. It turns out that the flow current of that phenomenon can be described quantitatively if only they were observed a sufficient number of times under unchanged conditions. So, for example, it is impossible to predict when throwing a coin, "Heads" or "Tails" will drop out. However, if one throws a coin very often, then it is possible to notice that the relation of the number thrown with loss of "Tails" to total number of throws differs a little from 1/2 and still deviates from ½ as many throwing are performed than.

The concept of probability gives mathematical modeling for the description of this type of random phenomena as objective reality. For experts who are into natural, technical, as well as social sciences it is important to know the bases of this theory as many real processes are susceptible to accidental impacts.

The random experiment or experience is a process outcomes are possible. It is impossible to predict what the result will be in advance. Experience is characterized by the fact that it is possible to repeat it as many times as you wish. Of particular importance are the many possible mutually exclusive outcomes of experience.

The possibility of excluding each other results of experience is called *Elementary Events*. The set of Elementary Events is designated through *E*. For example, casting a coin there is heads falling, i.e., H – «heads» is the elementary event, T – «tails» is the elementary event. For example, infinite Sample Space. Flip a coin until heads appears for the first time: $S = \{H, TH, TTN, TTTH, TTTTH, ...\}$.

Example 1. A one-time throwing of a playing dice. The possible results of this experiment excluding each other are in loss of one of the numbers 1, 2, 3, 4, 5, 6. The E set consists of six elementary events $e_1, e_2, e_3, e_4, e_5, e_6$, while the elementary event e_1 means: number 1 drops out.

Example 2. Simultaneous throwing of two dice. The E set of elementary events consists of 36 elements $e_{11}, e_{12}, \dots, e_{66}$, where by the elementary event $e_{i,j}$ means: on the first dice $\boldsymbol{i}$ drops out, and on the second $\boldsymbol{j}$.

Example 3. Determination of service duration of an electric lamp. Elementary events are all positive real numbers here. The E set, thus, consists of a positive valid half shaft.

Besides elementary events, it's often interesting to observe the events of more difficult nature, for example, in case of dice the event as "drops out even number" or in case of determination of duration of service of a lamp an event like "duration of service not less than 3000 hours".

Let's conduct an experiment and where *E* - set of its elementary events. Each subset $A \subseteq E$ is called *an Event*. The Event *A* takes place only in case when there is happening one of elementary events from which *A* consists.

Example 4. In the 1st example there are $A = \{e_2, e_4, e_6\}$ subsets forms an event «the even number drops out». The event «drops out even number» in case of dice takes place then and only then when there is one of the elementary events which contains subset *A*. In the 2nd example subset $A = \{e_{46}, e_{64}, e_{55}, e_{65}, e_{66}\}$ of a set *E* can be interpreted as an event «the sum of the dropped-out points is more or equal to 10». In the 3rd example $A = \{3\ 000, \infty\}$ can be interpreted as an event «the electric lamp serves more than 3000 hours».

E subset and consequently both the *E* set, and an empty set are interpreted according to the general definition as events. As *E* consists of all elementary events, and in each experience surely there is one of the elementary events, thus, *E* takes place always; such event is called *a Reliable Event*. We will designate a reliable event by letter *U*. The empty set doesn't contain elementary events and, therefore, never occurs; such event is called *an impossible event*, we will designate it as a letter *V*.

Empty set: The empty set is a set with no e1ements. We represent the *null* set with the symbol $\varnothing$. For any set *A*

$$A + \varnothing = A; \qquad A \cdot \varnothing = \varnothing .$$

Let $A_1, A_2, \dots, A_n$ events, i.e. subsets of some fixed *E* set of elementary events. Then association $A_1 \cup A_2 \cup \dots \cup A_n$ is again an event as it is an *E* subset. Association $A_1 \cup A_2 \cup \dots \cup A_n$ happens in only case when there is happening at least one of events $A_1, A_2, \dots, A_n$. The event $A_1 \cup A_2 \cup \dots \cup A_n$ is called *the sum* (*Union*) *of events*. It is often designated $A_1 + A_2 + \cdots + A_n$.

The union of two sets *A* and *B* is another set that includes all elements of *A* and all elements of *B*. We represent the union operator with this symbol $\cup$ or symbol +.

For example, if *A*={2,6,7,11} and *B*={2,3,7}, then

$$A \cup B = A + B = \{2,3,6,7,11\}.$$

In the same way the crossing $A_1 \cap A_2 \cap \dots \cap A_n$ of *A* events, is some event again. Crossing $A_1 \cap A_2 \cap \dots \cap A_n$ happens only in case when there are happening all *A* events at the same time. The event $A_1 \cap A_2 \cap \dots \cap A_n$ is often designated $A_1, A_2, \dots, A_n$ and called *intersection* of events. The intersection of two sets *A* and *B* is another set *C* such that all elements in *C* belong to *A* and to *B*. The intersection operator is symbolized as $\cap$. For example, if *A*={2,6,7,11} and *B*={2,3,7}, then

$A \cap B = AB = \{2,7\}$.

Example 5. In determining the life longevity A_1 there is an event "life longevity is between 0 and t_1", and - A_2 "life expectancy lies between t_1 and t_2", then $A_1 \cup A_2$ is an event "life expectancy between 0 and t_2".

Example 6. In the case of simultaneous throwing of two dices A_1 there is an event "sum of points more or equal to 11", A_2 - is the event "drops out identical quantity of points", then $A_1 \cap A_2$ is an event "two six drop out".

Example 7. In case of two dice A_1 - "the sum of the dropped-out points less or equal to 2", A_2 - the event "the sum of the dropped-out points is more or is equal to 5", then $A_1 \cap A_2$ - an impossible event $A_1 \cap A_2 = V$.

Two events A_1 and A_2 are called *antithetical* if they can't take place at the same time.

If A - some event, then addition $\overline{A} = E \backslash A$ also is an E subset i.e. some event. $\overline{A}$ occurs in only case when doesn't occur A; is called the event *opposite* (*additional*) to A. Events A and $\overline{A}$ are always incompatible: $A \cap \overline{A} = V$.

Example 8. If at some measurement A - an event "the measured size $\geq \alpha$", then $\overline{A}$ is an event "the measured size $< \alpha$".

For actions with random events (the sums, multiplication, additions) are fair the formulas of the elementary theory of sets:

The commutativity

$$A \cup B = B \cup A, \quad A \cap B = B \cap A,$$

The associativity

$$(A \cup B) \cup C = A \cup (B \cup C), \quad (A \cap B) \cap C = A \cap (B \cap C),$$

The distributivity

$$(A \cup B) \cap C = (A \cap C) \cup (B \cap C), \quad (A \cap B) \cup C = (A \cup C) \cap (B \cup C),$$

De Morgan's formulas

$$\overline{A \cup B} = \overline{A} \cap \overline{B}, \qquad \overline{A \cap B} = \overline{A} \cup \overline{B},$$

$$A \cup \overline{A} = U, \quad A \cap \overline{A} = V.$$

To construct the theory of chances, it is necessary to put events in compliance of probability which will give a quantitative assessment to a possibility of their implementation.

If E is incalculable (as, for example, in experience "establishing length of duty of an electric lamp"), then it is reasonable not to put probability in compliance for all E subsets. It is necessary to put the limit to a certain class of events. Events of this class are called random events. That is subject to probability theory.

If E of finite, then this class coincides with a class of all events. It turns out that the class of random events can always be chosen so that, first, there wouldn't be any mathematical difficulties at introduction of probabilities and, secondly, all events that interest us in practice would be in the chosen class, i.e. served as random events in mathematical sense. Therefore, for the scientist and the engineer it is important to know that all random events that interest him are also random in sense of the mathematical theory so the stated below is applicable to them.

Axioms of probability theory. Let the experiment be repeated n times and meanwhile being counted as many times as it happens. Suppose that it occurred m times. The ratio

$$\frac{m}{n} = W_n(A)$$

is called the *relative frequency* (or briefly - the frequency) of a random event A in n experiments. Where

$$0 \le W_n(A) \le 1.$$

Practice shows that as n increases, the frequency tends to a certain constant value. At the beginning of the concept of probability, a random event was tried to determine as the frequency limit. This led to theoretical and mathematical difficulties that could not be overcome. In modern theory, does not attempt to define the concept of probability. It is considered as the basic concept that satisfies certain axioms.

Let's consider the following properties of frequency, which can be regarded as experimental facts:

1. For large n, the frequency $W_n(A)$ oscillates less and less around a certain value.

2. $W_n(U) = 1$.

3. if A and B are incompatible

$$W_n(A \cup B) = W_n(A) + W_n(B).$$

Axioms of probability theory.

1. Each random event A is associated with a number $P(A)$, $0 \le P(A) \le 1$ that is called *probability A*.

2. The probability of a reliable event is equal to 1: $P(U) = 1$.

3. The axiom of additivity. If $A_1, A_2, \dots, A_n$ pairwise incompatible random events, i.e. $A_i \cap A_j = V$ at $i \neq j$, then

$$P(\bigcup_i A_i) = \sum_i P(A_i) \quad . \tag{1}$$

For a finite number of pairwise incompatible events $A_1, A_2, \dots, A_n$ the additivity axiom gives the relation

$$P(A_1 \cup A_2 \cup \dots \cup A_n) = P(A_1) + P(A_2) + \dots + P(A_n),$$

in particular

$$1 = P(U) = P(U \cup V) = P(U) + P(V) = 1 + P(V),$$

from which follows $P(V) = 0$, i.e. the probability of an impossible event is zero. From (1) follows that

$$P(A \cup \overline{A}) = P(A) + P(\overline{A}),$$

since A, $\overline{A}$ are inconsistent. Taking into account that $A \cup \overline{A} = U$ and axiom 2, we obtain

$$P(\overline{A}) = 1 - P(A) \quad . \tag{2}$$

For arbitrary (not necessarily pairwise incompatible) random events $A_1, A_2, \dots, A_n$ is inequality

$$P(A_1 \cup A_2 \cup \dots \cup A_n) \le \sum_{i=1}^{n} P(A_i) \quad . \tag{3}$$

The actual definition of the probability of some random event is purely theoretical and often impossible. In these cases, it is necessary to make a sufficient number of tests and take the relative frequency of the event in question as an approximation of probability.

So, for example, there is no mathematical or biological theory that allows one to calculate a priori the probability $P(A)$ of a random event "the birth of twins". To determine $P(A)$, it is necessary to use the statistics of a large number of births and to calculate how often this event occurred. Then the corresponding frequency can be approximately taken as the probability $P(A)$.

The definition of $P(A)$ in frequency is sometimes called the "*statistical definition*" of probability. This, however, is not about the definition of probability, but about its evaluation

The classical definition of the probability of an event. If the experience is that it is subdivided into only a finite number of elementary events, which are equally likely, then it is said that this is a classical case.

For experiments of this type, the theory of probabilities was developed by Laplace. Examples of such experiments are throwing a coin (two equally probable elementary events) or throwing a dice (six being equal elementary events). In the classical case, from the axioms for the probability $P(A)$ of the event A, we obtain

$$P(A) = \frac{number\ of\ elementary\ events\ favorable\ for\ A}{the\ number\ of\ all\ possible\ elementary\ events}. \quad (4)$$

In this case, under the elementary events favorable for A, such events are the implementation of which leads to the realization of A; In other words, these are the events of which A consists (understood as a subset of E). Laplace used the last formula to determine the probability.

Example 9. There are 4 blue marbles, 5 red marbles, 1 green marbles, and 2 black marbles in a bag. Suppose you select one marble at random. Find probability. *Solution.* Sample space: 12. There are 12 marbles total (4+5+1+2=12).

There are 2 black marbles in the bag 12 is your sample space:

P(black)=2/12=1/6.

There are 4 blue marbles in the bag 12 is your sample space:

P(blue)=4/12=1/3.

4 blue + 2 black = 6. We have 12 is your sample space:

P(blue or black) = 6/12=1/2.

There is 1 green, so 12 – 1 = 11 that aren`t green 12 is your sample space:

P(not green)=11/12.

I will definitely select a marble that is not purple because there are no purple marbles in the bag. Whenever the chance of something occurring is definite, the probability is 1. P(not purple)=1.

Example 10. Toss two fair coins and record the outcome. Find the probability of finding the exactly one head in the two tosses.

Solution. Sample space for the tossing two coins {HH, HT, TH, TT}. Here the letters H and T means a head or a tail respectively. Since the sum of the four simple events must be 1, each must have probability 1/4, then $P(A)$=1/4+1/4=1/2.

Example 11. The dice was thrown once is an event "an even number falls out". Favorable for A are elementary events e_2, e_4, e_6. There are six possible elementary events. Consequently, $P(A)$=3/6=1/2.

Example 12. Two dice at the same time. In this case, the winnings are paid if the sum of the dropped points ≥ 10. What is the probability of winning? There are 36 elementary events. Favorable for A are elementary events $e_{46}, e_{64}, e_{55}, e_{56}, e_{65}, e_{66}$. In this case e_{ij} means: on the first dice falls i, on the second j. Then $P(A)$=6/36=1/6.

In the classical case, combinatorial formulas are often used, for example, in order to calculate the number of all possible elementary events. In the box there are

N balls, exactly M of which are white. Let n balls be taken out from the box one by one without returning $(n \le M)$ balls. Then the probability that among these removed n balls there will be k white is

$$P_{N,M}(n,k) = \frac{C_M^k C_{N-M}^{n-k}}{C_N^n} \quad (k = 0,1,\ldots,n).$$

So, the general formula for calculation of the probability is

$$P(A) = \frac{C_m^k \cdot C_{n-m}^{m-k}}{C_n^m}.$$

Example 13. Suppose that there are 7 books. How many ways 4 books can be chosen out of 7?

Solution.

$$C_7^4 = \frac{7!}{4!(7-4)!} = \frac{7!}{4!3!} = 35.$$

Example 14. Suppose that 2 cards are drawn out from a well-shuffled deck of 36 cards. What is the probability that both of them are spades?

Solution. The number of ways is n=36; of drawing out 2 cards from a well-shuffled deck of 36 cards is C_{36}^2. Since 13 of the 36 cards are spades, the number of ways m of drawing 2 spades is C_{13}^2. Thus,

$$P(\text{getting 2 spades}) = \frac{m}{n} = \frac{C_{13}^2}{C_{36}^2} = \frac{78}{630} = 0{,}1238.$$

Example 15. Suppose that 3 people are selected at random from a group that consists of 6 men and 4 women. What is the probability that 1 man and 2 women are selected?

Solution. The number of ways of selecting 3 people from a group of 10 is C_{10}^3. One man can be selected in C_6^1 ways, and 2 women can be selected in C_4^2 ways. By the fundamental counting principle, the number of ways of selecting 1 man and 2 women is $C_6^1 \cdot C_4^2$. Thus, the probability that 1 man and 2 women are selected is

$$P = \frac{C_6^1 \cdot C_4^2}{C_{10}^3} = \frac{3}{10}.$$

Example 16. Lotto Game: guess "k numbers from n". For example, sport lotto 6 out of 49. What is the probability of getting the main prize? Indicate the number k. There is one favorable event in which there is main prize of the game. The number of all elementary events is equal to the number of possible samples of k numbers from n in order and without repeats, i.e. is equal to C_n^k. Thus, the probability of a major winnings in a sport lotto is

$$\frac{1}{C_{49}^{6}} = \frac{1}{13\,983\,816}.$$

Theorem (General Addition Rule). If A and B are two events in a sample space S, then:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

Example 17**.** Consider a pack of 52 playing cards. A card is selected at random. What is the probability that the card is either a diamond or a ten?

Solution. Let A be the event {a diamond is selected} and B be the event {a ten is selected}. The probability that it is a diamond is $P(A) = \frac{13}{52}$ since there are 13 diamond cards in the pack. The probability that the card is a ten is $P(B) = \frac{4}{52}$.

There are 16 cards that fall into the category of being either a diamond or a ten: 13 of these are diamonds and there is a ten in each of the three other suits. Therefore, the probability of the card being a diamond or a ten is $\frac{16}{52}$ not $\frac{13}{52} + \frac{4}{52} = \frac{17}{52}$. We say that these events are not mutually exclusive. We must ensure in this case not to simply add the two original probabilities; this would count the ten of diamonds twice - once in each category

We used the addition law. This example $P(A) = \frac{13}{52}$ and $P(B) = \frac{4}{52}$. The intersection event $A \cap B$ consists of only one member - the ten of diamonds - hence $P(A \cap B) = \frac{1}{52}$. Therefore

$$P(A \cup B) = \frac{13}{52} + \frac{4}{52} - \frac{1}{52} = \frac{16}{52} = 0{,}3077$$

as we have already argued.

Example 18. A card is drawn at random from a deck of 52 playing cards. What is the probability that it is an ace or a face card?

Solution. Let *F*={face card}, *A*={card is ace}. Then

$$P(F) = \frac{12}{52}; \qquad P(A) = \frac{4}{52}.$$

We have

$$P(F \cup A) = P(F) + P(A) - P(F \cap A) = \frac{12}{52} + \frac{4}{52} - 0 = \frac{16}{52} = 0{,}3077.$$

Conditional probability. The probability of some random event *A*, generally, changes if it is already known that some other random event *B* has occurred. The probability *A* provided that *B* has already occurred with probability $P(B) \neq 0$, is denoted $P(A/B)$ and is called the conditional probability *A* to condition *B*.

Example 19. If two dice were rolled at the same time. Let *A* be the phenomena "the sum of points ≥ 10", *B* the phenomena "the even sum of points". If it is known that *B* happened, then for *A* there are 18 possible elementary events (for example, e_{11} it is possible, but e_{12} not); of which are favorable for *A* are e_{46}, e_{55}, e_{66}. Consequently,

$$P(A/B) = 4/18 = 2/9.$$

Example 20. There are two boxes. In the first there are 5 white and 5 black balls, in the second - 1 white and 9 black. The experience is about taking random ball from one of the boxes. Let *B* be the event "the pulled ball is white", A_i - the event "the ball is taken out of the i-th box" (*i*=1, 2). Then

$$P(B/A_1) = \frac{5}{10} = \frac{1}{2}, \qquad P(B/A_2) = \frac{1}{10}.$$

Example 21. A manufacturer knows that the probability of an order being ready on time is 0,80 and the probability of an order being ready on time and being delivered on time is 0,72. What is the probability of an order being delivered on time, given that it is ready on time?

Solution. Let *R*: order is ready on time; *D*: order is delivered on time. *P*(*R*)=0,80; *P*(*R*,D)=0,72. Therefore:

$$P(D/R) = \frac{P(R,D)}{P(R)} = \frac{0{,}72}{0{,}80} = 0{,}90.$$

The conditional probability satisfies the following ratio checks:

$$P(A/B)=\frac{P(A\cap B)}{P(B)},\quad P(B)\neq 0, \tag{5}$$

$$P(B/A)=\frac{P(A\cap B)}{P(A)},\quad P(A)\neq 0. \tag{6}$$

Example 22. Pick a Card from a Deck. Suppose a card is drawn randomly from a deck and found to be an Ace. What is the conditional probability for this card to be Spade Ace? A = Spade Ace; B = an Ace; $A\cap B$= Spade Ace. We have

$$P(A)=1/52;\quad P(B)=4/52;\quad P(A\cap B)=1/52.$$

Hence,

$$P(A/B)=\frac{1/52}{4/52}=\frac{1}{4}.$$

If they are resolved regarding $P(A\cap B)$, then a multiplication rule is obtained:

$$P(A\cap B)=P(B)\cdot P(A/B)=P(A)\cdot P(B/A), \tag{7}$$

that is, the probability of the product of two random events is equal to the product of the probability of the event by the conditional probability of the other, provided that the first event occurred.

Example 23. Drawing a Spade Ace. Let A - an Ace; B - a Spade; $A\cap B$ - the Spade Ace

$$P(B)=13/52;\quad P(A/B)=1/13.$$

Hence,

$$P(A\cap B)=P(A/B)\cdot P(B)=\frac{1}{13}\cdot\frac{13}{52}=\frac{1}{52}.$$

Example 24. Selecting students. A statistics course has seven male and three female students. The professor wants to select two students at random to help her conduct a research project. What is the probability that the two students chosen are female?

Let A - the first student selected is female; B - the second student selected is female; $A\cap B$ - both chosen students are female

$$P(A)=3/10;\quad P(B/A)=2/9.$$

Hence

$$P(A \cap B) = P(B/A) \cdot P(A) = \frac{2}{9} \cdot \frac{3}{10} = \frac{1}{15}.$$

Example 25. A bag has 4 white cards and 5 blue cards. We draw two cards from the bag one by one without replacement. Find the probability of getting both cards white.

Solution. Let the total numbers of cards be 5 + 4 =9. Let A = event that first card is white and B = event that second card is white.

From question, $P(A) = \frac{4}{9}$. Now $P(B) = P(B/A)$, because the given events are dependent on each other. $P(B) = \frac{3}{8}$. So, $P(A \cap B) = \frac{4}{9} \cdot \frac{3}{8} = \frac{1}{6}$.

Two random events A and B are said to be *independent*, if the implementation of one does not affect the probability of implementing the other, i.e. if

$$P(A/B) = P(A), \tag{8}$$

Then the multiplication rule has the form

$$P(A \cap B) = P(A) \cdot P(B), \tag{9}$$

i.e. the probability of the product of two independent events is equal to the product of their probabilities. Since formulas (8) and (9) are equivalent, the relation (9) is often used to define the concept of "independence of two random events".

Random events $A_1, A_2, \dots, A_n$ are called *independent in aggregate* if for each m and every m-combination $i_1, i_2, \dots, i_m$, where $1 \le i_1 \le i_2 \le \dots \le i_m \le n$

$$P\left(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_m}\right) = P\left(A_{i_1}\right) \cap P\left(A_{i_2}\right) \cap \dots \cap P\left(A_{i_m}\right)$$

Random events $A_1, A_2, \dots, A_n$ are said to be *pairwise independent* if for arbitrary i, j $(i \ne j)$ A_i and A_j are independent.

Independence in the aggregate follows pairwise independence, but not vice versa.

The total probability. Suppose that a reliable event E can be represented as a sum of n pairwise incompatible events, i.e.

$$U = A_1 \cup A_2 \cup \dots \cup A_n,$$

where $A_i \cap A_j = V$ at $i \ne j$. Then for any random event B the relation

$$B = (A_1 \cap B) \cup (A_2 \cap B) \cup \ldots \cup (A_n \cap B)$$

According to the axiom of additivity, it follows that

$$P(B) = \sum_{i=1}^{n} P(B \cap A_i) .$$

If we apply formula (7), we obtain

$$P(B) = \sum_{i=1}^{n} P(A_i) \cdot P(B / A_i) . \qquad (10)$$

This formula is called the formula of *total probability*.

Example 26. Given: three boxes of one type in which are 2 white and 6 black balls. In the 2nd box are: 1 white and 8 black balls. Let's choose randomly the box and then a ball from there. Let's the event denote by B "white picked up ball", its probability $P(B)$. The chosen event "The box of 1st type" we denote by A_1, and by A_2 - "The box of 2nd type". Then

$$B = (B \cap A_1) \cup (B \cap A_2)$$

Since

$$A_1 \cap A_2 = V.$$

Consequently,

$$P(B) = P(A_1) \cdot P(B/A_1) + P(A_2) \cdot P(B/A_2)$$

But

$$P(A_1) = \frac{3}{4}, \qquad P(A_2) = \frac{1}{4}, \qquad P(B/A_1) = \frac{2}{8} = \frac{1}{4}, \qquad P(B/A_2) = \frac{1}{9}.$$

Finally

$$P(B) = \frac{3}{4} \cdot \frac{1}{4} + \frac{1}{4} \cdot \frac{1}{9} = \frac{31}{144}.$$

Example 27. Suppose that two factories supply light bulbs to the market. Factory *X*'s bulbs work for over 7000 hours in 99% of cases, whereas factory *Y*'s bu1bs work for over 7000 hours in 95% of cases. It is known that factory *X* supplies 60% of the tota1 bu1bs available. What is the chance that a purchased bulb will work for longer than 7000 hours?

Solution. Applying the law of tota1 probability, we have:

$$P(A) = P(A/B_X) \cdot P(B_X) + P(A/B_Y) \cdot P(B_Y) =$$

$$= \frac{99}{100} \cdot \frac{6}{10} + \frac{95}{100} \cdot \frac{4}{10} = \frac{594 + 380}{1000} = 0{,}974,$$

where

$P(B_X) = \frac{6}{10}$ is the probability that the purchased bulb was manufactured by factory *X*;

$P(B_Y) = \frac{4}{10}$ is the probability that the purchased bulb was manufactured by factory *Y*;

$P(A/B_X) = \frac{99}{100}$ is the probability that a bu1b manufactured by *X* will work for over 7000 hours;

$P(A/B_Y) = \frac{95}{100}$ is the probability that a bu1b manufactured by *Y* will work for over 7000 hours.

Thus, each purchased light bu1b has a 97,4% chance to work for more than 7000 hours.

Example 28. A disease called flu affects 3% of the population. There is a test to detect flu, but it is not perfect. For people with flu, the test is positive 90% of the time. For people without flu the test is positive 40% of the time. Suppose a randomly selected person takes the test and it is positive. What are the chances that a randomly selected person tests positive?

Solution. Let *A* represent a positive test result, *B* not having flu, *C* having flu, then we obtain *P*(*B*)=0,97; *P*(*C*)=0,03. The test specifications tell us: *P*(*A/B*)=0,4 and *P*(*A/C*)=0,9. We have

$$P(A) = P(A \cap B) + P(A \cap C) =$$

$$= P(B) \cdot P(A/B) + P(C) \cdot P(A/C) = 0{,}97 \cdot 0{,}4 + 0{,}03 \cdot 0{,}9 = 0{,}415.$$

Example 29. In an experiment on human memory, participants have to memorize a set of words (B_1), numbers (B_2) and pictures (B_3). These occur in the experiment with the probabilities

$$P(B_1) = 0{,}5; \quad P(B_2) = 0{,}4; \; P(B_3) = 0{,}1.$$

Then participants have recall the items (where A is the recall event). The results show that

$$P(A/B_1)=0,4;\quad P(A/B_2)=0,2;\ P(A/B_3)=0,1.$$

Compute $P(A)$, the probability of recalling an item. By the theorem of total probability:

$$\begin{aligned}P(A)&=\sum_{i=1}^{k}P(B_i)\cdot P(A/B_i)=\\&=P(B_1)\cdot P(A/B_1)+P(B_2)\cdot P(A/B_2)+P(B_3)\cdot P(A/B_3)=\\&=0,5\cdot 0,4+0,4\cdot 0,2+0,1\cdot 0,1=0,29.\end{aligned}$$

The term law of total probability is sometimes taken to mean the law of alternatives, which is a special case of the law of tota1 probability applying to discrete random variables.

Bayes formula. Suppose that the premises of the law of total probability are satisfied. Then we can calculate the probability of the event A_i provided, that event B has been occurred. For this purpose, the Bayesian formula, or the hypothesis probability formula could be used

$$P(A_i/B)=\frac{P(A_i)P(B/A_i)}{\sum_{j=1}^{n}P(A_j)P(B/A_j)}. \tag{11}$$

Example 30. The same experiment as it was described in Example 26. Let the white ball be removed. What is the possibility that it is removed from the 1st type box?

$$P(A_1/B)=\frac{P(A_1)P(B/A_1)}{P(A_1)P(B/A_1)+P(A_2)P(B/A_2)}=\frac{\frac{3}{4}\cdot\frac{1}{4}}{\frac{3}{4}\cdot\frac{1}{4}+\frac{1}{4}\cdot\frac{1}{9}}=\frac{27}{31}.$$

Example 31. While watching a game of hockey in the stadium, you observe someone who is clearly supporting Swifts in the game. What is the probability that they were actually born within 50 kilometers of Almaty?

Solution. Assume that:

- the probability that a randomly selected person in a typical local bar environment born within 50 kilometers of Almaty is 1/20, and 19/20;
- the chance that a person born within 50 kilometers of Almaty actually supports United is 7/10;
- the probability that a person not born within 50 kilometers of Almaty supports Swifts with probability 1/10 .

Define
- B - event that the person is born within 50 kilometers of Almaty,
- U - event that the person supports Swifts.

We want $P(B/U)$. By Bayes' Theorem,

$$P(B/U)=\frac{P(U/B)\cdot P(B)}{P(U)}=\frac{P(U/B)\cdot P(B)}{P(U/B)\cdot P(B)+P(U/\overline{B})\cdot P(\overline{B})}=$$

$$=\frac{\frac{7}{10}\cdot\frac{1}{20}}{\frac{7}{10}\cdot\frac{1}{20}+\frac{1}{10}\cdot\frac{19}{20}}=\frac{7}{20}\approx 0{,}269.$$

Example 32. 15% of a company's employees are mathematics and 20% are physics. 80% of the mathematics and 40% of the physics hold a managerial position, while only 20% of non-mathematics and non-physics have a similar position. What is the probability that an employee selected at random will be both a mathematic and a manager?

$$P(mathematic/manager)=\frac{0{,}15\cdot 0{,}8}{0{,}15\cdot 0{,}8+0{,}2\cdot 0{,}4+0{,}65\cdot 0{,}2}=0{,}3636.$$

Example 33. The entire output of a factory is produced on three machines. Three machines account for 20%, 30%, and 50% of the output, respectively. The fraction of defective items produced is this: for the first machine - 5%; for the second machine - 3%; for the third machine - 1%. If an item is chosen at random from the total output and is found to be defective, what is the probability that it was produced by the third machine?

Solution. Let A_i denote the event that a randomly chosen item was made by the *i-th* machine (for $i=1,2,3$). Let B denote the event that a randomly chosen item is defective. Then, we are given the following information:

$$P(A_1)=0{,}2;\quad P(A_2)=0{,}3;\quad P(A_3)=0{,}5.$$

If the item was made by machine A_1, then the probability that it is defective is 0,05; that is, $P(B/A_1)=0{,}05$. Overall, we have

$$P(B/A_1)=0{,}05; \quad P(B/A_2)=0{,}03; \quad P(B/A_3)=0{,}01.$$

To answer the original question, we first find *P*(*B*). That can be done in the following way:

$$P(B)=\sum_i P(B/A_i)\cdot P(A_i)=$$
$$=0{,}05\cdot 0{,}2+0{,}03\cdot 0{,}3+0{,}01\cdot 0{,}5=0{,}024.$$

Hence 2,4% of the total output of the factory is defective. We are given that *B* has occurred, and we want to calculate the conditional probability of A_3. By Bayes' theorem,

$$P(A_3/B)=\frac{P(B/A_3)\cdot P(A_3)}{P(B)}=\frac{0{,}01\cdot 0{,}50}{0{,}024}=\frac{5}{24}.$$

Given that the item is defective, the probability that it was made by the third machine is only 5/24. Although machine 3 produces half of the total output, it produces a much smaller fraction of the defective items.

Hence the knowledge that the item selected was defective enables us to replace the prior probability $P(A_3)=1/2$ by the smaller posterior probability $P(A_3/B)=5/24.$

Once again, the answer can be reached without recourse to the formula by applying the conditions to any hypothetical number of cases. For example, in 100,000 items produced by the factory: 20,000 will be produced by Machine *A*; 30,000 by Machine *B*; and 50,000 by Machine *C*. Machine *A* will produce 1000 defective items, Machine *B* - 900 and Machine *C* - 500. Of the total 2400 defective items, only 500, or 5/24 were produced by Machine *C*.

The Bernulli formula.

$$P_n(m)=C_n^m p^m q^{n-m}.$$

Here, C_n^m denotes the number of combinations of *n* elements taken *m* at a time. For large *n*, the calculation using this formula becomes difficult.

We can use the Bernulli's formula for independent events if probability of events occurring are

1) Less then *m* times

$$P(< m) = P_n(0) + P_n(1) + ... + P_n(m-1).$$

2) More than *m* times

$$P(> m) = P_n(m+1) + P_n(m+2) + ... + P_n(n).$$

3) No more then *m* times

$$P(\leq m) = P_n(0) + P_n(1) + ... + P_n(m).$$

4) At least *m* times

$$P(\geq m) = P_n(m) + P_n(m+1) + ... + P_n(n).$$

5) At least once

$$\text{Đ}_n(m \neq 0) = 1 - \text{Đ}_n(0).$$

Example 34. Calculate the probability of rolling 4 on a dice exactly 5 times in 25 trials.

Solution. *W*e have the following:

n = total trials = 25;

k = total successes = 5;

$n–k$ = total failures = 20;

p = 1/6 = 0,167; q = 5/6 = 0,833

$$C_{25}^{5} = \frac{25!}{5!\,20!} = 53\,130.$$

Therefore, probability will be:

$$P(25{,}5) = C_{25}^{5} \cdot p^5 \cdot q^{20} = 53130 \cdot 0{,}167^5 \cdot 0{,}833^{20} = 0{,}17844.$$

Thus, the probability is 0,17844. This way, we can calculate the probability of any event provided we know the number of trials and the probability of the event occurring in a single trial.

Example 35. A student is writing an exam of multiple choice questions. It contains a total of 15 questions, each of which has 4 possible answers. What is the probability that student gets exactly 11 correct answer?

Solution. Given that:

n = 15;

p = Probability of success = 1/4 = 0,25;

q = Probability of failure = 1 – 0,25 = 0,75;

k = 11.

The probability of exactly 11 correct answers:

$$P(15{,}11) = C_{15}^{11} \cdot 0{,}25^{11} \cdot 0{,}75^{4} = 0{,}0003089.$$

Example 36. What is the probability of getting heads exactly 3 times if you flip a fair coin 6 times?

We have k=3 number of successes; n=6 number of trials; p=0,5 probabi1ity of success; $q = 1 - p = 1 - 0{,}5 = 0{,}5$ probabi1ity of failure.

Probability of getting 3 heads:

$$P(X = 3) = C_6^3 \cdot 0{,}5^3 \cdot 0{,}5^{6-3} = 20 \cdot 0{,}125 \cdot 0{,}125 = 0{,}3125.$$

Example 37. The probability of a boy's birth is 0,515. How great is the probability that among 10 randomly chosen newborns there will be 6 boys?

The assumption of independence can be considered as fulfilled. Thus, for the desired probability we have

$$P(A) = P(10{,}6) = C_{10}^{6}(0{,}515)^{6}(0{,}485)^{4} \approx 0{,}2167.$$

Example 38. You are taking a 10 questions multiple choice test. If each question has four choices and you guess on each question, what is the probability of getting exactly 7 questions correct?

Solution.

n = 10;

k = 7;

n – k= 3;

p= 0,25 - probability of guessing the correct answer on a question;

q = 0,75 - probability of guessing the wrong answer on a question.

$$P(7\ correct\ guesses\ out\ of\ \ 10\ \ questions) =$$
$$= C_{10}^{7} \cdot 0{,}25^{7} \cdot 0{,}75^{3} = 0{,}0031.$$

Example 39. A student takes a multiple choice quiz with 4 possible answers to each of the 10 questions. If he guesses randomly, find the:

(a) probability he scores 7 out of 10;

(b) probability he scores 8 or better;

(c) probability he fails (6 or less).

Solution.

(a) probability he scores 7 out of 10:

$$P(10{,}7) = C_{10}^{7} \cdot 0{,}25^{7} \cdot 0{,}75^{3} = 0{,}003;$$

(b) probability he scores 8 or better:

$$P(\geq 8) = P(10) + P(9) + P(8) =$$
$$= C_{10}^{10} \cdot 0{,}25^{10} \cdot 0{,}75^{0} + C_{10}^{9} \cdot 0{,}25^{9} \cdot 0{,}75^{1} + C_{10}^{8} \cdot 0{,}25^{8} \cdot 0{,}75^{2} = 0{,}00042.;$$

(c) probability he fails (6 or less):

$$P(\leq 6) = 1 - [P(10{,}7) + P(\geq 8)] =$$
$$= 1 - [C_{10}^{7} \cdot 0{,}25^{7} \cdot 0{,}75^{3} + 0{,}00042] = 1 - 0{,}00342 = 0{,}9966.$$

RANDOM VARIABLE

The real variable, which, depending on the outcome of the experiment, i.e. depending on the case, takes different values, is called a *random variable*.

Let X be a random variable. *A distribution function* $F(x)$ of a random variable X is called a function

$$F(x) = P(X < x). \tag{12}$$

The value of the distribution function at a point x_0 is equal to the probability that a random variable takes a value less than x_0. In probability theory, a random variable is completely characterized by its distribution function, i.e. can be considered as given if its distribution function is given. Using the distribution function, you can specify the probability that a random variable falls into a given half-open interval

$$P(a \le X < b) = F(b) - F(a). \tag{13}$$

The distribution function $F(x)$ of an arbitrary random variables has the following properties:

1) $\lim\limits_{x\to+\infty} F(x) = 1, \quad \lim\limits_{x\to-\infty} F(x) = 0.$

2) $F(x)$ monotonically does not decrease, i.e. in cases where $x_1 < x_2$, takes place equality $F(x_1) \le F(x_2)$.

3) $F(x)$ is continuous from the left.

Discrete random variables. A random variable X is called *discrete* if it can take only a finite or countable set of values. Thus, it is characterized by the values $x_1, x_2, \ldots$ it can take, and the probabilities $p_i = P(X = x_i)$ with which it takes these values and which must satisfy the condition $\sum_i p_i = 1$.

A one-to one correspondence of sets x_i onto a set p_i is considered as a *function of the probability* of a discrete random variable. For a distribution function of a discrete random variable we have

$$F(x) = \sum_{x_i < x} p_i \tag{14}$$

The graphic representation of the series of the distribution of a discrete random variable is called the *polygon* of the distribution. The summation is implied over all ***i*** for which $x_i < x$. Thus, $F(x)$ is a step function with jumps in height p_i at the points x_i (Fig. 1).

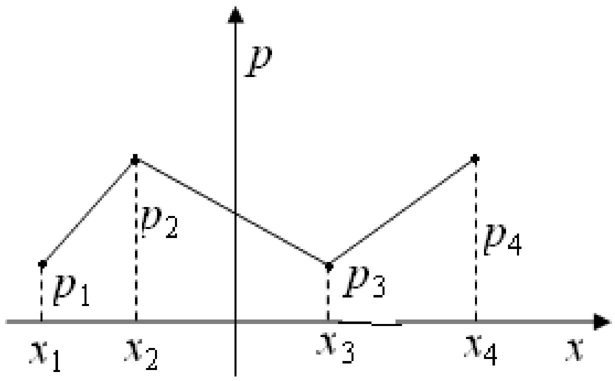

Figure 1.

Example 40. A discrete random variable X is set next distribution

X	0	10	20	30	40	50
P	0,05	0,15	0,3	0,25	0,2	0,05

Find: a) distribution function $F(x)$; b) the math expectation, dispersion.

Solution: (a) for a discrete random variable distribution function $F(x)$ to be for all values of i, that $x_i < x$ according to the formula:

$$F(x) = P(X < x) = \sum_{x_i < x} p_i = \sum_{x_i = x} P(X = x_i);$$

If $x \le 0$, then $F(x) = P(X < 0) = 0;$

If $0 < x \le 10$, then $F(x) = P(X = 0) = 0{,}05;$

If $10 < x \le 20$, then $F(x) = P(X = 0) + P(X = 10) = 0{,}05 + 0{,}15 = 0{,}2;$

If $20 < x \le 30$, then $F(x) = P(X = 0) + P(X = 10) + P(X = 20) =$
$= 0{,}2 + 0.3 = 0{,}5;$

If $30 < x \le 40$, then $F(x) = P(X = 0) + P(X = 10) + P(X = 20) + P(X = 30) =$
$= 0{,}5 + 0{,}25 = 0{,}75;$

If $40 < x \le 50$, then $F(x) = P(X = 0) + P(X = 10) + P(X = 20) + P(X = 30) +$
$+ P(X = 40) = 0{,}75 + 0{,}2 = 0{,}95;$

If $x < 50$, then $F(x) = P(X = 0) + P(X = 10) + P(X = 20) + P(X = 30) +$
$+ P(X = 40) + P(X = 50) = 0{,}95 + 0{,}05 = 1;$

In this way,

$$F(x) = \begin{cases} 0, & x \le 0 \\ 0.05, & 0 < x \le 10 \\ 0.2, & 10 < x \le 20 \\ 0.5, & 20 < x \le 30 \\ 0.75, & 30 < x \le 40 \\ 0.95, & 40 < x \le 50 \\ 1, & x > 50 \end{cases}$$

(b) we find the math expectation and the dispersion. The math expectation for discrete random variable is:

$$M(X) = \sum_i x_i p_i =$$

$$= 0 \cdot 0{,}15 + 10 \cdot 0{,}15 + 20 \cdot 0{,}3 + 30 \cdot 0{,}25 + 40 \cdot 0{,}2 + 50 \cdot 0{,}05 = 25{,}5.$$

The dispersion of the random variable *X* is obtained by the formula

$$D(X) = M(X^2) - [M(X)]^2.$$

We have

$$D(X) = 0^2 \cdot 0{,}05 + 10^2 \cdot 0{,}15 + 20^2 \cdot 0{,}3 + 30^2 \cdot 0{,}25 +$$
$$+ 40^2 \cdot 0{,}2 + 50^2 \cdot 0{,}05 = 805.$$

Example 41. Find the distribution of a discrete random variable *X*, which has only two possible values x_1 and x_2, and $x_1 < x_2$, knowing the expectation *M*(*X*)=0,24, the probability $p_1 = 0{,}6$ of possible value x_1.

Solution: The sum of the probabilities of all possible values of a discrete random variable is equal to one, so the probability that *X* takes the value x_2 is equal to 1 - 0,6 = 0,4. Then the law of the distribution of *X* is:

X	x_1	x_2
P	0,6	0,4

To find x_1 and x_2 be two equations, using the known values and formulas expectation and variance. To do this, write the law of distribution of X^2.

X^2	x_1^2	x_2^2
P	0,6	0,4

Thus,

$$M(X) = 0{,}6 \cdot x_1 + 0{,}4 \cdot x_2$$

and

$$D(X) = 0{,}6 \cdot x_1^2 + 0{,}4 \cdot x_2^2$$

Hence, we obtain the system:

$$\begin{cases} 0{,}6x_1 + 0{,}4x_2 = 1{,}4 \\ 0{,}6x_1^2 + 0{,}4x_2^2 - 1{,}4^2 = 0{,}24 \end{cases}.$$

Solving this system of equations, find two solutions:

$$x_1 = 1,\ x_2 = 2 \quad \text{and} \quad x_1 = 1{,}8\ \ x_2 = 0{,}8.$$

According to the problem $x_1 < x_2$, so the problem satisfies only the first solution $x_1 = 1,\ x_2 = 2$. Seeking the law of distribution of a discrete random variable X is:

X	1	2
P	0,6	0,4

*The event indic*ator. If A is some random event, where $P(A) = p$. Random value

$$X = \begin{cases} 1, & if\ \ \grave{A}\ occurs, \\ 0, & if\ \ \grave{A}\ does\ not\ occur \end{cases}$$

is called the indicator A (the random variable characteristic). The possible values are 0 and 1, the relevant probabilities

$$p_0 = P(X = 0) = 1 - p, \quad p_1 = P(X = 1) = p.$$

Binomial distribution. Let some experiment be repeated n times and individual experiments of this series do not depend on each other. Let event A occur or not occur in each experiment, and the probability of the realization in a separate experiment does not depend on the number of the experiment and is equal to p. Let $X^{(n)}$- be the number of occurrences of event A in such a series of n experiments. Obviously, the possible values $X^{(n)}$ of a random variable are numbers 0,1,2,...,n. Probabilities

$$P(n,k) = P(X^{(n)} = k)$$

are calculated by binomial law

$$P(n,k) = C_n^k p^k q^{n-k}, \quad q = 1 - p \quad (k = 0,1,...,n). \tag{15}$$

A random variable is called *binomially distributed* with parameters n and p, if possible values 0,1,2,...,n it takes probabilities $P(n,k)$ with the given formulas (15). The parameters n and p completely determine the binomial distribution.

Figure 2 shows the "polygons" of binomial distributions for n=20 and various p. The corresponding $P(n,k)$ are plotted along the ordinate and connected by a broken line. Since from $P(n,0)$, $P(n,k)$ can be easily calculated from the following recurrence formula:

$$\frac{P(n,k)}{P(n,k-1)} = \frac{(n-k+1)p}{kq}. \tag{16}$$

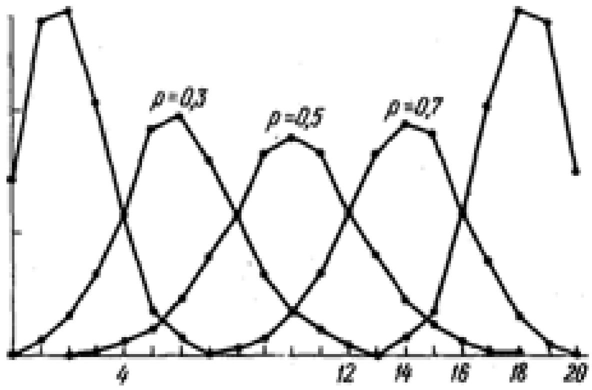

Figure 2.

Example 42. The probability of a boy's birth is equal to 0,515. How great is the probability that out of 10 randomly chosen newborns there will be 6 boys? The assumption of independence can be considered as fulfilled. Thus, for the required probability we have

$$P(A) = P(10,6) = C_{10}^{6}(0{,}515)^{6}(0{,}485)^{4} \approx 0{,}2167.$$

Example 43. There are *N* balls in the box, where *M* are white. From the box, the ball was taken out *n* times and after registering returned to the box again. What is the probability of the event that the white ball was registered *k* times? This probability $P(n,k)$ is calculated by using the binomial law (15):

$$P(n,k) = C_n^k \left(\frac{M}{N}\right)^k \left(1 - \frac{M}{n}\right)^{n-k} \qquad (k = 0,1,\dots,n) \tag{17}$$

The binomial law describes in the most general form the implementation of *runtime in a sample of* n *return*.

Example 44. If the probability of obtaining a defective product is 0,01. What is the probability that among a hundred products there will be not more than three defective ones? According to the binomial law and the law of addition, we get that

$$P(A) = C_{100}^{0}(0{,}01)^{0}(0{,}99)^{100} + C_{100}^{1}(0{,}01)^{1}(0{,}99)^{99} + C_{100}^{2}(0{,}01)^{2}(0{,}99)^{98} + $$
$$+ C_{100}^{3}(0{,}01)^{3}(0{,}99)^{97} = 0{,}9816.$$

Example 45. Your name is Asan and you are an expert penalty goal shooter. Your skill has been improved for the past 10 years, and now you are as good as you will ever be. Your success rate has been measured at 80%. Thus, *p*=0,8 and *q*=0,2. You take *n* = 6 shots on goal, so the possible values of *X* (the number of successes) are 0,1,2,3,4,5,6. Here is the probability for each value of *X*:

$$P(X = 0) = C_6^0 \cdot 0{,}8^0 \cdot 0{,}2^6 = 1 \cdot 1 \cdot 0{,}000064 = 0{,}000064;$$

$$P(X = 1) = C_6^1 \cdot 0{,}8^1 \cdot 0{,}2^5 = 6 \cdot 0{,}8 \cdot 0{,}00032 = 0{,}001536;$$

$$P(X = 2) = C_6^2 \cdot 0{,}8^2 \cdot 0{,}2^4 = 15 \cdot 0{,}64 \cdot 0{,}0016 = 0{,}01536;$$

$$P(X = 3) = C_6^3 \cdot 0{,}8^3 \cdot 0{,}2^3 = 20 \cdot 0{,}512 \cdot 0{,}008 = 0{,}08192;$$

$$P(X = 4) = C_6^4 \cdot 0{,}8^4 \cdot 0{,}2^2 = 15 \cdot 0{,}4096 \cdot 0{,}04 = 0{,}24576;$$

$$P(X = 5) = C_6^5 \cdot 0{,}8^5 \cdot 0{,}2^1 = 6 \cdot 0{,}32678 \cdot 0{,}2 = 0{,}393216;$$

$$P(X=6)=C_6^6\cdot 0{,}8^6\cdot 0{,}2^0=6\cdot 0{,}262144\cdot 1=0{,}262144.$$

Putting them all together gives the probability distribution for *X*:

x	0	1	2	3	4	5	6
$P(X=x)$	0,000064	0,001536	0,01536	0,08192	0,24576	0,393216	0,262144

We can use the probability distribution to find the probability that *X* is in a given range by adding the individual probabilities.

Probability of at least 5 successes	$P(5\le X\le 6)=P(5)+P(6)=$ $=0{,}393216+0{,}262144=0{,}65539$ or a 65,5% chance.
Probability of at most 2 successes	$P(0\le X\le 2)=P(0)+P(1)+P(2)=$ $=0{,}000064+0{,}001536+0{,}01536=0{,}01696$ or a 1,7% chance.
Probability of at least 3 successes	$P(3\le X\le 6)=1-(P(0)+P(1)+P(2))=$ $=1-0{,}01696=0{,}98304$ i.e. 1-Probability of at most 2 successes.
Since "at most 2 successes" and "at least 3 successes" are complementary events	=1-0,01696=0,98304.

We used the previous answer, and that was a easier than adding

$$P(X=3),\quad P(X=4),\quad P(X=5)\quad \text{and}\quad P(X=6).$$

Here is the probability distribution again:

x	0	1	2	3	4	5	6
$P(X=x)$	0,000064	0,001536	0,01536	0,08192	0,24576	0,393216	0,262144

Hypergeometric distribution. In the box there are *N* balls, where *M* are white. Let's take $n\ (\le M)$ balls one by one without returning (or simultaneously, which is the same thing). Then the probability that among these n balls removed there will be k white ones is

$$P_{N,M}(n,k) = \frac{C_M^k C_{N-M}^{n-k}}{C_N^n} \quad (k = 0,1,...,n). \tag{18}$$

A random variable is called *hypergeometric distribution*, if possible values 0,1,...,n it takes with probabilities $P_{N,m}(n,k)$ defined by the formula (18). The numbers *N*, *M*, *n* are the distribution parameters.

The hypergeometric distribution describes the implementation of the characteristic in the sample without a return. If *N* is very large in comparison with *n*, then it does not matter whether the balls return back or not, and formula (18) can be approximately replaced by formula (17) of the binomial distribution.

Poisson distribution. A random variable is said to be Poisson distributed if it takes a countable set of possible values 0, 1, 2, ... with probabilities

$$P(k) = \frac{\lambda^k}{k!} e^{-\lambda} \quad (k = 0,1,...). \tag{19}$$

The value λ is the *parameter of distribution*.

The Poisson distribution can be used as a good approximation of the binomial distribution if n is large and p is small. Then, in quality λ need to take np, that is $\lambda = np$.

Example 46. The car traveled 100 000 km. Let *X* be the number of punctures of the tire at this distance. Then *X* can be regarded as a random variable distributed according to Poisson's law (with a suitable λ), that is, the probability of three punctures of the bus is

$$P(A) = \frac{\lambda^3}{3!} e^{-\lambda}.$$

Example 47. Let's consider Example 44. We have *n*=100, *p*=0,01. Thus

$$\lambda = np = 100 \cdot 0{,}01 = 1,$$

$$P(A) = \frac{1^0}{0!}e^{-1} + \frac{1^1}{1!}e^{-1} + \frac{1^2}{2!}e^{-1} + \frac{1^3}{3!}e^{-1} = \frac{1}{e}\left(1 + 1 + \frac{1}{2} + \frac{1}{6}\right) = 0{,}9810,$$

which gives a good match with the exact value, but it is calculated much faster. The Poisson distribution is tabulated for various λ (Application 1).

Example 48. Let the probability of obtaining a defective product is 0,01. What is the probability that among a hundred products there will be no more than three defective ones?

We have n=100, p=0,01. Thus $\lambda = np = 100 \cdot 0{,}01 = 1$.

$$P(A) = \frac{1^0}{0!}e^{-1} + \frac{1^1}{1!}e^{-1} + \frac{1^2}{2!}e^{-1} + \frac{1^3}{3!}e^{-1} =$$
$$= \frac{1}{e}\left(1 + 1 + \frac{1}{2} + \frac{1}{6}\right) = 0{,}9810,$$

which gives a good match with the exact value, but it is calculated much faster.

Example 49. On a particular river, overflow floods occur once every 100 years on average. Calculate the probability of k=0,1,2,3,4,5 or 6 overflow floods in a 100-year interval, assuming the Poisson model is appropriate.

As the average event rate is one overflow flood per 100 years, $\lambda = 1$:

$$P(k\ overflow\ \ floods\ \ in\ 100\ \ years) = \frac{\lambda^k e^{-\lambda}}{k!} = \frac{1^k e^{-\lambda}}{k!};$$

$$P(k = 0\ overflow\ \ floods\ in\ 100\ \ years) = \frac{1^k e^{-1}}{0!} = \frac{e^{-1}}{1} = 0{,}368;$$

$$P(k = 1\ overflow\ \ floods\ in\ 100\ \ years) = \frac{1^k e^{-1}}{1!} = \frac{e^{-1}}{1} = 0{,}368;$$

$$P(k = 2\ overflow\ \ floods\ in\ 100\ \ years) = \frac{1^2 e^{-1}}{2!} = \frac{e^{-1}}{2} = 0{,}184.$$

The table below gives the probability for 0 to 6 overflow floods in a 100 year period.

k	0	1	2	3	4	5	6
$P(k)$	0,368	0,368	0,184	0,061	0,015	0,003	0,0005

Continuous random variables. A random variable is called *continuous* if its distribution function (*the integral distribution function*) can be represented in the form

$$F(x) = \int_{-\infty}^{x} f(t)dt \quad . \qquad (20)$$

The function $f(x)$ is called *the distribution density*. If

$$\lim_{x \to +\infty} F(x) = 1 \,,$$

then the condition must be fulfilled

$$\int_{-\infty}^{+\infty} f(x)dx = 1 \quad . \qquad (21)$$

With a given probability density, due to the fact that

$$P(a \le X < b) = F(b) - F(a)$$

and (20), the probability that a random variable falls within a given interval is equal to (Fig. 3)

$$P(a \le X < b) = F(b) - F(a) = \int_{a}^{b} f(x)dx \quad .$$

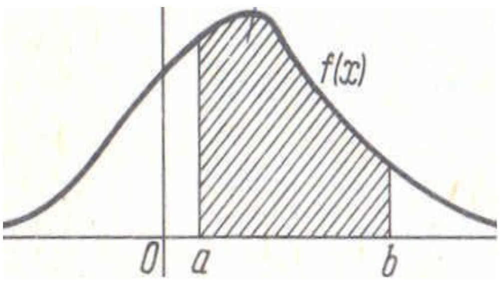

Figure 3.

The probability $P(X=a)$, i.e. the probability that a continuous random variable is equal to a given real number, is always equal to 0. Note that it does not follow from the equality $P(A)=0$ that A is an impossible event, although $P(V)=0$.

Uniform distribution. A random variable is called *uniformly distributed* on $[a,b]$ if its probability density on the $[a,b]$ is constant, and outside $[a,b]$ is equal to 0 (Fig. 4).

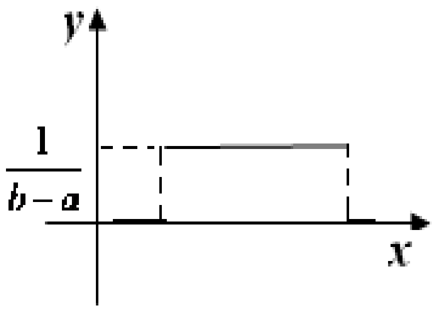

Figure 4.

Since

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

then

$$f(x) = \frac{1}{b-a}.$$

Uniform distribution on [*a,b*]:

$$M(X) = \int_a^b x\frac{1}{b-a}dx = \frac{b+a}{2},$$

$$D(X) = \int_a^b \left(x - \left(\frac{a+b}{2} \right) \right)^2 \frac{1}{b-a}dx = \frac{(b-a)^2}{12}.$$

Normal distribution (Gaussian distribution). A random variable is called *normal distribution* if it has following probability density

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(x-a)^2/2\sigma^2}, \qquad (23)$$

where a and σ are *the parameters of distribution*.

Function (23) is a bell-shaped curve. The parameter a is the maximum point through which the symmetry axis passes, the parameter σ is the distance from this axis to the inflection point.

If σ is small, the curve is high and pointed; If σ is large, it is wide and flat. Figure 5 shows the normal distribution for $a = 0$ and different σ.

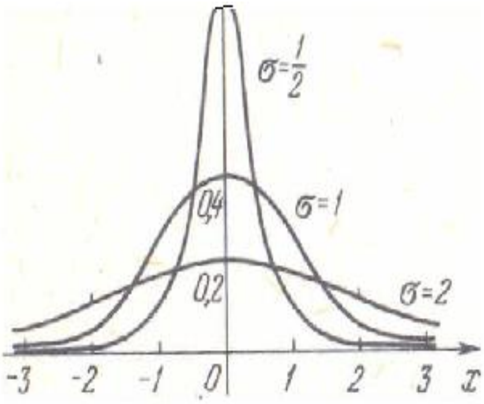

Figure 5.

If the random variable X has a normal distribution with parameters a and σ, then we say that X distributed normally according to law $N(x, a, \sigma)$, write as $X \in N(x, a, \sigma)$. Function

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2}, \quad (a = 0, \ \sigma = 1)$$

is called *the density of a normalized and centered normal distribution*. The probability density $\varphi(x)$ and the corresponding distribution

$$\Phi(x) = \int_{-\infty}^{x} \varphi(t) dt$$

tabulated (Application 2). The function $\Phi(x)$ is often called *the Gaussian error integral*

$$\Phi_0(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-t^2/2}\, dt,$$

$$\Phi(x) = \Phi_0(x) + \frac{1}{2}.$$

The special significance of the normal distribution in theory and practice is based to a large extent on the central limit theorem.

Example 50. Find the probability of getting a given interval (12,14) normally distributed random variable, if you know the expectation of a = 10 and standard deviation $\sigma = 2$.

Solution: we use the formula for the solution:

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

Substituting $\alpha = 12, \quad \beta = 14, \quad a = 10, \quad \sigma = 2$ obtain:

$$P(12<X<14) = \Phi(2) - \Phi(1).$$

From the Application 2 find the values of the functions

$$\Phi(2) = 0{,}4772 \quad \text{and} \quad \Phi(1) = 0{,}3413,$$

then the desired probability $P(12<X<14) = 0{,}1359$.

Exponential distribution. A random variable is called exponentially distributed if it has the following probability density

$$f(x) = \begin{cases} \lambda e^{\lambda x} & at \;\; x \geq 0, \\ 0 & at \;\; x < 0 \end{cases}$$

where λ the distribution parameter.

Example 51. The service life of the light bulb can be viewed with good approximation as an exponentially distributed value. Fig. 6 shows the probability density of the exponential distribution with $\lambda = 1$.

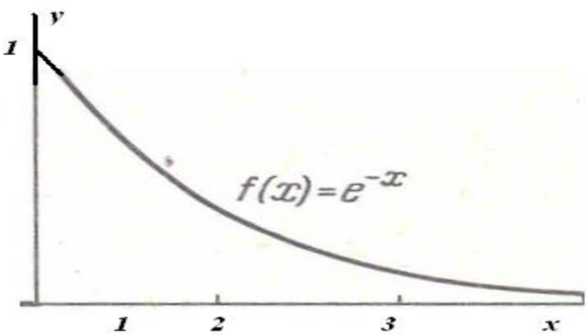

Figure 6.

MOMENTS OF DISTRIBUTION

Discrete time case. Let *X* be a discrete random variable with possible values of $x_1, x_2, \ldots$ and $p_k = P(X = x_k)$.

The number

$$v_i = \sum_i x_k^i p_k$$

in the case of absolute convergence of the series is called *the i-th initial moment* of the random variable *X* (or its distribution) (*i*=1,2,...).

The number

$$\mu_i = \sum_k (x_k - v_1)^i p_k$$

is called the *i-th central moment* of *X*.

Of particular importance are the first initial moment v_1 and the second central moment μ_2 .

Mathematical expectation. The first initial moment

$$v_1 = \sum_i x_k p_k$$

is called *the mathematical expectation* of *X* and is denoted by *M*(*X*).

The expectation determines the position of the distribution center in the following sense: if we assume p_k as the masses placed at the points x_k of the real axis, then *M*(*X*) is the coordinate of the center of gravity of this system.

Properties of mathematical expectation.

1) The mathematical expectation of the constant *C* (which can be regarded as a discrete random variable with one possible value a, which it takes with probability 1) is equal to this constant:

$$M(C)=C, \quad C \text{ - const.}$$

2) The mathematical expectation of a sum is equal to the sum of mathematical expectations:

$$M(X_1 + X_2) = M(X_1) + M(X_2) . \tag{24}$$

3) The mathematical expectation of a product of constant value by a random variable is equal to the product of the constant by the mathematical expectation of a random variable:

$$M(aX) = aM(X).$$

4) The mathematical expectation of the product of two independent random variables is equal to the product of their mathematical expectations

$$M(X_1 \cdot X_2) = M(X_1) \cdot M(X_2). \tag{25}$$

A binomial distribution with parameters *n*, *p*:

$$M(X) = \sum_{k=0}^{n} k C_n^k p^k q^{n-k} = np. \tag{26}$$

Hypergeometric distribution with parameters *N*, *M*, *n*:

$$M(X) = \sum_{k=0}^{n} k \frac{C_M^N C_{N-M}^{n-k}}{C_N^n} = n\frac{M}{N}. \tag{27}$$

Poisson distribution with parameter λ:

$$M(X) = \sum_{k=0}^{\infty} k \frac{\lambda^{\ddot{e}}}{k!} = \lambda. \tag{28}$$

Thus, the parameter λ here has a meaning of mathematical expectation.

Dispersion. The second central moment is called the *dispersion* of the random variable *X* and is denoted by *D*(*X*), that is,

$$D(X) = \sum_{k} [x_k - M(X)]^2 p_k = M(X - M(X))^2. \tag{29}$$

To calculate the dispersion, the following formula is often useful:

$$D(X) = M(X^2) - [M(X)]^2 \tag{30}$$

The square root of the dispersion is called *the spread*, or the standard deviation, or *the mean square deviation*, and is denoted by σ_X:

$$\sigma_X = \sqrt{D(X)}\,. \tag{31}$$

The value σ (or $D(X)$) is the measure of spread-out distribution with respect to the mathematical expectation.

Dispersion properties.

1) The dispersion of a constant is zero: $D(C) = 0$.

2) The dispersion of a product of a constant value by a random variable is equal to the product of the square of a constant value by the dispersion of the random variable:

$$D(CX) = C^2 D(X)$$

3) The dispersion of the sum of the constant C and the random variable is equal to the dispersion of the random variable:

$$D(C + X) = D(X)\,.$$

4) The dispersion of the sum of two independent random dispersions is equal to the sum of the dispersions of these quantities:

$$D(X + Y) = D(X) + D(Y)$$

Binomial distribution

$$D(X) = \sum_{k=0n}^{n} (k - np)^2 C_n^k p^k q^{n-k} = npq \tag{32}$$

$$\sigma_X = \sqrt{npq}. \tag{33}$$

Hypergeometric distribution:

$$D(X) = \frac{N-n}{N-1} n \frac{M}{N}\left(1 - \frac{M}{N}\right) \tag{34}$$

Poisson distribution:

$$D(X) = \lambda, \quad \sigma_X = \sqrt{\lambda} \tag{35}$$

Continuous case. Let X be a continuous random variable with probability density $f(x)$. Then

$$\nu_i = \int_{-\infty}^{+\infty} x^i f(x)dx \tag{36}$$

is called, in the case of absolute convergence of the integral, *the i-th initial moment* of the random variable X (i=1, 2, ...).

$$\mu_i = \int_{-\infty}^{+\infty} (x-\nu_1)^i f(x)dx$$

is called *the i-th central moment* of the random variable X.

The mathematical expectation. First initial moment

$$\nu_1 = \int_{-\infty}^{+\infty} xf(x)dx$$

is called *the mathematical expectation* of the random variable X.

The mathematical expectation $M(X)$ gives central position of gravity of the mass distribution, which is defined by the "mass distribution density".

The mathematical expectation in the continuous case has the same properties 1)-4), which were noted for the discrete case.

Uniform distribution on $[a,b]$:

$$M(X) = \int_a^b x\frac{1}{b-a}dx = \frac{b+a}{2}.$$

Normal distribution on $N(x,a,\sigma)$:

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

Consequently, the parameter a has the meaning of a mathematical expectation.

Exponential distribution:

$$M(X)=\int_0^{\infty} x\lambda e^{-\lambda x}dx=\frac{1}{\lambda}.$$

Dispersion. The second central moment

$$\mu_2 = D(X)=\int_{-\infty}^{+\infty}(x-M(X))^2 f(x)dx \tag{38}$$

is called the dispersion of the random variable *X*. The value $\sqrt{D(X)}=\sigma$ is called *the spread*, or the *standard deviation*, or the *mean square deviation* of the random variable *X*. The following formula

$$D(X)=M(X-M(X))^2=M(X^2)-[M(X)]^2. \tag{39}$$

The dispersion in the continuous case has the same properties 1)-4), which were noted for the discrete case.

Uniform distribution on [*a*,*b*]:

$$D(X)=\int_a^b\left(x-\left(\frac{a+b}{2}\right)\right)^2\frac{1}{b-a}dx=\frac{(b-a)^2}{12}.$$

Normal distribution on $N(x,a,\sigma)$:

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

Thus, the normal distribution is completely determined by specifying the mathematical expectation and the standard deviation.

Example 33. Exponential distribution:

$$D(X)=\frac{1}{\lambda^2}.$$

RANDOM VECTORS
(MULTIDIMENSIONAL RANDOM VALUES)

A combination $(X_1, X_2, \dots, X_n)$ of random variables is called *n-dimensional random vector*. Such a random vector can be characterized by its *n-dimensional distribution function*:

$$F(x_1, x_2, \dots, x_n) = P(X_1 < x_1, X_2 < x_2, \dots, X_n < x_n). \qquad (40)$$

A function $F(x_1, x_2, \dots, x_n)$ is often also called a vector distribution $(X_1, X_2, \dots, X_n)$ or a joint distribution of variables $X_1, X_2, \dots, X_n$. If we consider the variables $X_1, X_2, \dots, X_n$ as the coordinates of a point in n-dimensional Euclidean space, then the position of the point $(X_1, X_2, \dots, X_n)$ depends on the case, and the value of the function $F(x_1, x_2, \dots, x_n)$ is the possibility that the point is in a half-open parallelepiped

$$X_1 < x_1, X_2 < x_2, \dots, X_n < x_n$$

with edges parallel to the axes. The probability that point turns out in a parallelepiped $a_i \le X_i < b_i \quad (i = 1, 2, \dots, n)$ is determined by formula

$$P(a_1 \le X_1 < b_1, \dots, a_n \le X_n < b_n) =$$
$$= F(b_1, b_2, \dots, b_n) - \sum_{i=1}^{n} p_i + \sum_{1 \le i < j \le n} p_{ij} - \dots + (-1)^n F(a_1, \dots, a_n). \qquad (41)$$

Here

$$p_{i_1, i_2, \dots, i_k} = F(c_1, \dots, c_n) \quad (1 \le i_1 < i_2 < \dots < i_k \le n),$$

where $c_{i_1} = a_{i_1}, \dots, c_{i_k} = a_{i_k}$, and all the rest $c_j = b_j$.

For a two-dimensional random vector in particular we get

$$P(a_1 \le X_1 < b_1, a_2 \le X_2 < b_2) =$$
$$= F(b_1, b_2) - F(b_1, a_2) - F(a_1, b_2) + F(a_1, a_2).$$

Properties of the n-dimensional distribution function.

1)

$$\lim_{\substack{x_j \to +\infty \\ j=1,\dots n}} F(x_1,\dots,x_n)=1 \quad , \qquad \lim_{\substack{x_j \to -\infty \\ j=1,\dots n}} F(x_1,\dots,x_n)=0.$$

2) $F(x_1, x_2, \dots, x_n)$ *m*onotonically does not decrease in each variable.

3) $F(x_1, x_2, \dots, x_n)$ continuous on the left to each variable.

4) For arbitrary

$$a_i, b_i, \quad a_i < b_i \ \ (i=1,\dots,n),$$

the right-hand side of (41) is nonnegative.

Discrete random vectors. A random vector $(X_1, X_2, \dots, X_n)$ is called *discrete* if all its components are discrete random variables. If $x_k^{(j)}$ $(k = 1, 2, \dots)$ are the possible values of the j-th component, then the probability function $p_{i_1,i_2,\dots,i_n}$ the vector $(X_1, X_2, \dots, X_n)$ is defined as follows:

$$p_{i_1\dots i_n} = P(X_1 = x_{i_1}^{(1)}, X_2 = x_{i_2}^{(2)}, \dots, X_n = x_{i_n}^{(n)}).$$

The distribution function has a fair ratio

$$F(x_1, x_2, \dots, x_n) = \sum_{\substack{x_{i_k}^{(k)} < x_k \\ k=1,2,\dots,n}} p_{i_1\dots i_n}. \qquad (42)$$

Summation is made for all i_k, for which $x_{i_k}^{(k)} < x_k$, etc.

Polynomial distribution. Let in some experiment there always be one of the pairs of incompatible events $A_1, A_2, \dots, A_k$. Let $P_i = P(A_i)$. As $A_1 \cup A_2 \cup \dots \cup A_k = U$ and $A_i \cap A_j = V$ in $i \neq j$, then

$$\sum_{i=1}^{k} p_i = 1.$$

Let the experiment be performed n times, and the some of that experiments remain independent. Let us denote X_i the number of realizations $A_i (i = 1,2, \dots, k)$ in n experiments. Then each of the random variables X_i can only take a finite set

$0,1,2,\dots,n$ of values. Thus, the vector $(X_1, X_2, \dots, X_n)$ is a discrete random vector. For its probability function we have

$$p_{i_1 \dots i_k} = P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_k) = \frac{n!}{i_1! \dots i_k!} p_1^{i_1} \dots p_k^{i_k} , \qquad (43)$$

where $i_1 + i_2 + \dots + i_k = n$.

A random vector with probability function (43) is called *polynomially distributed.*

Continuous random vectors. A random vector is called *continuous* if its distribution function can be represented in the form

$$F(x_1, x_2, \dots, x_n) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} f(t_1, \dots, t_n) dt_1 \dots dt_n$$

$f(x_1, x_2, \dots, x_n)$is called *the density of the distribution* of a vector $(X_1, X_2, \dots, X_n)$ or also by the joint density of quantities $X_1, X_2, \dots, X_n$. The probability that a random vector $(X_1, X_2, \dots, X_n)$ is in the domain *G* of an *n*-dimensional space, can be written as follows

$$P((X_1, \dots, X_n) \in G) = \int \underset{G}{\dots} \int f(x_1, \dots, x_n) dx_1 \dots dx_n . \qquad (44)$$

Therefore, the density must satisfy

$$\int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} f(t_1, \dots, t_n) dt_1 \dots dt_n = 1. \qquad (45)$$

Uniform distribution. Vector $(X_1, X_1, \dots, X_n)$ is called distributed uniformly in the domain *G*, if it has a density that is constant in *G* and equal to 0 outside of *G*. This constant must be equal to $\frac{1}{V(G)}$, where $V(G)$ is the volume of the domain *G*.

Normal distribution. A vector $(X_1, X_2, \dots, X_n)$ is called *normally distributed* if it has a density of the form

$$f(x_1, \dots, x_n) = Ce^{-Q(x_1, \dots, x_n)} . \qquad (46)$$

Here Q is some positive definite quadratic form $x_1 - a_1, \dots, x_n - a_n$, $a_i -$ the constants, the constant *C* can be calculated from the condition (45). In the case *n*=2, the density of the normal distribution can be provided to the following form (Fig. 8).

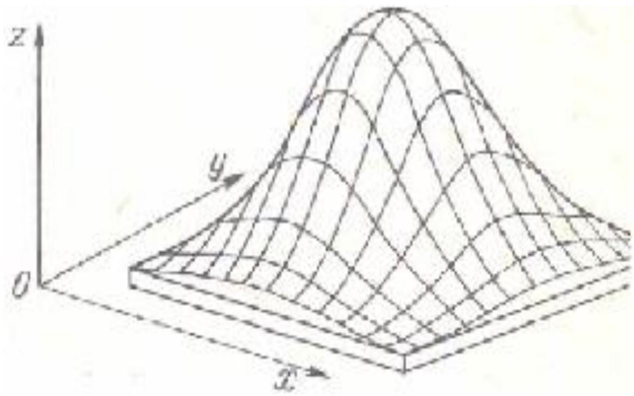

Figure 8.

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \cdot$$
$$\cdot \exp\left\{-\frac{1}{2(1-p^2)}\left[\frac{(x_1-a_1)^2}{\sigma_1^2} - 2p\frac{(x_1-a_1)(x_2-a_2)}{\sigma_1\sigma_2} + \frac{(x_2-a_2)^2}{\sigma_2^2}\right]\right\} \quad (47)$$

Boundary distributions. Let $F(x_1, x_2, \ldots, x_n)$ be the distribution function of a random vector $(X_1, X_2, \ldots, X_n)$. Then

$$P(X_{i_1} < x_{i_1}, \ldots, X_{i_k} < x_{i_k}) = F(c_1, \ldots, c_n) \quad (1 \le i_1 < \ldots < i_k \le n),$$

where $c_{i_1} = x_{i_1}, c_{i_2} = x_{i_2}, \ldots, c_{i_k} = x_{i_k}$, and all the others $c_j = +\infty$ is called *the k-dimensional boundary distribution* $F = (x_1, x_2, \ldots, x_n)$. It is a function of the distribution of a k-dimensional random vector $(X_{i_1}, X_{i_2}, \ldots, X_{i_k})$. In particular, when $k = 1$ the distributions of individual components are obtained:

$$F(+\infty, +\infty, \ldots, x_i, +\infty, \ldots, +\infty) = P(X_i < x_i) = F_i(x_i).$$

In the discrete case, the probability function of the k-dimensional boundary distribution is obtained by summing over indices which numbers are different from $i_1, i_2, \ldots, i_k$. In the continuous case, the density of the boundary distribution is obtained by integrating over variables which numbers are different from $i_1, i_2, \ldots, i_k$.

The total number of k-dimensional boundary distributions is C_n^k.

Example 52. Suppose we are given a two-dimensional discrete random variable (X_1, X_2) with a probability function

$$p_{ik} = P(X_1 = x_i^{(1)}, X_2 = x_i^{(2)}).$$

Then the probability functions of the individual components are obtained as boundary distributions

$$P(X_1 = x_i^{(1)}) = \sum_k p_{ik}, \quad P(X_2 = x_k^{(2)}) = \sum_l p_{lk}.$$

Example 53. Suppose we are given a three-dimensional random continuous quantity with density $f(x_1, x_2, x_3)$. Then, for example, the vector density (X_1, X_2) is obtained as a two-dimensional boundary distribution:

$$g(x_1, x_2) = \int_{-\infty}^{+\infty} f(x_1, x_2, x_3)\, dx_3.$$

The vector density 3th components is obtained as boundary distributions:

$$h(x_3) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x_1, x_2, x_3)\, dx_1\, dx_2.$$

Moments of a multidimensional random variable. Of particular interest are the first and second moments. If a random vector $(X_1, X_2, \dots, X_n)$ is discrete, then the numbers

$$v_j = \sum_{i_1, \dots, i_n} x_{i_j}^{(j)} p_{i_1 \dots i_n} \qquad (j = 1, \dots, n) \tag{48}$$

are called *the first initial moments* $(X_1, X_2, \dots, X_n)$. In the continuous case, the first initial moments are given by

$$v_j = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} x_j f(x_1, \dots, x_n) dx_1 \dots dx_n \qquad (j = 1, \dots, n), \tag{49}$$

where v_j - are the mathematical expectations of the individual components $v_j = M(X_j)$.

The second initial moments v_{ij} and the second central moments μ_{jk} are determined as follows:

Discrete case

$$v_{jk} = \sum_{i_1,\dots,i_n} x_{i_j}^{(j)} x_{i_k}^{(k)} p_{i_1\dots i_n}, \tag{50}$$

$$\mu_{jk} = \sum_{i_1,\dots,i_n} (x_{i_j}^{(j)} - v_j)(x_{i_k}^{(k)} - v_k) p_{i_1\dots i_n}. \tag{51}$$

Continuous case

$$v_{jk} = \int_{-\infty}^{+\infty}\dots\int_{-\infty}^{+\infty} x_j x_k f(x_1,\dots,x_n)dx_1\dots dx_n, \tag{52}$$

$$\mu_{jk} = \int_{-\infty}^{+\infty}\dots\int_{-\infty}^{+\infty} (x_j - v_j)(x_k - v_k) f(x_1,\dots,x_n)dx_1\dots dx_n. \tag{53}$$

It is said that the corresponding moments exist if the right sides of (48)-(53) converge absolutely. We have

$$v_{jk} = M(X_j X_k),$$

$$\mu_{jk} = M[(X_j - M(X_j)(X_k - M(X_k))] =$$
$$= M(X_j X_k) - M(X_j)M(X_k) = v_{jk} - v_j v_k.$$

The values μ_{jj} are equal to the variances of the individual components:

$$\mu_{jj} = D(X_j) = \sigma_j^2.$$

The quantity $\mu_{jj} = cov(X_j, X_k)$ is called *the covariance* (correlation moment) of random variables X_j, X_k and is denoted $cov(X_j, X_k)$. The matrix $\left\|\mu_{jk}\right\|_{j,k=1,\dots,n}$ is called *the covariance matrix* (*correlation*). Parameter

$$\rho_{jk} = \frac{\operatorname{cov}(X_j, X_k)}{\sqrt{D(X_j)D(X_k)}} = \frac{\operatorname{cov}(X_j, X_k)}{\sigma_j \sigma_k}. \qquad (54)$$

is called *the correlation coefficient* between X_j and X_k. It lies between -1 and $+1$.

Two random variables X and Y are called *noncorrelational,* if their correlation coefficient (their covariance) is equal to zero.

Example 54. The parameters in formula (47) for the density of a two-dimensional normal distribution have the following meanings:

$$a_1 = M(X_1),\ \ a_2 = M(X_2),\ \ \sigma_1 = \sqrt{D(X_1)},\ \ \sigma_2 = \sqrt{D(X_2)},$$

and

$$\rho = \frac{\operatorname{cov}(X_1, X_2)}{\sqrt{D(X_1)D(X_2)}}$$

is correlation coefficient between X_1 and X_2.

In the case of the n-dimensional distribution for the density

$$f(x_1,...,x_n) = Ce^{-Q(x_1,...,x_n)}$$

we have

$$f(x_1,...,x_n) = Ce^{-\frac{1}{2}\sum\limits_{i,j} b_{ij}(x_i - a_i)(x_j - a_j)},$$

where the parameters a_i, b_{ij} have the following meanings

$$a_i = M(X_i),\ \ b_{ij} = \frac{\Delta_{ij}}{\Delta},$$

where $\Delta = \det(\Delta_{ij})$, Δ_{ij} - matrix of algebraic add-ons of the covariance matrix.

Thus, the n-dimensional normal distribution, as in the one-dimensional case, is completely determined by specifying the first and second moments.

Conditional distributions. Let (X,Y) be a random vector, and $F(x,y)$ its distribution function.

Function

$$F(x/y) = \lim_{h \to 0} \frac{P(X < x, y \le Y < y + h)}{P(y \le Y < y + h)}. \qquad (55)$$

is called *the conditional distribution* of X under the condition that Y takes the value y. Is defined similarly $F(x/y)$.

In the discrete case, for the conditional probability function we have

$$P(X = x_i / Y = y_k) = \frac{p_{ik}}{\sum\limits_i p_{ik}},\ P(Y = y_k / X = x_i) = \frac{p_{ik}}{\sum\limits_k p_{ik}}, \tag{56}$$

In the continuous case for conditional densities, we have

$$f(x/y) = \frac{f(x,y)}{\int\limits_{-\infty}^{+\infty} f(x,y)dx}, \qquad f(y/x) = \frac{f(x,y)}{\int\limits_{-\infty}^{+\infty} f(x,y)dy}. \tag{57}$$

In the denominators of formulas (56) and (57) are the boundary distributions of the components that determine the condition.

Independence of random variables. Random variables $X_1, X_2, \dots, X_n$. are called *independent* if

$$F(x_1,\dots,x_n) = F_1(x_1)\dots F_n(x_n),$$

where $F_i(x_i)$ is the distribution function of the i-th component X_i (one-dimensional boundary distribution).

While in the general case of a joint distribution, if the distributions of individual components are known, nothing can be said, in the case of the independence of random variables, the function $F(x_1,\dots,x_n)$ is completely determined through the distributions of the individual components.

Example 55. Shooting at targets. Let’s assume that the vertical deviation X and the horizontal deviation Y have ~~a~~ normal distribution relative to the center

$$X \in N(x,0,\sigma_1), \quad Y \in N(y,0,\sigma_2).$$

It might be considered that X and Y are independent. Then for the density of the vector (X,Y) denoting the location of the bullet, we have

$$f(x,y) = f_1(x)f_2(y) =$$
$$= \frac{1}{\sqrt{2\pi}\,\sigma_1}\exp\left(-\frac{1}{2}\left(\frac{x}{\sigma_1}\right)^2\right)\cdot\frac{1}{\sqrt{2\pi}\,\sigma_2}\exp\left(-\frac{1}{2}\left(\frac{y}{\sigma_2}\right)^2\right) =$$
$$= \frac{1}{2\pi\,\sigma_1\sigma_2}\exp\left(-\frac{1}{2}\left[\left(\frac{x}{\sigma_1}\right)^2 + \left(\frac{y}{\sigma_2}\right)^2\right]\right).$$

Therefore, (X,Y) is normally distributed with parameters $a_1 = a_2 = 0;\ \sigma_1, \sigma_2;\ \rho = 0.$

For independent random variables, some very important and frequently used theorems on mathematical expectation and variance are valid.

If *X* and *Y* are independent, then

$$M(XY) = M(X)M(Y). \tag{59}$$

From this and the definition of covariance it follows that the independence of *X* and *Y* implies that *X* and *Y* are uncorrelated.

In the case of normal distribution, in correlation implies independence, which becomes obvious if in (47) we substitute $\rho = 0$ and apply the functional equality for the exponent.

The dispersion of the sum of independent random variables is equal to the sum of the dispersion, i.e.

$$D(X_1 + ... + X_n) = D(X_1) + ... + D(X_n),$$

if $X_1, X_2, \dots, X_n$ are independent.

Regression dependence. For many phenomena in nature and technology, stochastic (random) dependencies are typical. Between two random variables there is a stochastic dependence in the general case when there are some random factors that affect both random quantities and some factors acting only on the first or only the second random variable. Therefore, if

$$X = f(Z_1,...,Z_m,X_1,...,X_j),\ Y = g(Z_1,...,Z_m,Y_1,...,Y_k),$$

then *X* and *Y* are stochastically dependent. In the theory of regression, of particular importance is the objective: the prediction of the random variable *Y* that interest ~~to~~ us, if other random variables on which *Y* depends stochastically, have taken specific importance.

Regression lines. Curves in the plane *x*, *y* defined by equations

$$\overline{y}(x) = M(Y/X = x)$$

and

$$\overline{x}(y) = M(X/Y = y)$$

are called *regression lines* *Y* comparatively to *X* and *X* regarding to *Y*. In this case $M(Y/X = x)$ - the mathematical expectation of *Y* under the condition that *X* has taken the value of *x*. In the case of continuous *X* and *Y* we have

$$M(Y/X = x) = \int_{-\infty}^{+\infty} yf(y/x)dy, \qquad M(X/Y = y) = \int_{-\infty}^{+\infty} xf(x/y)dx.$$

In this case, $f(x/y)$ or $f(y/x)$ are conditional densities.

The regression lines have the following meaning: the best prediction of *Y*, if $X = x_0$, i.e. $\overline{y}(x_0)$. In this case, "best" means that for an arbitrary function *u*(*X*) fair the inequality

$$M(Y - u(X))^2 \geq M(Y - \overline{y}(X))^2.$$

This can also be expressed as follows: the regression function $\overline{y}(x)$ is a function that minimizes the average quadratic error of the prediction value *Y* based on the values of *X*. The corresponding way can be interpreted also $\overline{x}(y)$.

Direct regressions. Random variables *X* and *Y* are called linearly correlated if the regression lines are straight. These "direct regressions" are given by the following equations

regression of Y regarding to

$$X: \; y = \mu_Y + \beta_{Y/X}(x - \mu_X), \tag{60}$$

regression of *X* regarding to

$$Y: \; x = \mu_X + \beta_{X/Y}(y - \mu_Y). \tag{61}$$

values $\beta_{Y/X}$ and $\beta_{X/Y}$ are called (theoretical) *regression coefficients*. They are calculated as follows:

$$\beta_{Y/X} = \frac{\sigma_Y}{\sigma_X}\rho, \quad \beta_{X/Y} = \frac{\sigma_X}{\sigma_Y}\rho. \tag{62}$$

In this case, ρ there is *a correlation coefficient X* and Y:

$$\rho = \frac{\mathrm{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{\mu_{XY}}{\sigma_X \sigma_Y}. \tag{63}$$

And

$$\sigma_X = \sqrt{D(X)},\ \ \sigma_Y = \sqrt{D(Y)}.$$

The parameters μ_X, μ_Y entering into formulas (60), (61) are the mathematical expectations of *X* and *Y*:

$$\mu_X = M(X),\ \ \mu_Y = M(Y).$$

In the case when *X* and *Y* are not linearly correlated, using equations (60) and (61), using (62) and (63), we can form the equations of two straight lines. They are also called regression lines and in this case are linear approximations of the true regression lines.

Functions of random variables. Let a continuous random vector be given (X,Y) and $f(x,y)$ its density. It is required to find the distribution of random variables

$$X+Y,\ X\cdot Y,\ \frac{X}{Y}.$$

The sum $X+Y$ is also a continuous random variable, and its density

$$f(z) = \int_{-\infty}^{+\infty} f(x, z-x)dx.$$

If *X* and *Y* are independent, so that

$$f(x,y) = f_1(x) f_2(y),$$

then

$$f(z) = \int_{-\infty}^{+\infty} f_1(x) f_2(z-x)dx. \tag{64}$$

Consequently, the density of a sum is the folding of the densities of individual sums.

Example 56. If (X, Y) is distributed normally, and $f(x, y)$ given by (47). Then for the density $X + Y$ we are getting

$$f(z) = \frac{1}{\sqrt{2\pi(\sigma_1^2 + 2\rho\sigma_1\sigma_2 + \sigma_2^2)}} \exp\left(-\frac{1}{2}\frac{(z-(a_1+a_2))^2}{\sigma_1^2 + 2\rho\sigma_1\sigma_2 + \sigma_2^2}\right).$$

Consequently, the value $Z = X + Y$ is again normally distributed with parameters

$$\sqrt{\sigma_1^2 + 2\rho\sigma_1\sigma_2 + \sigma_2^2},\ a_1 + a_2.$$

In the case of independence, the reverse is also true: if the sum of two independent random variables is normally distributed, then the individual terms are normally distributed.

The product. If (X, Y) is a random vector and $Z = X \cdot Y$. Then for the density we have

$$f(z) = \int_{-\infty}^{+\infty} f\left(x, \frac{z}{x}\right)\frac{1}{|x|}dx.$$

The ratio. If (X, Y) is a random vector; the relation X/Y is some random continuous function with density function

$$f(z) = \int_{0}^{+\infty} xf(zx, x)dx - \int_{-\infty}^{0} xf(zx, x)dx.$$

CHARACTERISTIC FUNCTIONS

The characteristic function $\psi(t)$ of a random variable *X* is the mathematical expectation of a random variable e^{itX} :

$$\psi(t) = M(e^{itX}), \tag{65}$$

where t is a real parameter.

If *F*(*x*) is a distribution function of *X*, then

$$\psi(t) = \int_{-\infty}^{+\infty} e^{itx} dF(x) . \tag{66}$$

In the case of a discrete distribution

$$\psi(t) = \sum_{k=0}^{\infty} e^{itk} p_k \tag{67}$$

(67)- Fourier series with coefficients p_k.

In the case of continuous distribution

$$\psi(t) = \int_{-\infty}^{+\infty} e^{it} x_k f(x) dx \tag{68}$$

(68) the Fourier integral.

Example 57. If *X* obey Poisson's law with a parameter λ. Then the characteristic function is

$$\psi(t) = \sum_{k=0}^{\infty} e^{itk} \frac{\lambda^k}{k!} e^{-\lambda} = \exp(\lambda \cdot (e^{it} - 1)) . \tag{69}$$

Example 58. *X* is uniformly distributed on (-*a, a*). Then

$$\psi(t) = \int_{-a}^{+a} e^{itx} \frac{1}{2a} dx = \frac{\sin(at)}{at}.$$

Example 59. $X \in N(x, a, \sigma)$. Then the characteristic function is

$$\psi(t) = \frac{1}{\sqrt{2\pi}\,\sigma} \int_{-\infty}^{+\infty} \exp\left(itx - \frac{1}{2}\frac{(x-a)^2}{\sigma^2} \right) dx = \exp\left(iat - \frac{\sigma^2 t^2}{2} \right). \qquad (70)$$

Properties of Characteristic function.

1) The characteristic function is uniformly continuous on the entire real-axis.

2) For any characteristic function, we have

$$\psi(0) = 1, \quad |\psi(t)| \le 1 \quad (-\infty < t < +\infty).$$

3) If $Y = aX + b$ with constants a and b, then

$$\psi_Y(t) = \psi_X(at) e^{ibt},$$

where ψ_X is the characteristic function of *X*.

Using characteristic functions, it's possible to calculate the moments easily.

If the random variable *X* possess a moment of order *n*, then the characteristic function *X* is *n* times differentiable to *t* where

$$\psi^{(k)}(0) = i^k M(x^k) = i^k v_k. \qquad (71)$$

Example 60. If *X* normally distributed with parameters *a* and σ. We calculate *M*(*X*) and *D*(*X*). According to (70):

$$\psi(t) = \exp\left(iat - \frac{\sigma^2 t^2}{2} \right).$$

Then, by (71)

$$iv_1 = \psi'(0) = ia$$

and

$$-v_2 = \psi''(0) = -\sigma^2 - a^2.$$

Hence,

$$M(X) = v_1 = a$$

and

$$D(X) = v_2 - v_1^2 = \sigma^2 + a^2 - a^2 = \sigma^2.$$

For applied aims is significant the fact that the characteristic function of the sum of independent random variables is equal to the product of their characteristic functions.

If the random variables X_1 and X_2 are continuous, then the density of the sum, according to (64), is the convolution of both densities. This property, therefore, is nothing else than the convolution theorem for the Fourier transform.

Example 61. Applications of the convolution theorem. Let the random variable *X* be distributed binomially with the parameters *n* and *p*. It is required to find its characteristic function. As it is known, *X* can be interpreted as the number of realizations of event *A* in *n* independent trials, if the probability of carrying out *A* in each test is *p*. Therefore, *X* can be written as a sum:

$$X = X_1 + X_2 + ... + X_n.$$

where

$$X_j = \begin{cases} 0, \text{ if A is not realized in the j-th case} \\ 1, \text{ if A is realized in the j-th case.} \end{cases}$$

Under clause, X_j are independent random variables. Hence, according to the convolution theorem, we have

$$\psi_X(t) = \prod_{j=1}^{n} \psi_{X_j}(t)$$

and

$$\psi_{X_j}(t) = M\left(e^{itX_j}\right) = e^{it\cdot 0} \cdot q + e^{it\cdot 1} \cdot p = q + pe^{it}.$$

Thus

$$\psi_X(t) = (q + pe^{it})^n. \qquad (72)$$

The Reversal Formula and the Singularity Theorem. Let $F(x)$ be the distribution function, and $\psi(t)$ is the characteristic function of the random variable X. If x_1, x_2 the points of continuity $F(x)$, then

$$F(x_2) - F(x_1) = \frac{1}{2\pi} \lim_{c \to +\infty} \int_{-c}^{c} \frac{e^{-itx_1} - e^{-itx_2}}{it} \psi(t) dt. \qquad (73)$$

If the random variable X is continuous and $f(x)$ is the density of $F(x)$, then formula (73) is simplified:

$$f(x)=\frac{1}{2\pi}\int_{-\infty}^{+\infty}e^{-itx}\psi(t)dt.$$

Thus, the density is obtained from the characteristic function by the inverse Fourier transform. It follows from the Reversal Formula that the distribution function of a random variable is uniquely determined by its characteristic function. If, for example, in any way a characteristic function $\exp\left(iat-\frac{\sigma^2t^2}{2}\right)$ is obtained for X, then, according to the Singularity Theorem and formula (70), we have

$$X=N(x,a,\sigma).$$

Example 62. Let two independent random variables be normally distributed:

$$X=N(x,a_1,\sigma_1),\ Y=N(y,a_2,\sigma_2).$$

It is required to find the distribution for $X+Y$. As

$$\psi_X(t)=e^{ia_1t-\sigma_1^2t^2/2},\qquad \psi_Y(t)=e^{ia_2t-\sigma_2^2t^2/2}.$$

Then, according to the convolution theorem

$$\psi_{X+Y}(t)=\psi_X(t)\psi_Y(t)=e^{i(a_1+a_2)t-\frac{(\sigma_1^2+\sigma_2^2)t^2}{2}}.$$

Due to the Singularity Theorem, the only distribution having this distribution function is

$$N(x+y,a_1+a_2,\sigma=\sqrt{\sigma_1^2+\sigma_2^2}).$$

Thus, the random variable $X+Y$ is again distributed normally with parameters

$$a = a_1 + a_2 \quad \text{and} \quad \sigma = \sqrt{\sigma_1^2 + \sigma_2^2}.$$

Example 63. If X and Y are independent random variables that obey the Poisson law:

$$P(X = k) = \frac{\lambda_1^k}{k!} e^{-\lambda_1}, \qquad P(Y = k) = \frac{\lambda_2^k}{k!} e^{-\lambda_2}.$$

It is required to find the distribution of the sum $X+Y$. By the formula (69) we find

$$\psi_X(t) = e^{\lambda_1 (e^{it} - 1)}, \psi_Y(t) = e^{\lambda_2 (e^{it} - 1)}.$$

Due to independence and according to the convolution theorem

$$\psi_{X+Y}(t) = \psi_X(t)\psi_Y(t) = e^{(\lambda_1 + \lambda_2)(e^{it} - 1)}. \qquad (75)$$

Due to the Singularity Theorem, the only distribution having (75) as a characteristic function is a Poisson distribution with parameter $\lambda_1 + \lambda_2$. Thus, the sum of two independent random variables that obey to Poisson's law is again distributed according to Poisson's law with the parameter $\lambda_1 + \lambda_2$.

Note. Here the opposite is also true: if the sum of two independent random variables is distributed according to Poisson's law, then the terms are distributed according to Poisson's law.

Limit theorem for characteristic functions. A sequence $\{F_n(x)\}$ of distribution functions is called converging mainly on the distribution function $F(x)$ f, at all points of continuity

$$\lim_{n \to \infty} F_n(x) = F(x).$$

In the discrete case, convergence mainly $F_n(x)$ to $F(x)$ means that the corresponding probability functions converge:

$$p_k^{(n)} \to p_k$$

for all k.

In the continuous case, convergence essentially follows (if $f(x)$ are continuous), that

$$f_n(x) \to f(x)$$

for all x.

If the sequence $\{F_n(x)\}$ of distribution functions converges mainly to the distribution function $F(x)$, then the sequence of the corresponding characteristic functions $\{\psi_n(x)\}$ converges to $\psi(x)$ the characteristic function $F(x)$. This convergence is uniform in every finite interval.

Of greater significance is the converse theorem: if a sequence of characteristic functions converges $\{\psi_n(t)\}$ to a continuous function $\psi(t)$, then the sequence of the corresponding distribution functions $\{F_n(x)\}$ converges to some distribution function $F(x)$ and $\psi(t)$, is the characteristic function $F(x)$.

Note. The inverse theorem condition holds, in particular, if one of two conditions is met:

1) $\{\psi_n(t)\}$ converges uniformly on each final interval to the function $\psi(t)$.

2) $\{\psi_n(t)\}$ converges to the characteristic function.

Example 64. If

$$P(n,k) = C_n^k p_n^k q_n^{n-k}$$

is a sequence of binomial distributions with parameters n, p_n and

$$\lim_{n\to\infty} np_n = \lambda .$$

Which distribution does this sequence converge on? The corresponding sequence of characteristic functions

$$\psi_n(t) = (q_n + e^{it} p_n)^n$$

obviously converges to the function

$$\psi(t) = e^{\lambda(e^{it}-1)} .$$

Consequently, the limit distribution is a Poisson distribution:

$$p_\lambda(k) = \frac{\lambda^k}{k!} e^{-\lambda} \qquad (k = 0,1,\ldots) .$$

Generating functions. In the case of discrete random variables, which can take only the values 0,1,2, ..., often the generating functions are used instead of the characteristic functions.

If p_k is a probability function of some discrete random variable *X* of the indicated type, and z is a complex parameter. Then

$$\varphi(z)=\sum_k p_k z^k$$

is called *the generating function* of the random variable *X*.

The function $\varphi(z)$ is analytic in the disk $|z|<1$. Its limit $z\to e^{it}$ gives a characteristic function of *X*.

Generating functions have similar to those of characteristic functions.

Characteristic functions of multidimensional random variables. The *characteristic function of the n-dimensional random variable* is the mathematical expectation of a quantity

$$\exp\left(i\sum_{k=1}^{n}t_k X_k\right): \qquad \psi(t_1,\ldots,t_n)=M\left(\exp\left(i\sum_{k=1}^{n}t_k X_k\right)\right), \tag{76}$$

where $t_1, t_2, \ldots, t_n$ are the real parameters.

Example 65. The characteristic function of an n-dimensional normal distribution has the form

$$\psi(t_1,\ldots,t_n)=\exp\left(i\sum_k a_k t_k-\frac{1}{2}\sum_{j,k}b_{jk}t_j t_k\right), \tag{77}$$

where (b_{jk}) is the covariance matrix. In particular, for *n*=2 we have

$$\psi(t_1,t_2)=\exp\left[i(a_1t_1+a_2t_2)-\frac{1}{2}\left(\sigma_1^2t_1^2+2\rho\sigma_1\sigma_2t_1t_2+\sigma_2^2t_2^2\right)\right]. \tag{78}$$

If $X_1, X_2, \ldots, X_n$ are independent random variables, then

$$\psi(t_1,\ldots,t_n)=\prod_{k=1}^{n}\psi_k(t_k),$$

where ψ_k are the characteristic functions of the individual components X_k.

If

$$Z=X_1+X_2+\ldots+X_n$$

and

$$\psi(t_1,...,t_n)$$

are the characteristic function of the vector $(X_1, X_2, \dots, X_n)$, then

$$\psi_Z(t) = \psi(t_1,...,t_n).$$

LIMIT THEOREMS

Law of large numbers. A sequence $\{X_n\}$ of random variables is called convergent in probability to a random variable *X* if for any $\varepsilon > 0$ the equality

$$\lim_{n\to\infty} P(|X_n - X| < \varepsilon) = 1.$$

It is said that the sequence of random variables $X_1, X_2, \dots$ obeys the weak law of large numbers if for any $\varepsilon > 0$ the equality

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{k=1}^{n} X_k - \frac{1}{n}\sum_{k=1}^{n} M(X_k)\right| < \varepsilon\right) = 1. \qquad (79)$$

In other words, if

$$Z_n = \frac{1}{n}\sum_{k=1}^{n} X_k - \frac{1}{n}\sum_{k=1}^{n} M(X_k)$$

converges to 0 "in probability".

To infer the weak law of large numbers, Chebyshev's inequality is important. Let *X* be a random variable having finite variance. Then forany $\varepsilon > 0$ the inequality

$$P(|X - M(X)| \ge \varepsilon) \le \frac{D(X)}{\varepsilon^2}. \qquad (80)$$

Chebyshev's Theorem. If $X_1, X_2, \dots$ is a sequence of pairwise independent random variables whose variances are uniformly bounded, i.e.

$$D(X_k) \le C,$$

for each k, then this sequence obeys to the weak law of large numbers.

The Bernoulli theory is a direct consequence of this. Let *m* be the number of realizations of the event *A* in *n* independent trials, and suppose that in each that test *A* has probability p. Then the frequency

$$\frac{m}{n} = W_n(A)$$

tends "in probability" to p, i.e.

$$\lim_{n\to\infty} P(|W_n(A) - p| < \varepsilon) = 1 \tag{81}$$

for any $\varepsilon > 0$.

If $X_1, X_2, \ldots$ is a sequence of pairwise independent random variables with equal mathematical expectations

$$M(X_1) = M(X_2) = \ldots = a\,,$$

and are uniformly bounded $D(X_k)$, then for any $\varepsilon > 0$ fair equity

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{k=1}^{n} X_k - a\right| < \varepsilon\right) = 1\,. \tag{82}$$

The relation (82) is the theoretical basis of the rule of the arithmetic mean for measurements. Let it be necessary to measure the unknown quantity a. Because of random errors, the measurements are repeated n times, and the individual measurements are independent of each other; the k-th dimension can be described by the random variable *X*. If there is no systematic error during the measurement, then $M(X_k) = a$. Then, according to (82), by constructing the average of the arithmetic measured values, for sufficiently large n, with a probability arbitrarily close to 1, we obtain a value arbitrarily close to the required value of a.

A sequence $\{X_n\}$ of random variables is called almost convergent to a random variable *X* if

$$P\left(\lim_{n\to\infty} X_n = X\right) = 1\,.$$

It is said that the sequence $X_1, X_2, \ldots$ of random variables obeys the strong law of large numbers, if

$$P\left(\lim_{n\to\infty}\left(\frac{1}{n}\sum_{k=1}^{n}X_k - \frac{1}{n}\sum_{k=1}^{n}M(X_k)\right) = 0\right) = 1 \quad . \tag{83}$$

In other words, if

$$\frac{1}{n}\sum_{k=1}^{n}X_k - \frac{1}{n}\sum_{k=1}^{n}M(X_k)$$

“almost converges” to 0.

Kolmogorov's theorem. If a sequence $\{X_n\}$ of random events that are independent from each other satisfies the condition

$$\sum_{n=1}^{\infty}\frac{D(X_n)}{n^2} < +\infty \quad ,$$

then it obeys the strong law of large numbers.

On the connection between the strengthened and weak law, we can say the following: if a sequence of random variables obeys an amplified law, then it obeys the weak, but not vice versa.

Limit theorem de Moivre - Laplace.

Local limit theorem. If the probability of the occurrence of the event A in n independent experiments is constant and equal to p $(0 < p < 1)$, then the probability

$$P(n,k) = C_n^k p^k q^{n-k}$$

that in these experiments the event A occurs exactly k times, satisfies the relation

$$\lim_{n\to\infty}\frac{P(n,k)}{\frac{1}{\sqrt{2\pi}}\sqrt{npq}\, e^{-x^2/2}} = 1, \tag{84}$$

where

$$x = \frac{k-np}{\sqrt{npq}} \quad .$$

In other words: the binomially distributed random variable is asymptotically distributed normally with the parameters $a = np$ and $\sigma = \sqrt{npq}$.

Example 66. Let the probability of appearance in the production of a defective part is 0,005. How great is the probability that among the 10 000 items 40 will be defective? So, it should be determined $P(n,k)$ for n=10 000, k=40, p=0,005. According to (84)

$$P(n,k) \approx \frac{1}{\sqrt{2\pi}\sqrt{npq}} e^{-\frac{1}{2}\left(\frac{k-np}{npq}\right)^2}; \qquad \sqrt{npq} = 7{,}05; \ \frac{k-np}{\sqrt{npq}} = -1{,}42.$$

Consequently,

$$P(n,k) \approx \frac{1}{7{,}05\sqrt{2\pi}} e^{-\frac{1}{2}(-1{,}42)^2}.$$

From Application 1 we find $\varphi(1{,}42) = 0{,}1456$. Thus,

$$P(n,k) = \frac{0{,}1456}{7{,}05} \approx 0{,}0206.$$

When calculating from the exact formula, we get $P(n,k) = 0{,}0197$.

Example 67. Let the probability of appearance in the production of a defective part be 0,005. How great is the probability that among the 10 000 items 40 will be defective? So, it should be determined $P(n,k)$ for n=10 000, k=40, p=0,005. We have

$$P_n(k) \approx \frac{1}{\sqrt{npq}} \varphi(\tilde{o})$$

and

$$\sqrt{npq} = 7{,}05; \qquad \frac{k-np}{\sqrt{npq}} = -1{,}42..$$

From Application 1 we find $\varphi(1{,}42) = 0{,}1456.$. Thus,

$$P(10000{,}40) = \frac{0{,}1456}{7{,}05} \approx 0{,}0206.$$

Example 68. The shop produces 75% production premium. Find the probability that a run of 160 products will be 125 premium products.

Solution: Denoted by A events of randomly selected product premium. By condition

$$n = 160, m = 125, \ p = P(A) = 0.75; q = 1 - p = 0.25 .$$

Finding

$$np = 160 * 0.75 = 120$$

$$\sqrt{npq} = \sqrt{120 * 0.75} = \sqrt{30} \approx 5.477$$

$$x = \frac{m - np}{\sqrt{npq}} = \frac{125 - 120}{5.477} = \frac{5}{5.477} \approx 0.91$$

From the Application 1 we find: $\varphi(0,91) = 0,2637$. Applying local Laplace formula, we obtain the desired probability

$$P_{160}(125) = \frac{\varphi(x)}{\sqrt{npq}} = \frac{0,2637}{5,477} \approx 0,0481 .$$

The integral limit theorem. Let X be a binomially distributed random variable with parameters n and p. (Consequently, X can be interpreted as the number of realizations of the event A in n independent trials with $P(A) = p$ in a separate trial). Then, uniformly with respect to a and b ($-\infty \le a < b \le +\infty$), comparison is done

$$\lim_{n\to\infty} P\left(a \le \frac{X - np}{\sqrt{npq}} < b \right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx =$$

$$= \Phi_0(b) - \Phi_0(a). \qquad (85)$$

Example 69. Let there be a situation described in Example 51. We are looking for the probability that in a box with 10 000 parts there are not more than 70 defective ones:

$$P(X \le 70) = P\left(\frac{-50}{\sqrt{49{,}75}} \le \frac{X-np}{\sqrt{npq}} \le \frac{20}{\sqrt{49{,}75}}\right) =$$

$$= P\left(-7{,}09 \le \frac{X-np}{\sqrt{npq}} \le 2{,}84\right) \approx \frac{1}{\sqrt{2\pi}} \int_{-7{,}09}^{2{,}84} e^{-x^2/2} dx =$$

$$= \Phi_0(2{,}84) - \Phi_0(-7{,}09)$$

Since $\Phi_0(-x) = -\Phi_0(x)$, then

$$P(X \le 70) = \Phi_0(2{,}84) + \Phi_0(7{,}09)$$

From the Application 2 we find $\Phi_0(2{,}84) = 0{,}4977$; but $\Phi_0(7{,}09)$ not in the table, since it differs from 0,5 very little. Thus $P(X \le 70) \approx 0{,}9977$.

If the probability of occurrence of an event *p* is constant ($0 < p < 1$) and if events of *n* infinitely increases, then probability is written as $P(k_1 \le m \le k_2)$ or $P(k_1; k_2)$:

$$P_n(k_1, k_2) \approx \frac{1}{\sqrt{2\pi}} \int_{x_1}^{x_2} e^{-t^2/2} dt = \Phi(x_2) - \Phi(x_1)$$

or

$$P(k_1, k_2) = \Phi(x_2) - \Phi(x_1),$$

where

$$x_1 = \frac{k_1 - np}{\sqrt{npq}}, \qquad x_2 = \frac{k_2 - np}{\sqrt{npq}}.$$

$\Phi(x)$ is supporting function of Laplace function, this function values are given by Application 2.

Properties of $\Phi(x)$**:**

1) $\Phi(x)$ is odd function: $\Phi(-x)=-\Phi(x)$.

2) if $x \ge 5$, then $\Phi(x)=0{,}5$.

The sufficient accuracy of the formula is ensured when $\sqrt{npq} \ge 15$.

Example 70. Dice is thrown 144 times. Find the probability of getting the edge with six points from 20 to 25 times.

Solution: denoted by *A* event of getting the edge with six points in a single throw of the dice. Find the probability of the event *A*. By condition

$$n = 144,\ p = P(A) = \frac{1}{6},\ q = \frac{5}{6},\ m_1 = 20,\ m_2 = 25.$$

Perform the necessary calculations:

$$np = 144 \cdot \frac{1}{6} = 24; \qquad \sqrt{npq} = \sqrt{24 \cdot \frac{5}{6}} = \sqrt{20} = 4{,}427.$$

We compute x_1 and x_2 by formula:

$$x_1 = \frac{m_1 - np}{\sqrt{npq}} = \frac{20 - 24}{4.472} = -0.89;$$

$$x_2 = \frac{m_2 - np}{\sqrt{npq}} = \frac{25 - 24}{4.472} = 0.22$$

We apply the Laplace integral formula, the values of the Laplace found from Application 2:

$$P_{144}(20;25) \approx \Phi(x_2) - \Phi(x_1) = \Phi(0{,}22) - \Phi(-0{,}89) =$$

$$= \Phi(0{,}22) + \Phi(0{,}89) = 0{,}0871 + 0{,}3133 = 0{,}4004.$$

Central limit theorem. If $\{X_n\}$ is a sequence of independent random variables, and if

$$Z_n = \sum_{k=1}^{n} \frac{X_k - M(X_k)}{\sqrt{\sum_{k=1}^{n} D(X_k)}}. \tag{86}$$

Values are called normalized and centered sums ($D(Z_n) = 1,\ M(Z_n) = 0$). Let $\Phi_n(x)$ are distribution functions of Z_n, and $F_n(x)$ a distribution function of X_k. We denote by

$$B_n^2 = \sum_{i=1}^{n} D(X_i), \quad B_n > 0.$$

A necessary and sufficient condition for the equality

$$\lim_{n\to\infty} \hat{O}_n(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt \tag{87}$$

Is the following condition (Lindeberg)

$$\lim_{n\to\infty} \frac{1}{B_n^2} \sum_{k=1}^{n} \int_{|X-M(X_k)|>\varepsilon B_n} (X - M(X_k))^2 dF_k(x) = 0 \tag{88}$$

Note. The condition is satisfied particularly, if all X have the same distribution, for which the first and second moments are finite.

Condition (88) means that the individual terms

$$\frac{X_i - M(X_i)}{B_n}$$

of which Z_n consists according to formula (86), are uniformly small. In this case the meaning of the central limit theorem is this. If a random variable can be represented as the sum of a large number of independent terms, each of which contributes only an insignificant contribution to the sum, then this sum is distributed approximately normally.

Example 71. Assume that a measurement is made and X is a random measurement error. A random variable X appears as a result of the additive superposition of a large number of factors that do not depend on each other, which generate errors; Each of these factors has a small effect on the error. Thus, the value of X can be assumed to be distributed normally.

Example 72. Let X be the length of a birch leaf randomly selected from a many ripped leaves. Then X is a random variable obtained by imposing many small factors that do not depend on each other. Therefore, a normal distribution can be adopted for X.

Example of the 1-st midterm control (MC-1)

1. The frequency of boy's birth s equal to 0,51. How many boys will be among 700 newborns?

2. There are 10 black and 7 white balls in the box. 5 balls randomly are taken out from it. What is the probability that there will be 3 white and 2 black balls among them?

3-5. For each of the following tasks: (3, 4, 5) specify the formula (from A), B), C), D), E)) needed to solve it, and the name of this formula:

$$A)\quad p(A)=\sum_{i=1}^{n} p(B_i)p(A/B_i) \qquad B)\quad p(B_i/A)=\frac{p(B_i)\cdot p(A/B_i)}{\sum_{i=1}^{n} p(B_i)p(A/B_i)}$$

$$C)\quad p_n(m)=C_n^m p^m q^{n-m} \qquad D)\quad p_n(m)=\frac{\lambda^m}{m!}e^{-\lambda}$$

$$E)\quad p_n(k_1,k_2)=\Phi(x_2)-\Phi(x_1)$$

3. The probability of winning the lottery is 0.25. What is the probability that someone purchasing 8 tickets will win by 5 of them?

4. 20 machines work in the manufactory. Among them: 10 have A grade, 6 – B grade, and 4 – C grade. The probabilities that quality of the detail will be excellent for these machines are, respectively, 0.9; 0.8; and 0.7. What percentage of excellent details manufactory produces in general?

5. 100 trees have been planted. Find the probability that the number of engrafted trees is located from 80 to 90, if the probability of engraftment of a tree equals 0,9.

Example of the 2-st midterm control (MC-2)

1. Random value X is distributed uniformly with distribution density function

$$f(x)=\begin{cases}0, & if \quad x\le 2, x>9\\ 1/7, & if \quad 2<x\le 9\end{cases}.$$

Find:

a) The distribution function,

b) Dispersion,

c) The probability of falling into the interval (0;5).

2. The exponential distribution with parameter $\lambda = 5$ is given.

Find:

a) The distribution density function,

b) Mathematical expectation.

3. Mathematical expectation and the mean square deviation of normally distributed random variable are given: $M(X)=7, \sigma(X)=2$.

Find:

a) The distribution density function,

b) Dispersion.

4. A bus of some route goes on schedule with an interval of 7 minutes. Find the probability that a passenger approached the stop, will be waiting for the bus less than 4 minutes.

5. Using Chebyshev's inequality estimate the probability that $|X-MX|<\sqrt{0{,}1}$ if D(X)=0,16.

6. Find confidence interval for estimation of mathematical expectation a with reliability 0,96 of normally distributed random value, if $\sigma=4,\ \bar{x}_T=19,\ n=25$.

7. The random variable is given by distribution series

X	-3	0	1	4
P	0,1	0,3	0,4	0,2

Find:

a) Distribution function,

b) Mathematical expectation,

c) Dispersion,

d) Mode,

e) The probability of falling into the interval (-1;3);

f) Mean square deviation.

8. The random variable is given by the distribution function:

$$F(x)=\begin{cases} 0, & x\le 0,\\ \dfrac{x^2}{16}, & 0<x\le 4,\\ 1, & x>4. \end{cases}$$

Find:

a) The distribution density function,

b) Mathematical expectation,

c) Dispersion,

d) The probability of falling into the interval (1; 4),

e) Mean square deviation.

EXAMPLE OF TEST QUESTIONS ON PROBABILITY THEORY

1. What the probability of emergence at least of one event is equal to

A) $P(A) = p_1 \cdot p_2 \cdot ... \cdot p_n$

B) $P(A) = p_1 + p_2 + ... + p_n$

C) $P(A) = 1 - q_1 \cdot q_2 \cdot ... \cdot q_n$

D) $P(A) = q_1 \cdot q_2 \cdot ... \cdot q_n$

E) $P(A) = p_i + q_i$

2. To find formula of a total probability

A) $P(A) = P_A(B_1) + P_A(B_2)$

B) $P(A) = P_A(B_2) - P_A(B_1)$

C) $P(A) = \sum_{i=1}^{n} P(B_i) P_{B_i}(A)$

D) $P(A) = \sum_{i=1}^{n} P(B) P_{B_i}(B)$

E) $P(A) = P(A_i) \cdot P(B)$

3. What of inequalities fairly for probability of any random variable

A) $P(A) \geq 1$

B) $0 < P(A) < \infty$

C) $P(A) < 1$

D) $0 \leq P(A) \leq 1$

E) $0 \leq P(A) \geq 1$

4. To find Bernoulli's formula

A) $P_n(k) = k!$

B) $P_n(k) = C_n^k q p^{n-k}$

C) $P_n(k) = C_n^k p q^{n-k}$

D) $P_n(k) = C_n^k p^k q^{n-k}$

E) $P_n(k) = C_n^k p^{k-n} q^{n-k}$

5. The discrete random variable is set. To find a mathematical expectation

X	2	4	5	6	8
p	0,1	0,2	0,3	0,1	0,3

A) M(X)=1

B) M(X)=5,5

C) M(X)=6

D) M(X)=5

E) M(X)=2

6. To find Bayes's formula

A) $P_{B_i}(A) = P(A) + P_A(B_i)$

B) $P_A(B_i) = \dfrac{P(B_i)}{P(A)}$

C) $P_B(B) = \dfrac{P(A)}{P(B_i)}$

D) $P_B(B) = \dfrac{P(A)}{P(B_i)}$

E) $P_A(B_i) = \dfrac{P_{B_i}(A)P(B_i)}{P(A)}$

7. To find a mathematical expectation of a constant

A) M(C)=0

B) M(C)=p

C) M(C)=-p

D) $M(C) = C^2$

E) M(C)=C

8. To find dispersion of a constant

A) D(C)=1

B) D(C)=-C

C) $D(C) = -C^2$

D) $D(C) = C^2$

E) D(C)=0

9. As the mathematical expectation of a continuous random variable is defined

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

B) $M(x) = \int_{-\infty}^{\infty} xf(x)dx$

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

D) $M(x) = \int_{-\infty}^{\infty} dx$

E) $M(x) = 0$

10. Is called density of distribution of a continuous random variable

A) Distribution function

B) the mathematical expectation

C) first derivative of function of distribution

D) second derivative of function of distribution

E) integral from distribution function

11.A random variable is given by the distribution function $\begin{cases} 0, x \le 1, \\ \dfrac{(x-1)^2}{4}, 1 < x \le 3, \\ 1, x > 3 \end{cases}$

distribution density is a function of

A) $f(x) = \begin{cases} 0, x \le 1, \\ \dfrac{(x-1)}{2}, 1 < x \le 3, \\ 0, x > 3 \end{cases}$

B) $f(x) = \begin{cases} 0, \ \ x < 0, \ x > 3, \\ \dfrac{x-1}{2}, \ 0 < x < 3 \end{cases}$

C) $f(x) = \begin{cases} 0, x \le 1, \\ \dfrac{(x-1)}{2}, 1 < x \le 3, \\ 1, x > 3 \end{cases}$

D) $f(x) = \begin{cases} 0, x \le 0, \\ \dfrac{(x-1)}{2}, 0 < x \le 3, \\ 1, x > 3 \end{cases}$

12. A random variable is given by the distribution function $\begin{cases} 0, x \le 1, \\ \dfrac{(x-1)^2}{4}, 1 < x \le 3, \\ 1, x > 3 \end{cases}$ the probability of getting at the interval (0, 2) is

A) 0,25
B) 0,75
C) 0,5
D) 0
E) -0,25

13. A random variable is given by the distribution function $\begin{cases} 0, x \le 2, \\ \dfrac{(x-2)^2}{4}, 2 < x \le 4, \\ 1, x > 4 \end{cases}$

The expectation is equal to

A) 3,333
B) 3
C) 4
D) 2
E) 4,5

14. A random variable is given by the distribution function $\begin{cases} 0, x \le -1, \\ \frac{(x+1)^2}{16}, -1 < x \le 3, \\ 1, x > 3 \end{cases}$

distribution density is a function of

A) $f(x) = \begin{cases} 0, x \le -1, \\ \frac{(x+1)}{8}, -1 < x \le 3, \\ 0, x > 3 \end{cases}$

B) $f(x) = \begin{cases} 0, \ x < -1, \ x > 3, \\ \frac{x+1}{8}, \ -1 < x < 3 \end{cases}$

C) $f(x) = \begin{cases} 0, x \le -1, \\ \frac{(x+1)}{8}, -1 < x \le 3, \\ 1, x > 3 \end{cases}$

D) $f(x) = \begin{cases} 0, x \le 0, \\ \frac{(x+1)}{8}, 0 < x \le 3, \\ 1, x > 3 \end{cases}$

15. A random variable has a distribution density $\begin{cases} 0, x \le 1, x > 4, \\ \frac{1}{3}, 1 < x \le 4 \end{cases}$. The dispersion is

A) 0,75
B) 1
C) 2
D) 0

16. A bus route is on schedule with an interval of 10 minutes. If we assume that *t* - waiting time at bus stop distributed evenly in the specified range, the distribution function has the form

A) $F(x) = \begin{cases} 0, \ x \le 0, \\ \frac{x}{10}, \ 0 < x \le 10, \\ 1, \ \ x > 10 \end{cases}$

B) $F(x) = \begin{cases} 0, \ x \le 0, \\ \frac{x-10}{10}, \ 0 < x \le 10, \\ 1, \ \ x > 10 \end{cases}$

C) $F(x) = \begin{cases} 0, \ x \le 0, \ \ x > 10 \\ \frac{x}{10}, \ 0 < x \le 10, \end{cases}$

D) $F(x) = \frac{x}{10}$

17. The random variable is uniformly distributed on the interval [2, 6]. The dispersion is

A) 1,3
B) 1,2

C) 0
D) 1
E) 2

18. The random variable is uniformly distributed on the interval [2, 6]. The density distribution is given by

A) $f(x) = \begin{cases} 0, x \leq 2, \; x > 6 \\ 0.25, \; 2 < x \leq 6 \end{cases}$

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

C) $f(x) = \begin{cases} 0, x \leq 2, \; x > 6 \\ \dfrac{x-2}{4}, \; 2 < x \leq 6 \end{cases}$

D) $F(x) = \begin{cases} 0, x \leq 2, \\ \dfrac{x-2}{4}, \; 2 < x \leq 6, \\ 1, \; x > 6 \end{cases}$

19. The random variable obeys the exponential distribution with parameter $\lambda = 0{,}1$. The expectation is equal to

A) 10
B) 1
C) 5
D) 100
E)25

20. The random variable obeys the exponential distribution with parameter $\lambda = 0{,}2$. Root mean square deviation is

A) 5
B) 50
C) 10
D) 0,5
E)1/5

GLOSSARY

Almost convergent. A sequence $\{X_n\}$ of random variables is called *almost convergent* to a random variable *X* if $P\left(\lim_{n\to\infty} X_n = X\right) = 1$.

Antithetical (complementation) events. Two events A_1 and A_2 are called *antithetical (complementation)* if they can't take place at the same time.

Axioms of probability theory. 1) Each random event A is associated with a number P(A), $0 \le \mathrm{P(A)} \le 1$ that is called probability *A*; 2) The probability of a reliable event is equal to1: $P(U) = 1$; 3) The axiom of additivity: if $A_1, A_2, \dots, A_n$ pairwise incompatible random events.

Bayes formula, where $P(A_i)$ is the prior probability, $P(A_i/B)$ is the posterior probability

$$P(A_i / B) = \frac{P(A_i)P(B / A_i)}{\sum_{j=1}^{n} P(A_j)P(B / A_j)}.$$

Binomial distribution. A random variable is called *binomially distributed* with parameters *n* and *p*, if possible values 0,1,2,...,*n* it takes probabilities $P(n, k)$ with the given formulas

$$P(n,k) = C_n^k p^k q^{n-k}, \quad q = 1 - p \quad (k = 0,1,\dots,n).$$

Boundary distributions. Let $F(x_1, x_2, \dots, x_n)$ be the distribution function of a random vector $(X_1, X_2, \dots, X_n)$. Then

$$P(X_{i_1} < x_{i_1}, \dots, X_{i_k} < x_{i_k}) = F(c_1, \dots, c_n) \quad (1 \le i_1 < \dots < i_k \le n),$$

where $c_{i_1} = x_{i_1}, c_{i_2} = x_{i_2}, \dots, c_{i_k} = x_{i_k}$, and all the others $c_j = +\infty$ is called *the k-dimensional boundary distribution* $F = (x_1, x_2, \dots, x_n)$.

Generating functions. If p_k is a probability function of some discrete random variable *X* of the indicated type, and *z* is a complex parameter. Then we called *the generating function* of the random variable *X:*

$$\varphi(z) = \sum_k p_k z^k.$$

Centered normal distribution. Function

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2}, \quad (a = 0, \ \sigma = 1)$$

is called *the density of a normalized and centered normal distribution*.

Classical definition of the probability of an event. $P(A) = \frac{m}{n}$, where m – the number of elementary events favorable for A; n – the number of all positive elementary events.

Continuous variable. A random variable is called *continuous* if its distribution function (*the integral distribution function*) can be represented in the form

$$F(x) = \int_{-\infty}^{x} f(t)dt.$$

Central limit theorem. If $\{X_n\}$ is a sequence of independent random variables, and if

$$Z_n = \sum_{k=1}^{n} \frac{X_k - M(X_k)}{\sqrt{\sum_{k=1}^{n} D(X_k)}}.$$

Conditional distributions. Let (X,Y) be a random vector, and F(x,y) its distribution function. Function

$$F(x/y) = \lim_{h \to 0} \frac{P(X < x, y \le Y < y+h)}{P(y \le Y < y+h)}$$

is called *the conditional distribution* of X under the condition that Y takes the value y.

Continuous random vectors. A random vector is called *continuous* if its distribution function can be represented in the form

$$F(x_1, x_2, ..., x_n) = \int_{-\infty}^{\tilde{o}_1} ... \int_{-\infty}^{x_n} f(t_1, ..., t_n) dt_1 ... dt_n.$$

Correlation coefficient. Parameter

$$\rho_{jk} = \frac{\mathrm{cov}(X_j, X_k)}{\sqrt{D(X_j)D(X_k)}} = \frac{\mathrm{cov}(X_j, X_k)}{\sigma_j \sigma_k}.$$

is called *the correlation coefficient* between X_j and X_k.

Covariance (correlation moment). The quantity $\mu_{jj} = cov(X_j, X_k)$ is called *the covariance* (correlation moment) of random variables X_j, X_k, where the values μ_{jj} are equal to the variances of the individual components: $\mu_{jj} = D(X_j) = \sigma_j^2$.

Covariance matrix (correlation). The matrix $\|\mu_{jk}\|_{j,k=1,...,n}$ is called *the covariance matrix* (*correlation*).

Characteristic functions of multidimensional random variables. The *characteristic function of the n-dimensional random variable* is the mathematical expectation of a quantity ($t_1, t_2, ..., t_n$ are the real parameters)

$$\exp\left(i\sum_{k=1}^{n} t_k X_k \right): \qquad \psi(t_1,\ldots,t_n) = M\left(\exp\left(i\sum_{k=1}^{n} t_k X_k \right)\right).$$

Chebyshev's inequality. Let X be a random variable having finite variance. Then forany $\varepsilon > 0$ the inequality

$$P(\left|X - M(X)\right| \ge \varepsilon) \le \frac{D(X)}{\varepsilon^2}.$$

Chebyshev's Theorem. If $X_1, X_2, \ldots$ is a sequence of pairwise independent random variables whose variances are uniformly bounded, i.e. $D(X_k) \le C$, for each k, then this sequence obeys to the weak law of large numbers.

Convergent in probability to a random variable. A sequence $\{X_n\}$ of random variables is called *convergent* in probability to a random variable X if for any $\varepsilon > 0$ the equality

$$\lim_{n\to\infty} P(\left|X_n - X\right| < \varepsilon) = 1.$$

Continuous random variables. A random variable is called *continuous* if its distribution function (*the integral distribution function*) can be represented in the form

$$F(x) = \int_{-\infty}^{x} f(t)dt.$$

Discrete variable. A random variable X is called *discrete* if it can take only a finite or countable set of values.

Discrete random vectors. A random vector $(X_1, X_2, \ldots, X_n)$ is called *discrete* if all its components are discrete random variables.

Dispersion properties. 1) $D(C) = 0$; 2) $D(CX) = C^2 D(X)$; 3) $D(C + X) = D(X)$; 4) $D(X + Y) = D(X) + D(Y)$.

Distribution function of a random variable. Let X be a random variable. *A distribution function* $F(x)$ of a random variable X is called a function $F(x) = P(X < x)$.

Formula of total probability of the event A_i, provided that event B has been occurred:

$$P(B) = \sum_{i=1}^{n} P(A_i) \cdot P(B / A_i)$$

Elementary Events. The possibility of excluding each other results of experience is called Elementary Events.

Empty set. The empty set is a set with no e1ements. We represent the null set with the symbol $\varnothing$.

Event. Let's conduct an experiment and where E - set of its elementary events. Each subset $A \subseteq E$ is called, where the Event A takes place only in case when there is happening one of elementary events from which A consists.

Exponential distribution. A random variable is called exponentially distributed if it has the following probability density (where λ the distribution parameter)

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & at \quad \tilde{o} \geq 0, \\ 0 & at \quad x < 0 \end{cases}.$$

Hypergeometric distribution. A random variable is called *hypergeometric distribution*, if possible values 0,1,...,n it takes with probabilities $P_{N,m}(n,k)$ defined by the formula:

$$P_{N,M}(n,k) = \frac{C_M^k C_{N-M}^{n-k}}{C_N^n} \quad (k = 0,1,\ldots,n).$$

Impossible event. The empty set $\varnothing$ doesn't contain elementary events and therefore, never occurs; such event is called *an impossible event*.

Independence of random variables. Random variables $X_1, X_2, \ldots, X_n$ are called *independent* if $F(x_1, \ldots, x_n) = F_1(x_1) \ldots F_n(x_n)$, where $F_i(x_i)$ is the distribution function of the i-th component X_i (one-dimensional boundary distribution).

i*-th initial moment of *X. The number $v_i = \sum_i x_k^i \, p_k$ in the case of absolute convergence of the series is called *the i-th initial moment* of the random variable X (or its distribution) (i=1,2,...).

i*-th central moment of *X. The number $\mu_i = \sum_k (x_k - v_1)^i \, p_k$ is called the *i-th central moment* of X.

Independent events. Two random events A and B are said to be *independent*, if the implementation of one does not affect the probability of implementing the other; i.e. if $P(A/B) = P(A)$.

Integral limit theorem de Moivre-Laplace. Let X be a binomially distributed random variable with parameters n and p. Then, uniformly with respect to a and b ($-\infty \leq a < b \leq +\infty$), comparison is done

$$\lim_{n\to\infty} P\left(a \leq \frac{X - np}{\sqrt{npq}} < b\right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx =$$

$$= \Phi_0(b) - \Phi_0(a).$$

Kolmogorov's theorem. If a sequence $\{X_n\}$ of random events that are independent from each other satisfies the condition

$$\sum_{n=1}^{\infty}\frac{D(X_n)}{n^2}<+\infty,$$

then it obeys the strong law of large numbers.

Law of large numbers. It is said that the sequence of random variables $X_1, X_2, \ldots$ obeys the weak law of large numbers if for any $\varepsilon > 0$ the equality

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{k=1}^{n}X_k-\frac{1}{n}\sum_{k=1}^{n}M(X_k)\right|<\varepsilon\right)=1.$$

Linearly correlated. Random variables X and Y are called *linearly correlated* if the regression lines are straight.

Local limit theorem de Moivre-Laplace. If the probability of the occurrence of the event A in n independent experiments is constant and equal to p $(0 < p < 1)$, then the probability $P(n,k)=C_n^k p^k q^{n-k}$ that in these experiments the event A occurs exactly k times, satisfies the relation

$$\lim_{n\to\infty}\frac{P(n,k)}{\frac{1}{\sqrt{2\pi}}\sqrt{npq}\,e^{-x^2/2}}=1, \quad \text{where} \quad x=\frac{k-np}{\sqrt{npq}}.$$

Mean square deviation. The square root of the dispersion is called *the spread*, or the standard deviation, or *the mean square deviation* $\sigma(x) = \sqrt{D(X)}$.

***n*-dimensional random vector**. A combination $(X_1, X_2, \ldots, X_n)$ of random variables is called *n-dimensional random vector*.

Noncorrelational random variables. Two random variables X and Y are called *noncorrelational,* if their correlation coefficient (their covariance) is equal to zero.

Normal distribution. A random variable is called *normal distribution* if it has following probability density (where a and σ are *the parameters of distribution*)

$$f(x)=\frac{1}{\sqrt{2\pi}\,\sigma}e^{-(x-a)^2/2\sigma^2}.$$

Random variable. The real variable, which, depending on the outcome of the experiment, i.e. depending on the case, takes different values, is called a *random variable*.

Reliable Event. As E consists of all elementary events, and in each experience surely there is one of the elementary events, thus, E takes place always; such event is called *a Reliable Event*.

Poisson distribution. A random variable is said to be *Poisson distributed* if it takes a countable set of possible values 0, 1, 2, ... with probabilities (the value λ is the *parameter of distribution,* $\lambda=np$)

$$P(k) = \frac{\lambda^k}{k!}e^{-\lambda} \qquad (k = 0,1,...).$$

Probability into a given half-open interval. The probability of random variable falls into a given half-open interval $P(a \le X < b) = F(b) - F(a)$.

Polygon of the distribution. The graphic representation of the series of the distribution of a discrete random variable is called the *polygon* of the distribution.

Regression lines. Curves in the plane *x*, *y* defined by equations $\overline{y}(x) = M(Y/X = x)$ and $\overline{x}(y) = M(X/Y = y)$ are called *regression lines Y* comparatively to *X* and *X* regarding to *Y*.

Simple event. The outcome of the experiments called the Simple Event.

Uniform distribution. A random variable is called *uniformly distributed* on [*a*,*b*] if its probability density on the [*a*,*b*] is constant, and outside [*a*,*b*] is equal to 0.

Application 1

The Table of Value Function $\varphi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$

	0	1	2	3	4	5	6	7	8	9
0,0	0,3989	3989	3989	3988	3986	3984	3982	3980	3977	3973
0,1	3970	3965	3961	3956	3951	3945	3939	3932	3925	3918
0,2	3910	3902	0894	3885	3876	3867	3857	3847	3836	3825
0,3	3814	3802	3790	3778	3765	3752	3739	3726	3712	3697
0,4	3683	3668	3652	3637	3621	3605	3589	3572	3555	3538
0,5	3521	3503	3485	3467	34448	3429	3410	3391	3372	3352
0,6	3332	3312	3292	3271	3251	3230	3209	3187	3166	3144
0,7	3123	3101	3079	3056	3034	3011	2989	2966	2943	2920
0,8	2897	2874	2827	2803	2780	2756	2756	2732	2709	2685
0,9	2661	2637	2613	2589	2565	2541	2516	2492	2468	2444
1,0	0,2420	2396	2371	2347	2323	2299	2275	2251	2227	2203
1,1	2179	2155	2131	2107	2083	2059	2036	2012	1989	1965
1,2	1942	1919	1895	1872	1849	1826	1804	1781	1758	1736
1,3	1714	1691	1609	1647	1626	1604	1582	1561	1539	1518
1,4	1497	1476	1456	1435	1415	1394	1374	1354	1334	1315
1,5	1295	1276	1257	1238	1219	1200	1182	1163	1145	1127
1,6	1109	1092	1074	1057	1040	1023	1006	0989	0973	0957
1,7	0940	0925	0909	0893	0878	0863	0848	0833	0818	0804
1,8	0709	0775	0761	0748	0734	0721	0707	0694	0681	0669
1,9	0656	0644	0632	0620	0608	0596	0584	0573	0562	0551
2,0	0,0540	0529	0519	0508	0498	0488	0478	0468	0459	0449
2,1	0440	0431	0422	0413	0404	0396	0387	0379	0371	0363
2,2	0355	0347	0339	0332	0325	0317	0310	0303	0297	0290
2,3	0283	0277	0270	0264	0258	0252	0246	0241	0235	0229
2,4	0224	0219	0213	0208	0203	0198	0194	0189	0184	0180
2,5	0175	0171	0167	0163	0158	0154	0151	0147	0143	0139
2,6	0136	0132	0129	0126	0122	0119	0116	0113	0110	0107
2,7	0104	0101	0099	0096	0093	0091	0088	0086	0084	0081
2,8	0079	0077	0075	0073	0071	0069	0067	0065	0063	0061
2,9	0060	0058	0056	0055	0053	0051	0050	0048	0047	0046
3,0	0,0044	0043	0042	0040	0039	0038	0037	0036	0035	0034
3,1	0033	0032	0031	0030	0029	0028	0027	0026	0025	0025
3,2	0024	0023	0022	0022	0021	0020	0020	0019	0018	0018
3,3	0017	0017	0016	0016	0015	0015	0014	0014	0013	0013
3,4	0012	0012	0012	0011	0011	0010	0010	0010	0009	0009
3,5	0009	0008	0008	0008	0008	0007	0007	0007	0007	0006
3,6	0006	0006	0006	0005	0005	0005	0005	0005	0005	0001
3,7	0004	0004	0004	0004	0004	0004	0003	0003	0003	0003
3,8	0003	0003	0003	0003	0003	0002	0002	0002	0002	0002
3,9	0002	0002	0002	0002	0002	0002	0002	0002	0001	0001

Application 2

The Table of Value Function $\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_0^x \exp\left(-\frac{t^2}{2}\right)dt$

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
0,00	0,0000								
0,01	0,0040	0,51	0,1950	1,01	0,3438	1,51	0,4345	2,02	04783
0,02	0,0080	0,52	0,1985	1,02	0,3461	1,52	0,4357	2,04	0,4793
0,03	0,0120	0,53	0,2019	1,03	0,3485	1,53	0,4370	2,06	0,4803
0,04	0,0160	0,54	0,2054	1,04	0,3508	1,54	0,4382	2,08	0,4812
0,05	0,0199	0,55	0,2088	1,05	0,3531	1,55	0,4394	2,10	0,4821
0,06	0,0239	0,56	0,2123	1,06	0,3554	1,56	0,4406	2,12	0,4830
0,07	0,0279	0,57	0,2157	1,07	0,3577	1,57	0,4418	2,14	0,4838
0,08	0,0319	0,58	0,2190	1,08	0,3599	1,58	0,4429	2,16	0,4846
0,09	0,0359	0,59	0,2224	1,09	0,3621	1,59	0,4441	2,18	0,4854
0,10	0,0398	0,60	0,2257	1,10	0,3643	1,60	0,4452	2,20	0,4861
0,11	0,0438	0,61	0,2291	1,11	0,3665	1,61	0,4463	2,22	0,4868
0,12	0,0478	0,62	0,2324	1,12	0,3686	1,62	0,4474	2,24	0,4875
0,13	0,0517	0,63	0,2357	1,13	0,3708	1,63	0,4484	2,26	0,4881
0,14	0,0557	0,64	0,2389	1,14	0,3729	1,64	0,4495	2,28	0,4887
0,15	0,0596	0,65	0,2422	1,15	0,3749	1,65	0,4205	2,30	0,4893
0,16	0,0636	0,66	0,2454	1,16	0,3770	1,66	0,4515	2,32	0,4898
0,17	0,0675	0,67	0,2486	1,17	0,3790	1,67	0,4525	2,34	0,4904
0,18	0,0714	0,68	0,2517	1,18	0,3810	1,68	0,4535	2,36	0,4909
0,19	0,0753	0,69	0,2549	1,19	0,3830	1,69	0,4545	2,38	0,4913
0,20	0,0793	0,70	0,2580	1,20	0,3849	1,70	0,4554	2,40	0,4918
0,21	0,0832	0,71	0,2611	1,21	0,3869	1,71	0,4564	2,42	0,4922
0,22	0,0871	0,72	0,2642	1,22	0,3883	1,72	0,4573	2,44	0,4927
0,23	0,0910	0,73	0,2673	1,23	0,3907	1,73	0,4582	2,46	0,4931
0,24	0,0948	0,74	0,2703	1,24	0,3925	1,74	0,4591	2,48	0,4934
0,25	0,0987	0,75	0,2734	1,25	0,3944	1,75	0,4599	2,50	0,4938
0,26	0,126	0,76	0,2764	1,26	0,3962	1,76	0,4608	2,52	0,4941
0,27	0,1064	0,77	0,2794	1,27	0,3980	1,77	0,4616	2,54	0,4945
0,28	0,1103	0,78	0,2823	1,28	0,3997	1,78	0,4625	2,56	0,4948
0,29	0,1141	0,79	0,2852	1,29	0,4015	1,79	0,4633	2,58	0,4951
0,30	0,1179	0,80	0,2881	1,30	0,4032	1,80	0,4641	2,60	0,4953
0,31	0,1217	0,81	0,2910	1,31	0,4049	1,81	0,4649	2,62	0,4956
0,32	0,1255	0,82	0,2939	1,32	0,4066	1,82	0,4656	2,64	0,4959

0,33	0,1293	0,83	0,2967	1,33	0,4082	1,83	0,4664	2,66	0,4961
0,34	0,1331	0,84	0,2995	1,34	0,4099	1,84	0,4671	2,68	0,4963
0,35	0,1368	0,85	0,3023	1,35	0,4115	1,85	0,4678	2,70	0,4965
0,36	0,1406	0,86	0,3051	1,35	0,4131	1,86	0,4686	2,72	0,4967
0,37	0,1443	0,87	0,3078	1,37	0,4147	1,87	0,4693	2,74	0,4969
0,38	0,1480	0,88	0,3106	1,38	0,4162	1,88	0,4699	2,76	0,4971
0,39	0,1517	0,89	0,3133	1,39	0,4177	1,89	0,4706	2,78	0,4973
0,40	0,1554	0,90	0,3159	1,40	0,4192	1,90	0,4713	2,80	0,4974
0,41	0,1591	0,91	0,3186	1,41	0,4207	1,91	0,4719	2,82	0,4976
0,42	0,1628	0,92	0,3212	1,42	0,4222	1,92	0,4726	2,84	0,4977
0,43	0,1664	0,93	0,3238	1,43	0,4236	1,93	0,1732	2,86	0,4979
0,44	0,1700	0,94	0,3264	1,44	0,4251	1,94	0,4738	2,88	0,4980
0,45	0,1736	0,95	0,3289	1,45	0,4265	1,95	0,4744	2,90	0,4981
0,46	0,1772	0,96	0,3315	1,46	0,4279	1,96	0,4750	2,92	0,4982
0,47	0,1808	0,97	0,3340	1,47	0,4292	1,97	0,4756	2,94	0,4984
0,48	0,1844	0,98	0,3365	1,48	0,4306	1,98	0,4761	2,96	0,4985
0,49	0,1879	0,99	0,3389	1,49	0,4319	1,99	0,4767	2,98	0,4986
0,50	0,1915	1,00	0,3413	1,50	0,4332	2,00	0,4772	3,00	0,49865
						3,2	0,49931	3,8	0,499928
						3,4	0,49966	4,00	0,499968
						3,6	0,499841	4,50	0,499997
								5,00	0,499997

REFERENCES:

1. Гмурман В.Е. Теория вероятностей и математическая статистика. М.: Высшая школа, 2013.

2. Вентцель Е.С. Задачи и упражнения по теория вероятностей. – М.: Высшая школа, 2010

3. Гусак А.А. Вышая математика. Минск: Театра Системс, 2007, ч. 2.

4. Гмурман В.Е. Руководство к решению задач по теории вероятностей и математической статистике. М.: Высшая школа, 2003.

5. Лунгу К.Н. и др. Сборник задач по высшей математике. 2 курс – М.: Айрис-пресс, 2011. – 576 с.

6. Khasseinov K. Canons of mathematics. - M.: Nauka, 2007. - 591 p.

7. Письменный Д. Конспект лекций по теории вероятностей и математической статистике, случайные процессы. -М.: Айрис-Пресс, 2006.- 288 с.

8. Кремер Н.Ш. Теория вероятностей и математическая статистика. - М.: ЮНИТИ-ДАНА, 2004.

9. General course of higher mathematics for economists. Textbook edited by V. I. Ermakov. – M.: INFRA, 1999.

10. Kremer N. Sh. Higher mathematics for economists: Textbook for higher educational institutions. – M., 1997.

11. Kremer H. Sh. Mathematical statistics. –M.: Ekologicheskoe obrazovanie, 1992.

12. Bugrov Ya. S., Nikolskii S. M. Calculus. –M.: Nauka, 1988.

CONTENTS

Iskakova Akzholtay Kurmantaevna

Kokazhaeva Amangul Basarbekobna

Baisalova Manshyk Zhymamyratovna

PROBABILITY THEORY LECTURES

Signed in print 15.11.2020
Circulation ___ copies
volume 6,0 ed.sheets.
Price ____

Format 60x84 1/16
Typographical paper №1
Ordering_____ tg.

Multiple copying Office «Kizdar university» of
Kazakh National Women's Pedagogical University
050000, Almaty c., 116, Gogol st.