Contents

1 Introduction to Probability Theory **2**

1.1 Random Experiment. Sample Space. Random Events. 2

1.2 Operations with Events 4

1.3 Collectively Exhaustive Events 7

2 Probability. Different Definitions. Probability Properties **8**

2.1 Classical Definition of Probability 8

2.2 Probability Properties 9

2.3 Definition of Probability in Case of Finite Number of Outcomes That are Not Equally Likely 11

2.4 Geometric Probability 12

2.5 Conditional Probabilities and Bayes' Formula 14

2.6 Calculating Conditional Probability According to Classical Definition 15

2.7 Independent Events 16

2.8 Total Probability Formula 16

2.9 Bernoulli Process and Formula 18

3 Random Variables **19**

3.1 Discrete Random Variable 19

3.2 Random Variable Distribution 20

3.3 Distribution Function of Discrete Random Variable 21

3.4 Independence Between Two Discrete Random Variables 22

3.5 Numerical Characteristics of Discrete Random Variables 23

3.5.1 Expected Value 23

3.5.2 Properties of Expected Value 25

3.5.3 Variance 26

3.5.4 Properties of Variance 27

1 Introduction to Probability Theory

We are ruled by chance. This famous philosophical statement has probably summed up the reflections of many generations on the subject of random events that happen all the time. How to live and survive in the world of randomness? Are we able to find a pattern in any perceived randomness? If so, how to use this knowledge wisely? These questions made people follow what's happening and gather knowledge of random events and their chances of happening, in other words, of the event probabilities.

Let's introduce the main concepts and ideas of a mathematical model used to study random events in the probability theory. This exact model allows analyzing and generalizing the data collected as a result of observations. As common in math papers, let's introduce the fundamental concepts using strict definitions. Note that the naturality of these definitions is straightforward.

1.1 Random Experiment. Sample Space. Random Events.

Определение 1.1.1 *A random experiment is an experiment that produces a result from the set of all possible outcomes. However, the exact result is unknown before the experiment.*

Определение 1.1.2 *Different disjoint (those that cannot happen at the same time) results of the experiment are called elementary events or outcomes.*

Определение 1.1.3 *The sample space of an experiment is a set of all possible outcomes or results of that experiment. The sample space is designated by big omega.*

Here are some examples of random experiments and their sample spaces:

1. Let the random experiment be the toss of a coin. Then, sample space big omega is heads and tails.

2. Let the random experiment be the toss of a die. Then, sample space big omega consists of six elementary events (outcomes): one, two, and so on, six.

3. In the same way, we can think of an exam as an experiment. Big omega consists of the elementary events: 2, 3, 4, 5. Obviously, such examples are limitless. What they have in common is that the number of possible outcomes in each experiment is finite. However, it is not always true. Here's a classical example of an infinite sample space.

4. Let's liken the thread to a line segment that starts and ends at 0 and 1 respectively. Then, breaking point x lies on the line segment. Therefore, big omega is the entire closed line segment from 0 to 1 [0,1], or the infinite set.

Определение 1.1.4 *A random event is any subset of a sample space.*

In other words, a random event is any set of the elementary events of the given experiment.

Events are usually designated by capital Latin letters: $A, B, C, \ldots$ If necessary, they are written in words.

Определение 1.1.5 *We say that event A has occurred if the outcome in the experiment is in A.*

Пример 1.1.1 *Let the experiment be the toss of a die. Then, sample space big omega consists of elements 1, 2, ..., 6, each corresponding one side of the die. According to the introduced definition (and the sample space), event A of getting an odd number consists of 3 elements 1, 3, 5. The event B of getting 2 has only one outcome:* 2.

Let's assume that the outcome in the experiment is 1. *Then, event A has occurred because elementary event* 1 *is in A, and event B has not occurred because the equivalent elementary event is not in B.*

Now, suppose that we are getting 2. *Then, event B has occurred, but A hasn't.*

Определение 1.1.6 *Event big omega consisting of all elementary events of an experiment is called a sure event.*

A sure event is an event that is going to happen, no matter what.

Определение 1.1.7 *An impossible event is an event that does not contain any of the possible outcomes. It is an empty set of outcomes.*

Such an event never occurs as a result of the experiment. An impossible event is designated by $\varnothing$. Its symbol is the same as the one of an empty set.

Пример 1.1.2 *A die is tossed. The event of getting less than 7 that consists of the elementary events 1, 2, ..., 6 and does not differ from sample space big omega, is a sure event. The event of getting a number greater than or equal to 7 does not include elementary events. Therefore, it is an impossible event in the experiment.*

Определение 1.1.8 *Let A and B be the subsets of big omega or, what is the same, let A and B be the events. A is a special case of B if A is a subset of B.*

It is also said that A leads to B. It means that if A occurs, B also occurs. The pictorial representation of events A and B is shown on the screen.

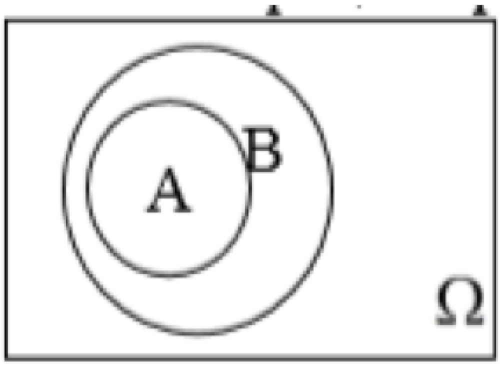

Figure 1: A is a subset of B

Пример 1.1.3 *A die is tossed. There are three events. Event A is getting 1. Event B is getting an odd number. Event C is getting 3 or 5. Obviously, A and C lead to B.*

1.2 Operations with Events

Next, to create compound events using simple events, we will consider operations with events. It will allow us to compute the probabilities of compound events based on the probabilities of simple events according to certain rules. Besides, when considering events A and B, we also want to consider such events as or A, or B; both A and B; not A; A, but not B; and so on.

Определение 1.2.1 *The union of events A and B is the event consisting of all outcomes that are either in A or in B.*

The union can be written as A + B. The pictorial representation is shown on the screen.

Event A + B occurs if at least one of the events occurs. In other words, A + B occurs if only A happens, or only B happens, or both of them occur at the same time.

Определение 1.2.2 *The intersection of two events A and B is an event consisting of all outcomes that are in both A and B.*

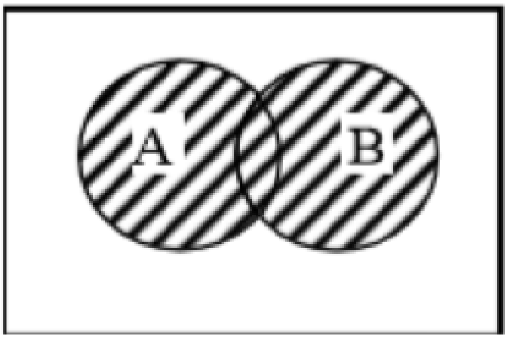

Figure 2: The union of events A and B

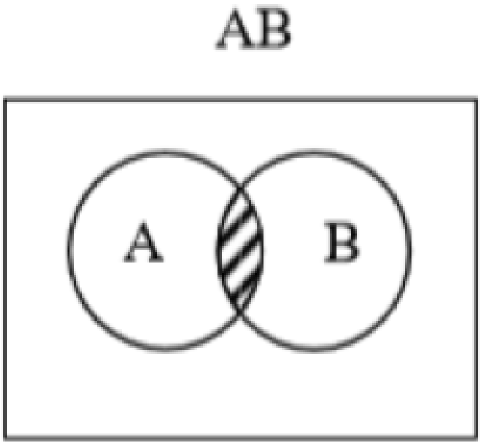

Figure 3: The intersection of events A and B

The intersection can be written as AB. The pictorial representation is shown on the screen.

AB occurs if and only if both A and B occur together.

Определение 1.2.3 *A and B are mutually exclusive (or disjoint) events if they cannot happen at the same time. In other words, their intersection is impossible.*

The pictorial representation is shown on the screen.

Определение 1.2.4 *The difference between two events A and B is the event consisting of all outcomes that are in A but not in B.*

The difference can be written as $A \backslash B$ (the difference between A and B). The pictorial representation is shown on the screen.

Event $A \backslash B$ occurs if and only if event A occurs, but event B does not occur.

Определение 1.2.5 *The absolute complement of A (the one that is the opposite of A) is the event consisting of all outcomes that are not in A.*

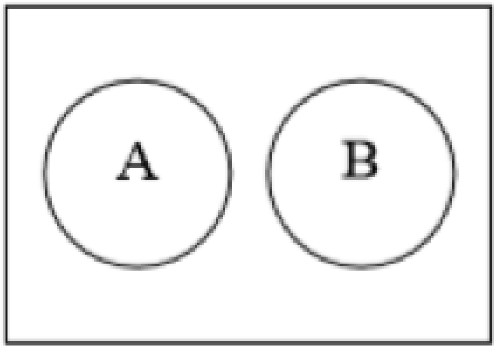

Figure 4: Disjoint events A and B

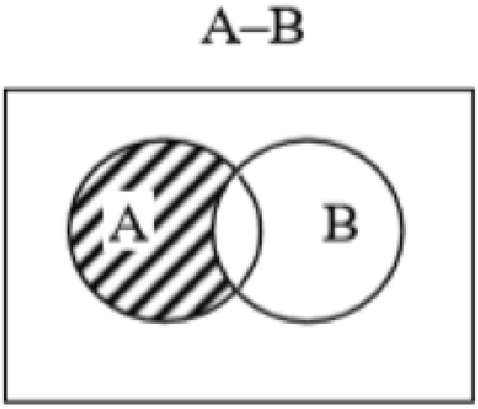

Figure 5: The difference between A and B

The absolute complement of A is denoted by [A overline] ($\overline{A}$). The pictorial representation is shown on the screen.

Event $\overline{A}$ occurs if and only if event A does not occur, that is, [A overline equals the difference between big omega and A] $\overline{A} = \Omega \backslash A$.

Пример 1.2.1 *Let the experiment be the toss of a die. Let's consider events. A is getting 1, B is getting an odd number, C is getting 3 or 5, D is getting an even number.*

1. *Let's find the union. Clearly, A plus B equals B, A plus C equals B, A plus D equals 1,2,4,6, and B plus D is a sure event.*

2. *Next, we are going to find the intersection of events. AB is an impossible event because we cannot get 1 and (3 or 5) at the same time. AD is also an impossible event because 1 is an odd number and we cannot get both an odd and an even number at the same time. 3 and 5 are odd, so intersection BC leads to event C.*

3. *Let's find the difference between the events. The difference between A and B is an impossible event because 1 is odd. A and C do not intersect. There-*

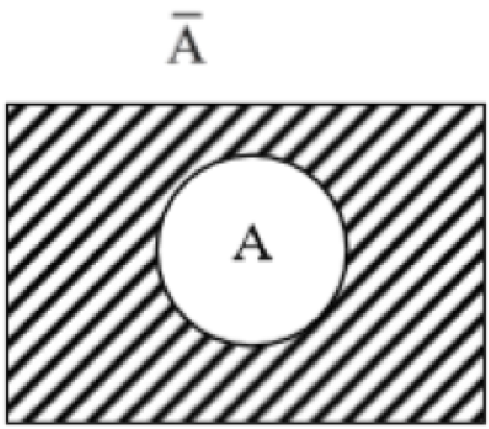

Figure 6: The absolute complement of A

fore, their difference is the same as A. The same applies to A and D. The situation with B and A is different. Their difference leads to C. In this case, A and C do not intersect either. Therefore, their complement is C.

4. *Let's find the absolute complements. The absolute complement of A means that we are getting a number other than one. This event consists of elementary events of getting 2, 3, 4, 5, 6. The absolute complement of B that is the opposite of the event of getting an odd number is the event of getting an even number, namely, event D. The absolute complement of C means that we are not getting 3 and 5, that is, we are getting either 1, or an even number. Therefore, this event consists of elementary events 1, 2, 4, 6. The opposite is also true. The opposite of D is B.*

1.3 Collectively Exhaustive Events

To sum up this section, let's introduce the concept of collectively exhaustive events.

Определение 1.3.1 *Let's consider events [big H one, big H two, and so on, big H n], which are the subsets of set big omega. Events [big H one, big H two, and so on, big H n]* $H_1, H_2, \ldots, H_n$ *are collectively exhaustive if one and only one of the events occurs as a result of the experiment (either* H_1*, or* H_2*, …, or* H_n*).*

The pictorial representation of the collectively exhaustive events given that $n = 5$ is shown on the screen.

In other words, events [big H one, big H two, and so on, big H n] are collectively exhaustive if and only if: a) Their union contains a sure event. b) They do not intersect (which means that no more than one event can occur at a given time).

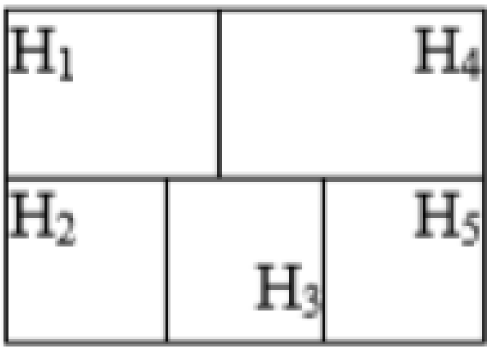

Figure 7: Collectively exhaustive events

Пример 1.3.1 *Let the experiment be the toss of a die. Events:* $A = \{getting\ 1\}$, $B = \{getting\ an\ odd\ number\}$, $C = \{getting\ 3\ or\ 5\}$. $D = \{getting\ an\ even\ number\}$.

1. *You can see that B and D are collectively exhaustive events because we can get either an even or odd number.*

2. *The events* A, C, *and* D *are collectively exhaustive because they are pairwise disjoint and their union gives big omega.*

2 Probability. Different Definitions. Probability Properties

Probability is a measure quantifying the likelihood that events will occur. Let's define the term probability that allows modeling a large number of random events.

2.1 Classical Definition of Probability

We are going to consider the experiment with a finite number of outcomes being equal to n. Assume that the outcomes are equipossible, which means they are equally likely to occur.

Определение 2.1.1 *Let event* A *consist of* m *outcomes. These* m *outcomes are called favorable to event* A.

Определение 2.1.2 *The probability* P *of* A *of getting event* A *is the ratio of the number of equally likely outcomes favorable to event* A *to the number of all possible outcomes, that is, to the ratio* m *to* n. *The notation is shown on the screen.*

Пример 2.1.1 *Let's look at the example. Let the experiment be the toss of a die. As you've already learned, big omega consists of the elementary events, namely, getting 1, getting 2, and so on, getting 6. A total of 6, therefore, $n = 6$. If we assume that the die is fair, outcomes will be equally likely. Let's find the probability of getting an odd number. We will designate this event by A. Then, A consists of the following outcomes: getting 1, getting 3, getting 5. The number of favorable events is $m = 3$. Now we can find the probability of event A as the ratio of 3 to 6, that is, 0.5.*

Пример 2.1.2 *Assume that the final task includes* 40 *task variants with linear normalization of parameters and* 60 *task variants with exponential normalization of parameters. A variant is selected at random. What is the probability of selecting a task variant with linear normalization? What is the probability of selecting a task variant with exponential normalization?*

We consider the events. A is selecting a task variant with linear normalization; B is selecting a task variant with exponential normalization. The number of all equally likely outcomes in the experiment:

$$n = 40 + 60 = 100.$$

40 *outcomes are favorable to A,* 60 *are favorable to B. Therefore,*

$$\mathsf{P}(A) = \frac{40}{100} = 0.4,$$

$$\mathsf{P}(B) = \frac{60}{100} = 0.6.$$

2.2 Probability Properties

The introduced concept of probability has many properties that are useful for solving tasks in practice. Let's look closely at them. We will formulate them in the following sequence. First, three main properties, then the rest of them that are easily proved based on the first three.

Let [big omega] Ω be a sample space of an experiment, and A and B the events.

1. The probability of the sure event is 1.

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

2. The probability of any event is non-negative.

$$\mathsf{P}(A) \geq 0.$$

3. If A and B are disjoint ($A \cdot B = \varnothing$), the probability of the union of disjoint events is the sum of their probabilities.

$$\mathsf{P}(A + B) = \mathsf{P}(A) + \mathsf{P}(B).$$

   The proof of these properties follows from the provided definition of probability.

   Let's formulate the rest properties.

4. If event A leads to B ($A \subset B$), the probability of A is not greater than the probability of B.
$$\mathsf{P}(A) \le \mathsf{P}(B).$$

5. The probability of any event lies between 0 and 1.

$$0 \le \mathsf{P}(A) \le 1.$$

6. The probability of the complement of A equals 1 minus the probability of event A.
$$\mathsf{P}(\overline{A}) = 1 - \mathsf{P}(A).$$

7. The probability of an impossible event is zero.

$$\mathsf{P}(\varnothing) = 0.$$

8. For any A and B (possibly joint), the probability of their sum is the sum of their probabilities minus the probability of their product.

$$\mathsf{P}(A + B) = \mathsf{P}(A) + \mathsf{P}(B) - \mathsf{P}(AB).$$

   Let's explain the last property. Let's liken A and B to some regions in the plane. Their areas are respectively equal to the probabilities of A and B. To find the union area, we calculate the sum of the areas of A and B, and their intersection area is included in the sum twice. Therefore, one such area should be excluded from this sum.

The last statement is often termed an addition theorem on probability. You can see that it's a generalization of the third main probability property.

Here's an important note.

Замечание 2.2.1 *Since collectively exhaustive events are pairwise disjoint and their union equals a sure event, the sum of probabilities of collectively exhaustive events equals one. It allows considering collectively exhaustive events as elementary ones (recall that only one collectively exhaustive event can ever occur as a result of the experiment).*

Let's consider several tasks that can be solved using probability properties.

Пример 2.2.1 *A bookcase contains* 6 *programming books and* 10 *statistics books. Among them,* 7 *statistics books and* 4 *programming books are in Russian (the rest of them in English).*

	Language: Russian	*Language: English*	*Total*
Statistics	7	3	10
Programming	4	2	6
Total	11	5	16

If a book is chosen at random, what is the chance that it is a statistics book or a book in English?

Assume that A is choosing a statistics book, and B is choosing a book in English. The task is to find the probability of the union of these events. The events are joint, because, for example, some statistics books are in English. Therefore, we will turn to the addition theorem on probability shown on the screen. All the probabilities on the right side of the formula can be found using the classical definition and the given table. The total number of outcomes in the experiment equals the number of books, that is, 16. *The number of favorable events for statistics books is* 10, *for books in English* 5, *for statistics books in English* 3. *Hence, the probability of the union is ten-sixteenth plus five-sixteenth minus three-sixteenth. The result is 0.75.*

$$\mathsf{P}(A+B) = \frac{10}{16} + \frac{5}{16} - \frac{3}{16} = \frac{12}{16} = 0.75.$$

Thus, the chance that a random book is a statistics book or a book in English equals 0.75

2.3 Definition of Probability in Case of Finite Number of Outcomes That are Not Equally Likely

Despite the popularity of the classical model, it is not always applicable. For example, it is not reasonable to assume that the probabilities of meeting and not meeting an elephant in a city are equally likely. Let's exclude the condition requiring that elementary events should be equally likely. This condition is necessary for the classical definition.

Let's consider a random experiment with n outcomes and sample space big omega consisting of small omega 1, small omega 2, and so on, small omega n ($\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$). Assume that each omega i (ω_i) has been assigned the probability [p of omega i] $\mathsf{P}(\omega_i)$ based on some considerations. Thus, all [i equal to 1, 2, and so on, n] $i = 1, 2, \ldots, n$ have been assigned such non-negative values [p of omega i] $\mathsf{P}(\omega_i)$ that their sum is equal to 1.

$$\sum_{i=1}^{n} \mathsf{P}(\omega_i) = 1.$$

Определение 2.3.1 *The probability of event A is number [p of A] ($\mathsf{P}(A)$) that is equal to the sum of the probabilities of all outcomes in A.*

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

We can easily prove that the three main probability properties are satisfied. Therefore, other properties are also satisfied.

Пример 2.3.1 *A student takes a test. Possible outcomes are all the grades. The sample space consists of four elements, namely, getting 2, getting 3, getting 4, or getting 5)* $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\} = \{2, 3, 4, 5\}$. *It is not a first year for that student, and the grading history allows making some assumptions about probable results. Assume that we know the following probabilities of the elementary events for that student: [p of omega 1] or the probability that the student will get a bad grade (that is, 2) equals 0.02, and so on. You can see that the sum of these numbers equals one. Therefore, we can work within the framework of the introduced definition.*

We need to determine the probability of the following events: The student will pass the exam and get a good mark, 4 *or* 5. *The student will pass the exam.*

Let's consider the events. A is getting 4 *or* 5 *that equals omega 3, omega4; B is passing the exam = omega2, omega3, omega4. According to the introduced definition, the probability of event A is the sum of the probabilities of the elementary events omega 3 and omega 4 and is equal to 0.9. The probability of event B is the sum of the probabilities of the elementary events omega 2, omega 3, and omega 4 and is equal to 0.98.*

2.4 Geometric Probability

The discussed definitions of probability assume that a sample space is finite. Let's generalize the concept of probability for such a case when a sample space is a limited region for which a measure has been defined. A measure can be, for example, the length (for a straight line), area (for the plane), or volume (for a space region). The result of the experiment will be a random point within the region. In that case, events are different subsets that lie within that region and have a measure.

Note that the geometric interpretation allows illustrating the operations with events using Venn diagrams (what we've already done).

As in the classic scheme, we assume that any point in the given region has the same chance of being chosen. From the geometric standpoint, we will interpret equipossibility as follows. Chances to choose points within the regions are equally likely. Since a random event is a subset of the sample space (with a measure), it is natural to require the subset to have a corresponding measure. Let's look at the concrete examples.

Пример 2.4.1 *In the coordinate plane, we select a random point with coordinates x and y within the square with the side of 2. We consider event A, that is, choosing a point for which the sum of its squared coordinates is less than or equal to one. In the example, sample space [big omega] Ω is a set of points of the square with the side of 2, and its measure is the area of the square, that is, 4.*

$$S(\Omega) = 2^2 = 4.$$

A measure of event A approximately equals 3.14, that is, the area of the incircle with a unit side length.

$$S(A) = \pi r^2 \approx 3.14.$$

In order not to designate a measure on the straight line (the length) of set A by $l(A)$, a measure in the plane (the area) of set by $S(A)$, a measure of a space region (volume) of set A by $V(A)$, we will always write [lambda of A] $\lambda(A)$.

So, let sample space [big omega] Ω be a limited region, and [lambda of big omega] $\lambda(\Omega)$ its measure. We can consider the experiment of choosing a point from [big omega] Ω at random. Let's determine the probability of random event that [A is in big omega] $A \subset \Omega$ that will occur as a result of this trial if point [small omega is in A] $\omega \in A$ is chosen at random. Assume that event A has the corresponding measure: $\lambda(A)$.

Определение 2.4.1 *Probability [p of a] $\mathsf{P}(A)$ of event A is the ratio of the measure of event A to the measure of sample space [big omega] Ω:*

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

It is a geometrical definition of probability. Let's compute the geometric probability of the event described in the given example. To find the geometric probability of the event of getting inside the circle as a result of the experiment that consists of throwing the point into the square, we need to divide the circle area by the square area. According to the numerical data, this ratio approximately equals 0.785.

Let's take the area as an example. It follows from the measure properties that, first, [P of big omega equals 1] $\mathsf{P}(\Omega) = 1$, second, [P of A is greater than or equal to 0] $\mathsf{P}(A) \geq 0$ because the ratio of the areas cannot be negative, and third, for disjoint A, B, the shown expression is true.

$$\mathsf{P}(A+B) = \frac{\lambda(A+B)}{\lambda(\Omega)} = \frac{\lambda(A)+\lambda(B)}{\lambda(\Omega)} = \frac{\lambda(A)}{\lambda(\Omega)} + \frac{\lambda(B)}{\lambda(\Omega)} = \mathsf{P}(A) + \mathsf{P}(B).$$

So, the first three probability properties are satisfied. As has been noted, it means that the other properties are satisfied too.

Geometric probability is used to solve the tasks of different nature. The only thing that matters is to properly build a geometric model. Let's look at the example that is a variation of a well-known meeting problem.

Пример 2.4.2 *Alex and John decide to meet somewhere between 5 p.m. and 6 p.m. They cannot text or call each other. Whoever arrives first will not wait for the other for more than 30 minutes. In any case, each should go when it's 6 p.m. What is the probability that they will meet?*

Let x be the time of Alex's arrival, and y of John's. If we place the origin at point $(17, 17)$, the sample space is a set of points of the square with vertices (0,0), (1,0), (0,1), (1,1) or a Cartesian product of the line segment with itself (the line segment starts and ends at 0 and 1).

Then, event A (meeting) is a set of points from [big omega] Ω whose coordinates (x, y) satisfy the condition [the modulus of the difference between x and y that is less than or equal to one-half] $|x-y| \leq 1/2$. From the geometric standpoint, these are the points of the square $[0,1] \times [0,1]$ that lie between the straight lines [y equals x plus one-half] $y = x + 1/2$ and [y equals x minus one-half] $y = x - 1/2$. The area of the region corresponding to event A will be equal to the difference between the area of the square (with the side of 1) and the sum of areas of two right triangles with legs of 0.5. Since the square area equals 1, the area of the desired set equals the difference between 1 and 0.25, that is, 0.75

So, the probability equals 0.75.

2.5 Conditional Probabilities and Bayes' Formula

Sometimes we have to consider event probabilities based on a part of the sample space instead of all of it. This is the case when we know that a certain event has occurred. Due to that, the conditions of the next trial change along with the sample space. For example, a basket contains balls of different colors. If a ball of a particular color is picked up, the number of balls and the numbers of colored balls in the basket will change. Therefore, the conditions of each drawing will not be the same.

Определение 2.5.1 *Let B be some event, and [P of B is not zero]* $\mathsf{P}(B) \neq 0$*. The conditional probability of event A under condition B is written as P of A the vertical bar B (*$\mathsf{P}(A|B)$*) and defined as the ratio of the probability of AB to the probability of B.*

We can also describe it as the conditional probability of A given that B has occurred. We can also say: the probability of A given B.

The formula for the probability of the intersection of events follows from the definition of the conditional probability (given that the conditional probability has been defined):

$$\mathsf{P}(AB) = \mathsf{P}(A|B)\mathsf{P}(B).$$

Замечание 2.5.1 *The multiplication theorem on probability can be extended to* n *factors:*

$$\mathsf{P}\left(A_1A_2A_3 \cdot \ldots \cdot A_{n-1}A_n\right) = \mathsf{P}\left(A_1\right)\mathsf{P}\left(A_2|A_1\right)\mathsf{P}\left(A_3|A_1A_2\right)\cdot\ldots\cdot\mathsf{P}\left(A_n|A_1A_2\ldots A_{n-1}\right).$$

It can be easily proved by the induction on the number of factors.

2.6 Calculating Conditional Probability According to Classical Definition

Let's consider a random experiment with n equipossible outcomes. Let m outcomes be favorable to event B, and k outcomes be favorable to event AB. Then,

$$\mathsf{P}(B) = \frac{m}{n}, \mathsf{P}(AB) = \frac{k}{n},$$
$$\mathsf{P}(A|B) = \frac{\mathsf{P}(AB)}{\mathsf{P}(B)} = \frac{k}{m}.$$

To solve problems with equipossible outcomes, we can calculate the conditional probability by the formula shown on the screen:

$$\mathsf{P}(A|B) = \frac{k}{m},$$

where k is the number of outcomes favorable to AB, and m is the number of outcomes favorable to B.

Замечание 2.6.1 *The conditional probability [P of A given B]* $\mathsf{P}(A|B)$ *is the same as the probability defined by new sample space [omega prime] coinciding with B with new events [A prime] equal to the product of A and B. All probability properties are preserved.*

Пример 2.6.1 *A die is tossed. Let's consider the events. A is getting 3; B is getting an odd number. We need to find the probability of A given B (*$\mathsf{P}(A|B)$*).*

Since the outcomes are equipossible, we can use the formula for the conditional probability of equipossible events. The number of outcomes favorable to B is 3*. The number of outcomes favorable to the product of* AB *is* 1*. Hence, the probability of A given B equals 1/3 (*$\mathsf{P}(A|B) = 1/3$*).*

2.7 Independent Events

Intuitively, two events are independent if one event does not influence the probability of another event. Let's formally define it.

Let [big omega] Ω be a sample space, and A and B be the members of [big omega] $A \subset \Omega, B \subset \Omega$.

Определение 2.7.1 *A and B are independent if the probability of them both is the product of the probabilities of each occurring.*

It's interesting to compare probability multiplication formulas for dependent and independent events.

This comparison leads us to the conclusion that, for independent events, the probability of event A given B equals the probability of event A. The same applies to B.

The obtained results agree with the expected ones. Let's prove that independent events exist in reality. Let's consider a deck of cards problem.

Пример 2.7.1 *There are 52 cards in a deck of cards. We pull one card. Let's consider two events. A is pulling an ace, and B is drawing an odd-numbered card. We need to prove that A and B are independent events.*

The total number of outcomes is 52*. The number of outcomes favorable to A is* 4*, for B* 13*, and for the intersection of* AB 1*. Then, the probability of event A equals [4 over 52]* $\frac{4}{52}$*, of event B [13 over 52]* $\frac{13}{52}$*, of event AB [1 over 52]* $\frac{1}{52}$*.*

The calculations show that the probability of both events is the product of the probabilities of each occurring. Therefore, A and B are independent.

2.8 Total Probability Formula

There could be a case when the probability of event A is unknown but the conditional probabilities of event A are known or easy to calculate given that events [H i] have occurred and that events [H i] are collectively exhaustive. Moreover, the probabilities of events [H i] are known for all H_i of some collectively exhaustive events H_i. We also know the probabilities of events [H i] $\mathsf{P}(H_i)$ for all i. In this case, we can calculate [the probability of A] $\mathsf{P}(A)$ by the formula shown on the screen:

The probability of event A is a sum of products of the probabilities of events [H i] and the conditional probability of A given [H i].

$H_1, H_2, \dots, H_n$ are collectively exhaustive events, and the probability of [H i] is not zero for all i from 1 to n.

Note that collectively exhaustive events are said to be hypotheses, and their probabilities are called prior probabilities of hypotheses. It follows from the definition of collectively exhaustive events and probability properties that a sum of prior probabilities of hypotheses equals one. Hence, it is always as shown on the screen.

$$\sum_{i=1}^{n} \mathsf{P}\left(H_i\right) = 1.$$

The conditional probability of A given [H i] $\mathsf{P}(A|H_i)$ in the total probability formula is said to be the probability of event A given that hypothesis H_i is true.

The total probability formula helps to solve many problems related to finding probabilities. Let's prove it with concrete examples.

Пример 2.8.1 *A school purchased computer monitors.* 45% *of the computer monitors are made by the first manufacturer,* 30% *by second, and* 25% *by third. The probability of producing a defective item for the first manufacturer is* 0.05*, for second,* 0.01*, and for third,* 0.04*. We take one monitor at random. We need to find the probability that this monitor is defective.*

Let's consider the events.

$$A = \{\textit{the monitor is defective}\},$$

$$H_1 = \{\textit{the monitor is made by the first manufacturer}\},$$

$$H_2 = \{\textit{the monitor is made by the second manufacturer}\},$$

$$H_3 = \{\textit{the monitor is made by the third manufacturer}\}.$$

*We know the percentage of purchased monitors to manufacturers (*45%, 30%*, and* 25%*). They allow us to calculate prior probabilities of hypotheses H_1, H_2, and H_3. Take a look at the screen.*

$$\mathsf{P}(H_1) = 0.45, \mathsf{P}(H_2) = 0.30, \mathsf{P}(H_3) = 0.25,$$

The probability of selecting a defective monitor given that it is made by a specific manufacturer coincides with the defect rate declared by the manufacturers. The conditional probabilities are shown on the screen.

Next, we can use the total probability formula to find that the probability of event A equals 0.355:

2.9 Bernoulli Process and Formula

Определение 2.9.1 *The Bernoulli process is a sequence of trials for which the following conditions are met:*

- *Each trial has exactly two outcomes, success and failure.*
- *The outcome of one trial does not influence the outcome of another trial.*
- *The probability of success is the same every time the experiment is conducted.*

We can consider the example of a series of observations (trials) described by the Bernoulli process.

Пример 2.9.1 *The experiment is rolling a die three times. Events: A is getting 6, and [A overline] is getting a number other than 6. The outcome favorable to event A will be a success, and outcome favorable to event [A overline]* $\overline{A}$ *will be a failure. Then, the probability of A equals one-sixth (*$\mathsf{P}(A) = \frac{1}{6}$*). All three conditions of the Bernoulli process are satisfied. Thus, we have* 3 *trials with the probability of success equal to one-sixth* $\mathsf{p} = \frac{1}{6}$.

The Bernoulli process is particularly interesting because it allows us to determine the probability of m successful events in a sequence of n trials. To determine the probability, we use the so-called Bernoulli formula in this case.

We will consider a Bernoulli process that includes n trials with the probability of success p. The probability of m successful events in a sequence of n trials will be written as [p n of m] $\mathsf{P}_n(m)$. It can be calculated by the Bernoulli formula:

$$\mathsf{P}_n(m) = C_n^m \mathsf{p}^m \mathsf{q}^{n-m},$$

where n is the number of trials, m is the number of successes in n trials, p is the probability of a successful event in a separate trial, and [q equals 1 minus p] $\mathsf{q} = 1 - \mathsf{p}$.

n choose m (C_n^m) is the number of combinations of n that can be taken from m. It's calculated as the ratio of n factorial to m factorial and to n minus m factorial:

n factorial is the product of all natural numbers from 1 to n. Zero factorial is assumed to be one.

Пример 2.9.2 *The experiment is tossing a fair coin three times. The task is to estimate the probability of getting heads 2 times. Let event A mean that a head is obtained when tossing once. Then the complement is tails.*

The probability of event A is one-half:

We will interpret the outcome favorable to event A as a success, and to event [A overline] $\overline{A}$ as a failure. We believe that the coin is fair, and the probability of a successful event is one-half. $\mathsf{P}(A) = \frac{1}{2}$. *So, here we have a Bernoulli process consisting of* 3 *trials with the probability of a success being one-half ($\frac{1}{2}$) in each of them. We need to find the probability of 2 successes in a series of 3 trials ($\mathsf{P}_3(2)$). According to the Bernoulli formula, the probability equals 3 choose 2 multiplied by squared p by q raised to the power of 3 minus 2 and equals three-eighths.*

So, the probability of getting heads twice when tossing a fair coin three times equals three-eighths ($\frac{3}{8}$).

$\mathsf{P}(A) = 0.6.$
0.31

3 Random Variables

The concept of a random variable is one of the basic concepts of probability theory. Without it, we cannot correctly formulate problems and describe methods of mathematical statistics that form the basis for the applied statistics that are collecting, processing, and analyzing data. Later, we will become acquainted with the concept of a random variable, but first, we need to find out what discrete random variables are. Skipping ahead, we would like to note that random variables are often used in statistics to interpret collected quantitative data.

3.1 Discrete Random Variable

Let's consider a random experiment. Assume that its sample space is finite, that is, big omega (omega) equals omega (omega) 1, omega 2, and so on, omega n. ($\Omega = \{\omega_1, \omega_2, \omega_3, \ldots, \omega_n\}$).

Numerical function [ksi] ξ is defined on this sample space. Since its arguments are the outcomes of the random experiment, the values will also be random. In that sense, [ksi] ξ is a random variable because the exact value it may take cannot be predicted. Numbers [x i] that are equal to [ksi of omega i] $x_i = \xi(\omega_i)$ will be called the realizations of [ksi] ξ. Since, according to the function definition, each [omega 1] ω_i corresponds to the only one [ksi of omega i] $\xi(\omega_i)$, the set of values of this function will also be either finite or countable.

Определение 3.1.1 *A random variable is a function defined on a sample space whose outcomes are numeric values.*

Определение 3.1.2 *A random variable is discrete if it is taking a finite or countable set of values.*

Замечание 3.1.1 *Note that random variables are often designated by small Greek letters such as ksi, eta, and so on ($\xi, \eta, \ldots$).*

Пример 3.1.1 *If we are tossing a die, a random variable can be the scores on the die. If we are tossing two dice, the sum of the scores on the dice is also a random variable. A random variable can be the duration of a random conversation (in minutes), the number of coins in the pocket of a stranger, and many more.*

3.2 Random Variable Distribution

To describe a random variable, we not only need to list its possible values but also to determine probabilities that the random variable will take these values. Let's consider an exam consisting of three questions. In this case, a random variable is the number of questions that a student has answered correctly. All the answers can be correct, or only two of them or one can be correct. It is also possible that the student will fail to answer any of the questions. Thus, a random variable can take values $0, 1, 2, 3$. For all students taking an exam, the set of values is the same. However, the probability of the correct answers depends on how well the student studied, and it will vary for different students.

Определение 3.2.1 *Let [ksi] ξ be a discrete random variable. The relation that establishes a connection between the possible values of a random variable and the corresponding probabilities is called the law of the distribution of random variable [ksi] ξ or simply the distribution of random variable [ksi] ξ.*

The distribution of a random variable is often written in the form of a table. Let [x 1, x 2, and so on, x n] $x_1, x_2, \ldots, x_n$ be all the possible values of random variable [ksi] ξ, and [p i] be the probabilities that [ksi] takes value [x i] ($\mathsf{p}_i = \mathsf{P}(\xi = x_i)$) where [i goes from 1 to n] ($(i = 1, 2, \ldots, n)$).

Then distribution [ksi] ξ can be written in the form of the table you see on the screen. The first row contains values of random variable ksi, and the second row includes the probabilities of these values.

ξ	x_1	$\ldots$	x_n
P	p_1	$\ldots$	p_n

Замечание 3.2.1 *The notation [p of ksi that equals x i] $\mathsf{P}(\xi = x_i)$ is used to write that the probability of the event whose elements are those and only those elementary events in which random variable [ksi] ξ takes values [x i] x_i. Since the first row includes all the possible values of the random variable and these*

values are mutually exclusive, then the events whose probabilities are the probabilities that the random variable will take values [x i] ($\mathsf{P}(\xi = x_i)$) *are collectively exhaustive. Therefore, the sum p i from 1 to n equals 1.*

$$\sum_{i=1}^{n} \mathsf{p}_i = 1.$$

Пример 3.2.1 *Assume a box contains 1 ruble, 2 rubles, 5 rubles, and 10 rubles. How much money can we get if we take one coin from a box at random? Let's write the distribution for this random variable. It takes all its values with the same probability that is equal to one-fourth.*

$$\begin{array}{c|c|c|c|c} \xi & 1 & 2 & 5 & 10 \\ \hline \mathsf{P} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{array}.$$

Пример 3.2.2 *Now, let's imagine that we have two boxes each containing one coin of each value. How much money can we get if we take one coin from each box at random?*

Box 1 / Box 2	*1*	*2*	*5*	*10*
1	2	3	6	11
2	3	4	7	12
5	6	7	10	15
10	11	12	15	20

The distribution of a sum of coin values is shown on the screen.

$$\begin{array}{c|c|c|c|c|c|c|c|c|c|c} \xi & 2 & 3 & 4 & 6 & 7 & 10 & 11 & 12 & 15 & 20 \\ \hline \mathsf{P} & \frac{1}{16} & \frac{2}{16} & \frac{1}{16} & \frac{2}{16} & \frac{2}{16} & \frac{1}{16} & \frac{2}{16} & \frac{2}{16} & \frac{2}{16} & \frac{1}{16} \end{array}.$$

Why are the probabilities different? The thing is that the sum equal to 2 *is obtained for one elementary event* $1{+}1$ *(1 from the first box and 1 from the second box). The probability of this event in the classic scheme is equal to one-sixteenth* ($\frac{1}{16}$)*. The sum equal to 3 is obtained for 2 elementary events:* $1 + 2$ *and* $2 + 1$*. Therefore, the probability of this event equals two-sixteenth* ($\frac{2}{16}$)*. Similarly, we obtain the rest probabilities. A useful check is to verify that their sum equals one.*

3.3 Distribution Function of Discrete Random Variable

Assume we have discrete random variable [ksi] ξ. Let's consider numerical function f that transforms the value of this random variable. Besides, function [f of ksi] ($f(\xi)$) is also a discrete random variable. Its distribution law can be found based on the distribution of [ksi] ξ. Values of [f of ksi] ($f(\xi)$) are obtained by plugging values [x i] x_i of the random variable ξ as an argument of function f. The obtained values preserve the probabilities of arguments x_i. When values f of several arguments coincide, this value is assigned the probability of the sum of the probabilities of these arguments.

Пример 3.3.1 *Let the distribution of random variable [ksi] ξ be given in the table*

ξ	-3	3	7
P	0.3	0.5	0.2

We can consider linear function [f of ksi that equals ksi plus 5] $f(\xi) = \xi + 5$. Its distribution is shown on the screen.

$\xi + 5$	2	8	12
P	0.3	0.5	0.2

Let's consider quadratic function [f of ksi equals squared ksi] $f(\xi) = \xi^2$. Its distribution is shown on the screen.

ξ^2	9	49
P	0.8	0.2

Note that event [squared ksi equals 9] ($\xi^2 = 9$) consists of two disjoint events [ksi equals 3] $\xi = 3$ or [ksi equals negative 3] $\xi = -3$, which means that the probability of squared ksi to take the value 9 is the sum of the probabilities of ksi to take the values 3 and negative 3 ($\mathsf{P}(\xi^2 = 9) = \mathsf{P}(\xi = 3) + \mathsf{P}(\xi = -3)$).

Later, we will often use transformations (usually linear) of random variables.

3.4 Independence Between Two Discrete Random Variables

Let [ksi] ξ be a discrete random variable with values [x i] x_i; and [eta] η be a discrete random variable with values [y i] y_i.

Определение 3.4.1 *Random variables [ksi] ξ and [eta] η are independent if, for any values [x i] x_i and [y i] y_i, events [ksi equals x i] $\xi = x_i$ and [eta equals y i] $\eta = y_i$ are independent, that is,*

$$\mathsf{P}\left(\xi = x_i, \eta = y_i\right) = \mathsf{P}\left(\xi = x_i\right) \cdot \mathsf{P}\left(\eta = y_i\right).$$

Пример 3.4.1 *The experiment is rolling a die twice. The number of scores on the second roll does not depend on the scores on the first roll. Thus, if we say that random value [ksi] ξ is the number of points on the first roll, and random value [eta] η is the number of points on the second roll, these random values are independent.*

Пример 3.4.2 *Assume once again that there's a box containing one coin of each value, that is, 1 ruble, 2 rubles, 5 rubles, and 10 rubles. We take two coins from the box, one by one. Let [ksi] ξ be the value of the first coin, and [eta] η of the second one.*

1. *If we place the first coin back to the box, random variables [ksi] ξ and [eta] η will be independent.*

2. *If we keep the first coin, random variable [eta] η will depend on [ksi] ξ. It's as simple as that.*

3.5 Numerical Characteristics of Discrete Random Variables

Let's consider the main characteristics of discrete random variables in probability theory. Further on, we will try to assess these theoretical characteristics based on available experimental data.

3.5.1 Expected Value

Let the distribution of random variable [ksi] ξ be given in the table shown on the screen.

ξ	x_1	$\dots$	x_n
P	p_1	$\dots$	p_n

.

Определение 3.5.1 *The expected value of random variable [ksi] ξ is number [E ksi] that equals the sum of all possible values of the random variable times the corresponding probabilities.*

The expected value is often termed an average probability value. An average value usually means the arithmetic mean (the sum of numbers divided by the count of numbers).

The arithmetic mean will surely coincide with the expected value if all [p i equal to 1 over n] $\mathsf{p}_i = \frac{1}{n}$. If [p i are equal to m i over n and the sum of m i from 1 to n equals n $\mathsf{p}_i = \frac{m_i}{n}$, $(\sum_{i=1}^{n} m_i = n)$, to obtain the traditional arithmetic mean from the expected value, we need to take the value of term x_i in the sum of values m_i times. In this case, the number of values will be equal to n. Thus, in the formula for the expected value, the probabilities are weights that indicate how often the random variable takes the corresponding value. Note that the expected value is also called the center of mass.

Пример 3.5.1 *Let's consider four random variables. Their distributions are shown on the screen:*

$$\begin{array}{c|c|c} \xi_1 & 2 & 3 \\ \hline \mathsf{P} & 0.5 & 0.5 \end{array}.$$

$$\begin{array}{c|c|c} \xi_2 & 2 & 3 \\ \hline \mathsf{P} & 0.7 & 0.3 \end{array}.$$

$$\begin{array}{c|c|c} \xi_3 & 2 & 5 \\ \hline \mathsf{P} & 0.9 & 0.1 \end{array}.$$

$$\begin{array}{c|c|c} \xi_4 & 5 & 8 \\ \hline \mathsf{P} & 0.7 & 0.3 \end{array}.$$

Note that all given random variables take only two values. That's why the taken values coincide for the first two, but the probabilities to take these values are different. Meanwhile, the values taken by the second and fourth are different, but the probabilities to take the first and second values coincide.

Let's calculate their expected values (we will sometimes call them the expectation for brevity). Expectation ksi 1 equals 2.5, expectation ksi 2 equals 2.3, and so on.

$$\mathsf{E}\xi_1 = 2 \cdot 0.5 + 3 \cdot 0.5 = 2.5.$$

$$\mathsf{E}\xi_2 = 2 \cdot 0.7 + 3 \cdot 0.3 = 2.3.$$

$$\mathsf{E}\xi_4 = 2 \cdot 0.9 + 5 \cdot 0.1 = 2.3.$$

$$\mathsf{E}\xi_4 = 5 \cdot 0.7 + 8 \cdot 0.3 = 5.9.$$

Замечание 3.5.1 *It turns out that the expected value of the first two random variables is different, although the values are the same. The expected value of the second and third has matched, and the values and their probabilities have been different. This example shows that the expected value characterizes the value of the random variable on average, and the coincidence of the random variables does not follow from the equation of expected values. In other words, their laws of distribution.*

However, the mean value is often useful. Let's look at some examples.

Пример 3.5.2 *Let's write a distribution series for the random variable that defines how much time student spends on social media. We will round the time to integer values. Assume that the student spends 2 hours on social media with the probability of five-sixteenth* $\frac{5}{16}$, *3 hours with the probability of one-eighth* $\frac{1}{8}$, *1 hour with the probability of one-half* $\frac{1}{2}$, *and 0 hours with the probability of one-sixteenth* $\frac{1}{16}$ *(when there's no Internet). The distribution series is shown on the screen.*

ξ	0	1	2	3
P	$\frac{1}{16}$	$\frac{1}{2}$	$\frac{5}{16}$	$\frac{1}{8}$

We need to calculate the expected value for this random variable. It will allow us to define how much time the student spends on social media on average. It turns out that social media take 1.5 hours on average.

Замечание 3.5.2 *Based on [E ksi]* $\mathsf{E}\xi$, *it wouldn't be true to say that the student is going to spend exactly 1.5 hours on social media someday. However, we can carry out a long-term assessment. For example, the student will spend 365 by 1.5 that is 547.5 hours on this enjoyable pastime. Let's divide it by* 24 *hours, and we will get* 22 *days a year.*

Пример 3.5.3 *In some lottery, a winner will get an apartment that costs* 4 *million rubles.* 5 *million tickets were sold. Each ticket sells for* 500 *rubles. How much money will the company make from each player? Let [ksi]* ξ *be a random variable reflecting the amount of money paid (or received). Recall that we spent* 500 *rubles per ticket. We will take this into account when formulating the law of distribution of the prize. The distribution of the random variable is shown on the screen.*

ξ	-500	3999500
P	4999999/5000000	1/5000000

Then,

$$\mathsf{E}\xi = -500 \cdot 4.999.999/5.000.000 + 3.999.500 \cdot 1/5.000.000 = -499.2.$$

That's the average amount of money that each lottery player will lose.

3.5.2 Properties of Expected Value

1. The expected value of the constant equals that constant.

2. The constant factor can be brought outside of the expected value sign.

3. For any random variables [ksi] ξ and [eta] η, the expected value of the sum equals the sum of the expected values.

Замечание 3.5.3 *Note that, apart from the expected value, other numerical characteristics are used to estimate a random variable. These characteristics are the mode and median. We'll talk about them in more detail a bit later.*

3.5.3 Variance

In addition to the expected value, we will also consider another important numerical characteristic of a random variable. It's a measure of the deviation of the random variable values from the average, that is, from the expected value. The next examples can well demonstrate the importance of this characteristic.

Пример 3.5.4 *Consider 2 groups of people. Let random variables [ksi 1]* ξ_1 *and [ksi 2]* ξ_2 *show the height of a random person from the corresponding group.*

ξ_1	182	184	186
P	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

ξ_2	162	178	192	204
P	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$

Consider the laws of distribution of people by height in group 1 and group 2. Surprisingly, the expected value, that is, the average height in these groups is the same (184 *cm). We see that the height is homogeneous in group 1, and group 2 shows the large data spread.*

Пример 3.5.5 *Let's consider two groups of students who have just taken an exam. Compare their grades. In the first group, the grades are* $3, 4, 4, 4, 4$*, in the second,* $2, 3, 4, 5, 5$.

In both groups, the expected value is 3.8. *The first group is more homogeneous, and the second demonstrates the large data spread.*

Замечание 3.5.4 *To evaluate the homogeneity (or the spread), we need to calculate the variance. It allows us to estimate how close the random variable values are to its mean value.*

Определение 3.5.2 *The variance of random variable [ksi]* ξ *is number [D ksi] that equals the expected value of the squared deviation of random variable ksi from its expected value.*

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

It can be proved that the variance equals the difference between the expected value of squared random variable ksi and its squared expected value.

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

This formula is commonly used to find the variance.

Пример 3.5.6 ***The remaining part of the example.*** *We can calculate the variance for the first group of students:*

To begin with, let's write the distributions for squared random variable ksi 1 and find its expectation. It will be 14.6, then, the variance will be 0.16.

The variance calculated for the second group in the same way will be 1.36. You can see that the variance (the spread characteristic) of the second group is larger.

3.5.4 Properties of Variance

1. The variance of a constant is 0.

2. The constant factor can be brought outside of the variance sign if squared.

3. Adding a constant value, c, to a random variable does not change the variance.

4. If random variables [ksi] ξ and [eta] η are independent, the variance of their sum (difference) equals the sum of variances.

The square root of the variance that is the residual standard error (RSE) is used as a linear measure of the deviation from the average.

Определение 3.5.3 *Number [sigma that equals the square root of the variance] $\sigma = \sqrt{\mathsf{D}\xi}$ is called residual standard error (RSE) of random variable [ksi] ξ. The residual standard error is also termed the standard deviation.*