Probability theory and Mathematical Statistics

Vilius Stakėnas

2012

I. Probability space

2 / 369

1.1 Introduction

3 / 369

Experiments, outcomes and events

Probability theory explores the **experiments with non-predictable outcomes.**
We say, that the outcomes of these experiments are random.
It is usual to denote the set of all possible outcomes of the experiment (the sample space) by Ω and the specific outcomes by $\omega, \omega_1, \omega_2, \ldots$

4 / 369

Toss of a coin

Example 1. *Toss of a coin*
The experiment: we toss a coin. The are two possible outcomes: H (head) and T (tail).

Indeed, there are more possible outcomes: the coin can remain standing on the edge or can be lost. Because these outcomes are rare we suppose they never happen.

The sample space is $\Omega = \{H, S\}$.

▲

5 / 369

Three tosses

Example 2. *Three tosses of a coin*
If we perform three tosses of a coin, the outcome of the experiment will be a string of three symbols corresponding to the results of the tosses. The sample space

$$\Omega = \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}.$$

▲

Up to the first success

Example 3. *Up to the first success*
Suppose the head in the throw of a coin means success. We perform the throwing of a coin up to the first success. It can appear in the first toss, in the second, ... It can never occur (at least theoretically).
The sample space of this experiment is infinite:

$$\Omega = \{H, TH, TTH, TTTH, \ldots, TTTT...\}.$$

▲

Braking the car

Example 4. *Braking the car*
Our experiment – braking the car. We are interested in the braking distance. The outcomes can be denoted by non-negative numbers.
The sample space $\Omega = (0; +\infty)$.
Of course, the breaking distance can't be very large for the usual car. But suppose we are breaking the space shuttle!

▲

8 / 369

The very beginning

We start exploring the experiment as follows:

- define the sample space Ω of the experiment;
- associate all other events related to the experiment with the subsets $A \subset \Omega$ of the sample space.

We shall call these subsets themselves the events.

9 / 369

Some definitions

Definitions

Let Ω be a sample space of the experiment, and $A \subset \Omega$ some event related to the experiment.
If $A = \Omega$, we say, that the event A is certain.
If $A = \emptyset$, we say that A is impossible.
The event

$$\bar{A} = \{\omega \in \Omega : \omega \notin A\}$$

is called the complement of A
The complement of the event A is the event, which occurs if and only if A does not occur.

Quick exercises

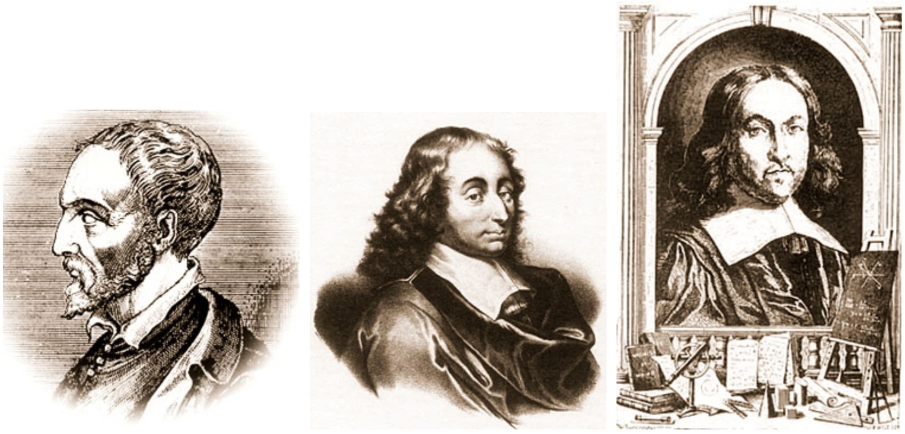
Problem 1. *Experiment - the exam in Probability theory. Which outcomes of interest could be fixed? Propose at least two variants of possible sets of outcomes.*

Quick exercises

Problem 2. *Eight short distance runners will start in the semi-final. Four of them with the best results will run in the final. We are interested only in results of the runner R in the two final contests of the game. Define the set of possible outcomes of interest.*

1.2. The classical definition of the probability

The founders



The founders of probability computation: G. Cardano, B. Pascal, P. Fermat

The classical definition of the probability

For a finite set B we denote its cardinality by $|B|$.

The classical definition

Definition 1. *Consider the experiment with the finite set Ω of outcomes which are equally likely. The probability of the event $A \subset \Omega$ is defined by*

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

Simple properties of probability

Theorem 1. Let Ω consists of finite number of outcomes which are equally likely. The probability has the following properties:

1. $0 \leqslant P(A) \leqslant 1.$
2. $P(\emptyset) = 0, \quad P(\Omega) = 1;$
3. for any event $A \subset \Omega$ we have

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

The counting rule

The counting rule

Let the set U consists of M elements and the set V of N elements. Then there are $M \cdot N$ different pairs $\langle u, v\rangle$ where $u \in U$ and $v \in V$.

The counting rule

The generalized counting rule

Let the sets $U_1, U_2, \ldots, U_k$ consist of $N_1, N_2, \ldots, N_k$ elements respectively. Then there are $N_1 \cdot N_2 \cdots N_k$, different tuples $\langle u_1, u_2, \ldots, u_k\rangle$, where $u_1 \in U_1, \ldots, u_k \in U_k$.

Ordered selections with repetitions

Suppose $U_1 = U_2 = \ldots = U_k = U, |U| = N$. Then there are

$$\underbrace{N \cdot N \cdots N}_{k} = N^k.$$

different tuples $\langle u_1, u_2, \ldots, u_k \rangle, u_i \in U$. A tuple $\langle u_1, u_2, \ldots, u_k \rangle$ can be considered as selection of copies of elements from U.

Example

Example 5.
A symmetrical dice is rolled four times. What is the probability to get the faces with the odd, even, odd, even points respectively?

▲

Example

Example 6.
A symmetrical dice is rolled four times. What is the probability to get the faces with six points at least once?

▲

Example

Example 7. *Random walk*
Let $A_1, A_2, \ldots, A_n$ be some different points in the space. A particle appears at a random point of this set and wanders from point to point making a choice every second: whether it remains at the same point or jumps to a different randomly chosen point.
What is the probability that in r seconds the particle will visit at least one of the points $A_1, A_2, \ldots, A_m (m < n)$?
What is the probability that in r seconds the particle makes at least one return to an already visited point?

▲

Permutations

Let us consider ordered selections of different elements from the set U.
Let there are N elements in the set U.
A row of k different elements from the set U is called
permutation of length k without repetition from N elements.
Denote the number of distinct permutations of length k by A_N^k.

$$
\begin{aligned}
A_N^2 &= N \cdot (N-1), \quad A_N^3 = N \cdot (N-1) \cdot (N-2), \ldots, \\
A_N^k &= \underbrace{N \cdot (N-1) \cdot (N-2) \cdots (N-k+1)}_{k}.
\end{aligned}
$$

24 / 369

Example

Example 8.
There are 52 playing cards. What is the probability that 5 randomly drawn card will be all spades?

▲

25 / 369

Example

Example 9. *Birthsday problem*
What is the probability that in the group of r persons there are at least two celebrating their birthsday at the same day of the year?
What is the smallest value of r such that the probability to find at least to persons in the group of r persons having the same birthsday is larger than $\frac{1}{2}$?

▲

Combination

A subset of k elements chosen from a set having n elements is called **combination of k elements from a set of n elements**.

The number of distinct combinations C_n^k

$$A_n^k = C_n^k \cdot k!,$$

$$C_n^k = \frac{A_n^k}{k!} = \frac{\overbrace{n \cdot (n-1) \cdot (n-2) \cdots (n-k+1)}^{k}}{\underbrace{k \cdot (k-1) \cdot (k-2) \cdots 1}_{k}}.$$

Example

Example 10.
From the set of 52 playing cards 5 cards were chosen randomly. What is the probability to get exactly two hearts among them?

▲

Example

Example 11.
There are N white and M black balls (N good and M defective articles on the shelf). If k balls are chosen randomly, what is the probability to get this way exactly n white balls?

▲

Random walk

Example 12. *Random walk of a particle*
A particle starts at the point $O(0;0)$ and travels to the point $(n;n)$. It choses one of the shortest paths: i.e. from the point $(a;b)$ moves to either the point $(a+1;b)$ or $(a;b+1)$. What is the probability that the particle visits the point $A(x,y)$ $(x,y<n)$?

▲

Example

Example 13. *Random walk*
A particle starts at the point $O(0;0)$. It choses direction randomly and moves either to the point $(1;1)$ or $(1,-1)$. It proceeds in the same way: from the point (u,v) it jumps to either the point $(u+1,v+1)$ or $(u+1,v-1)$. What is the probability that after n steps it occurs at the point $(n;x)$?

▲

Problems

Problem 3. *The are n musical songs in the playlist, they are played in random order. The same song can be played repeatedly. What is the probability that the first song will be played exactly after m other songs?*

Problems

Problem 4. *There are n musical songs in the playlist, they are played in random order with the exception that the same song can not be played again if it was just finished to play. What is the probability that the first song will be played exactly after m other songs?*

Problems

Problem 5. *The game with a symmetrical coin: if it lands on the head, the first player gets one point, if the coin lands on the tail, the second player gets the point. What is the probability that after* 4 *tosses both players will have* 2 *points each? What is the probability that after five tosses the first player will be ahead with one point?*

Problems

Problem 6. *There are* n *photos selected for the exposition,* m *of them are landscapes. The photos will be exposed in random order. What is the probability that all landscapes will be hang up ir a successive row?*

Problems

Problem 7. *The same set of n messages was sent to m persons. Each of them deleted one messeage randomly. What is the probability that the same message was deleted at least twice?*

Problems

Problem 8. *There are n floors in the building, there are m men in the elevator. Each of them can get out in each floor. What is the probability that exactly two men will go out in the same floor and all other men will go out alone in different floors?*

Example

Example 14. *When the clock stops?*
When the batteries will run down, the clock will stop. What is the probability that this occurs between 2 and 3 PM?

▲

The sample space Ω is the circle. The event of interest is an arc A.

$$P(A) = \frac{\text{length of the arc} A}{\text{length of } \Omega} = \frac{1}{12}.$$

Definition of geometrical probability

Definition

Let the outcomes of the experiment are represented by the points of some geometrical area Ω, all outcomes are equally likely and the geometrical measure of Ω is positive. Then the probability of the event $A \subset \Omega$ is

$$P(A) = \frac{\textit{geometrical measure of } A}{\textit{geometrical measure of } \Omega}.$$

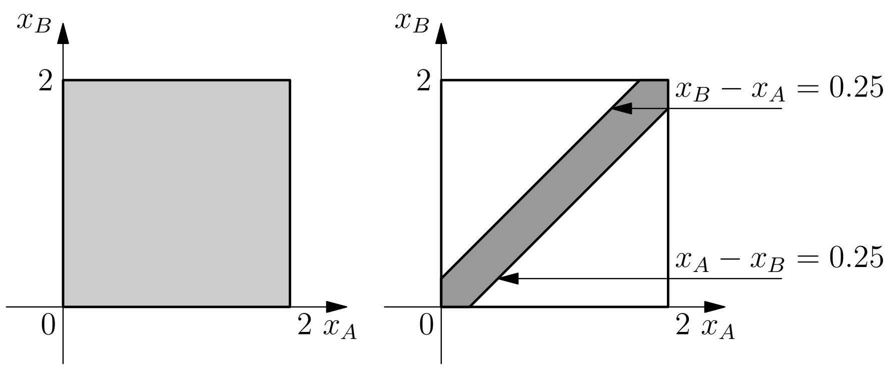
Note that only this definition implies a restriction to the the events under consideration: the corresponding sets should be measurable.

Example

Example 15. *Phone calls*
Two friends promised to call to the third between the 2 and 3 PM o'clock. What is the probability that the interval between two calls will not exceed 15 minutes?

▲

Solution



Hence,

$$P(U) = \frac{2^2 - 1,75^2}{2^2} = \frac{15}{64} \approx 0,234.$$

Example

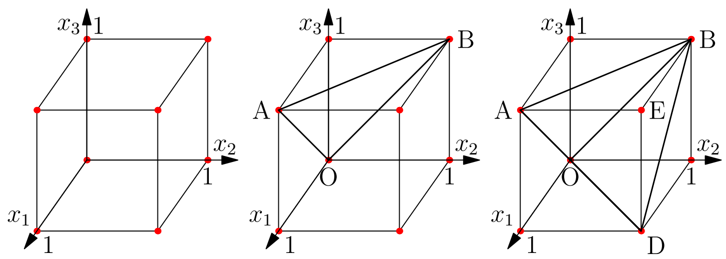
Example 16. *Constructing a triangle*
The lengths of three randomly chosen sides of a triangle are the numbers from the interval $[0;1]$. What is the probability that the triangle with the chosen sides exist?

▲

Let the lengths of the sides be x_1, x_2, x_3. For a triangle to exist it is necessary and sufficient that the triangle inequalities hold:

$$x_1 + x_2 \geqslant x_3, \quad x_1 + x_3 \geqslant x_2, \quad x_2 + x_3 \geqslant x_1.$$

Solution



$$P(T) = \frac{\text{volume of } OABDE}{\text{volume of the unite cube}}.$$

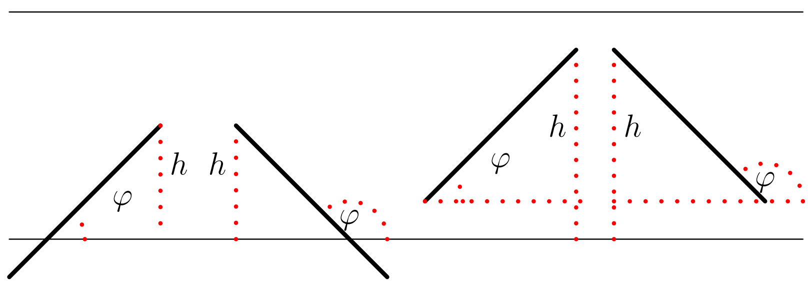
How to compute the volume of the solid? Substract from the volume of the cube the volumes of three pyramids which are cut off:

$$P(T) = \frac{1 - 3 \cdot \frac{1}{6}}{1} = \frac{1}{2}.$$

The Buffon's problem

Example 17. *Buffon's problem*
An infinite system of parallel lines is drawn on the plane. The distance between any neighouring lines is equal to 1. A needle of length l $(l < 1)$ is thrown on the plane. What is the probability that the needle intersects some line?

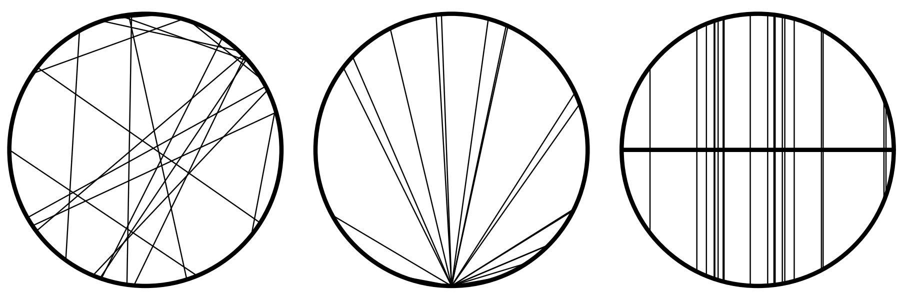
▲



Bertrand's problem

Example 18. *Bertrand's problem*
In the unite circle a chord is randomly chosen. What is the probability that the length of the chord exceeds the length of the equilateral triangle inscribed into circle?

▲

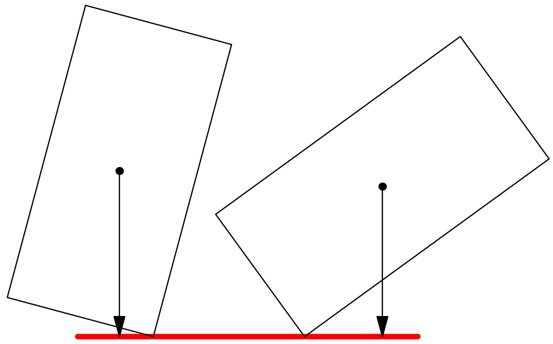


The answer depends on the interpretation of random choice!

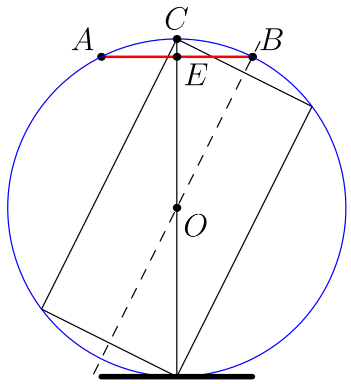
Thick coin

Example 19. *Landing on the edge*
Radius of the symmetrical coin is r, thickness is h. What is the probability that if we toss this coin it will land on the edge?

▲



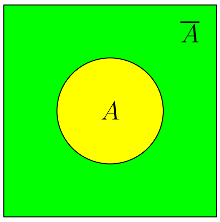
Thick coin



$$P = \frac{2\pi \cdot OC \cdot CE}{2\pi \cdot OC^2}$$

Representation of events

Events are represented by subsets of the set of outcomes. The operations with the subsets are interpreted as operations with events. Let Ω be the set of all outcomes and $\mathcal{A}$ the system of events associated with the experiment, i.e. some system of subsets of Ω.
Every event A in the family has its complement $\overline{A}$.

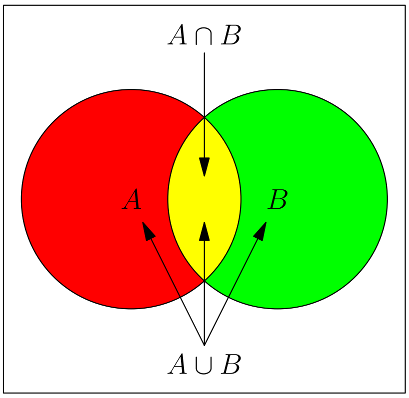


Intersection and union of events

Definition 2. *Let A and B are two events related to the same experiment, i.e. they are represented by the subset of the same sample space. The event, represented by the outcomes, which belong to both events A, B, is called intersection and denoted by $A \cap B$.*
The event represented by the outcomes, which belong to at least one of the events A, B, is called union and denoted by $A \cup B$.

Intersecion and union

The events and the operations with them are clearly illustrated by the diagrams.



If all outcomes from A belong to B too, then we shall write $A \subset B$.

$$A \subset A \cup B, \quad A \cap B \subset A, \quad A \cap B \subset B.$$

Properties of operations with the events

Theorem 2. Let A, B, C be the events related to the same experiment, Ω is the certain, $\emptyset$ the impossible event. The following statesments are true:

1. $\overline{\emptyset} = \Omega, \ \overline{\Omega} = \emptyset, \ \overline{\overline{A}} = A;$
2. $\emptyset \cup A = A, \ \Omega \cup A = \Omega, \ \emptyset \cap A = \emptyset, \ \Omega \cap A = A;$
3. $A \cup A = A, \ A \cap A = A;$
4. $A \cap \overline{A} = \emptyset, \ A \cup \overline{A} = \Omega;$
5. $(A \cup B) \cap C = (A \cap C) \cup (B \cap C);$
6. $\overline{A \cup B} = \overline{A} \cap \overline{B}, \ \overline{A \cap B} = \overline{A} \cup \overline{B}.$

Example

Example 20.
Using the properties of operations we can simplify the complicated expressions. For example,

$$D = \overline{(\overline{A} \cup B) \cup C} = \overline{(\overline{A} \cup B)} \cap \overline{C} = A \cap \overline{B} \cap \overline{C}.$$

▲

The probabilistic model of the experiment

The probabilistic model (the probabilistic space) consists of

- the sample space $\Omega \neq \emptyset$;
- system of events $\mathcal{A}$ related to experiment;
- definition of probability $P : \mathcal{A} \to [0; 1]$.

What requirements should be set on $\mathcal{A}$ and P?

Algebra of random events

It is quite natural to consider the system of events $\mathcal{A}$ having the following properties

- $\emptyset \in \mathcal{A}, \quad \Omega \in \mathcal{A}$,
- for each $A \in \mathcal{A}$ we have also $\overline{A} \in \mathcal{A}$,
- for any sequence of events (finite or infinite) $A_1, A_2, \ldots \in \mathcal{A}$ we have $\cap_i A_i, \cup_i A_i \in \mathcal{A}$.

These conditions are not independent: some of them follow from other ones.

σ-algebra of events

Definition 3. *Let Ω be non-empty set. A system of subsets $\mathcal{A}$ is called σ-algebra, if the following conditions are satisfied:*

- $\Omega \in \mathcal{A}$;
- *if* $A \in \mathcal{A}$, *then* $\overline{A} \in \mathcal{A}$;
- *if* $A_i \in \mathcal{A}, i = 1, 2, \ldots,$ *then*

$$\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}.$$

Further properties

Theorem 3. Let $\mathcal{A}$ be some σ-algebra of subsets of the set Ω. The following propositions are true:

- if $A_i, i \in I$, is some countable system (finite or infinite) of sets from $\mathcal{A}$, then $\cap_{i \in I} A_i \in \mathcal{A}$;
- if $A, B \in \mathcal{A}$, then $A \backslash B \in \mathcal{A}$.

Here $A \backslash B = A \cap \overline{B}$; this set is called the difference of A and B.

58 / 369

Example

Example 21.
Let $\Omega = \{1, 2, 3, 4, 5\}$, and we are especially interested in the events $\{1, 2, 3\}, \{3, 4, 5\}$.
The smallest σ-algebra, which includes these events is:

$$\mathcal{A} = \{\emptyset, \{1, 2, 3\}, \{3, 4, 5\}, \{3\}, \{4, 5\}, \{1, 2\}, \{1, 2, 4, 5\}, \Omega\}.$$

▲

59 / 369

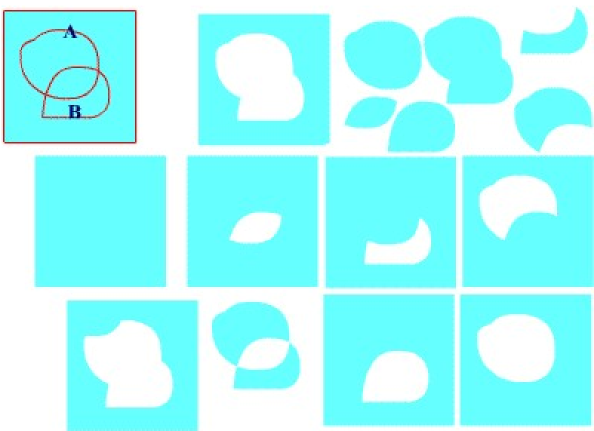
Generated σ-algebra

Definition 4. *Let $\mathcal{S}$ be some system of subsets of the non-empty set Ω. If a σ-algebra $\mathcal{A}$ satisfies the condition $\mathcal{S} \subset \mathcal{A}$ and has the property: for any σ- algebra $\mathcal{A}'$, $\mathcal{S} \subset \mathcal{A}'$, we have $\mathcal{A} \subset \mathcal{A}'$, then σ- algebra $\mathcal{A}$ is called σ-algebra, generated by $\mathcal{S}$ and denoted by $\mathcal{A} = \sigma(\mathcal{S})$.*

Think about $\sigma(\mathcal{S})$ as a smallest σ-algebra, which covers the system $\mathcal{S}$.

Generated σ-algebra

Theorem 4. For any system of subsets $\mathcal{S}$ $\sigma(\mathcal{S})$ exists.



Important example

Example 22. *Borelian σ-algebra*
Let $\mathcal{S}$ be the system of intervals $[a, b) \subset \mathbb{R}$.
Then $\sigma(\mathcal{S})$ is called Borelian σ-algebra. We shall denote it by $\mathcal{B}$.

If $\mathcal{S}$ is the system of all n-dimensional intervals $[a_1, b_1) \times \ldots \times [a_n, b_n)$, then $\sigma(\mathcal{S})$ is called Borelian σ-algebra of the space $\mathbb{R}^n$. Notation: $\mathcal{B}^n$.
▲

The sets from the Borelian σ-algebra are called Borelian. All "good" sets

are Borelian!
There are a lot of non-Borelian sets, but we do’nt encounter them in practice. Think for analogy about the rational and irrational numbers!

62 / 369

1.6. Probability

63 / 369

Definition of probability

Definition 5. *Let $\mathcal{A}$ be a σ-algebra of the subsets of the sample space Ω. A function $P : \mathcal{A} \to [0, 1]$ is called probability (or probabilistic measure), if*

- $P(\Omega) = 1$;
- *if* $A_i \in \mathcal{A}, A_i \cap A_j = \emptyset$ $(i, j = 1, 2, \ldots, i \neq j)$, *then*

$$P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i).$$

The second condition in the definition of P is called σ-additivity.

64 / 369

Probabilistic space

Definition 6. *The triple $\langle \Omega, \mathcal{A}, P \rangle$, where $\mathcal{A}$ is a σ-algebra of subsets of the sample space Ω and P – probability, is called probabilistic space.*

Discrete probabilistic space

Constructing the discrete probabilistic space

Let Ω be a discrete sample space, i.e. it is finite or infinite but countable, $\Omega = \{\omega_1, \omega_2, \ldots\}$.
The σ-algebra $\mathcal{A}$ of events may be taken consisting of all subsets $A \subset \Omega$.
We need to define the probabilities of the outcomes

$$P(\omega_1) = p_1, \; P(\omega_2) = p_2, \ldots \quad (0 < p_i < 1)$$
$$p_1 + p_2 + \ldots = 1.$$

Then the probability of an arbitrary event $A \subset \Omega$ is defined by

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

Example

Example 23. *Blossoms of a plant*
We wait for the blossoms of some plant. Hence, the experiment is growing of the plant. The outcomes we have chosen are:

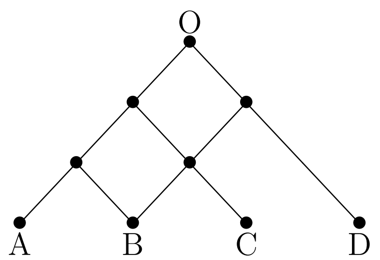
$$\begin{aligned}
\omega_0 &= \{\text{there will no blossoms}\},\\
\omega_1 &= \{\text{the plant will have exactly one blossom}\},\\
\omega_2 &= \{\text{the plant will have exactly two blossoms}\},\\
\omega_3 &= \{\text{the plant will have exactly three blossoms}\},\\
\omega_4 &= \{\text{the plant will have at least four blossoms}\}.
\end{aligned}$$

How to define the probabilities? No recipes. An advice: consult the botany books. May be you shall find some statistical data from the previous observations?

▲

Example

Example 24.
The experiment – journey down from the point O choosing the directions at all crossroads randomly.



The sample space $\Omega = \{\omega_A, \omega_B, \omega_C, \omega_D\}$, here the outcome ω_X means that we arrived to point X.
The probabilities are

$$P(\omega_A) = \frac{1}{6},\ P(\omega_B) = \frac{3}{6},\ P(\omega_C) = \frac{2}{6},\ P(\omega_D) = \frac{1}{6}.$$

▲

Some equalities

Suppose that the probabilistic space $\langle\Omega, \mathcal{A}, P\rangle$ is given.

Theorem 5. The following statements are true:

1. $P(\emptyset) = 0$;
2. for any countable system (finite or infinite) of disjoint events A_i $(i \in I)$ we have
$$P(\bigcup_{i\in I} A_i) = \sum_{i\in I} P(A_i),$$
3. $P(A \backslash B) = P(A) - P(A \cap B)$; in particular $P(\overline{A}) = 1 - P(A)$;
4. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$;

Inequalities

Theorem 6. Let A_i $(i \in I)$ be some countable set of events and A be some event such that $A \subset \cup_i A_i$. Then

$$P(A) \leq \sum_{i\in I} P(A_i).$$

In particular case for $A = \cup_i A_i$,

$$P(\bigcup_{i\in I} A_i) \leq \sum_{i\in I} P(A_i).$$

Optimal choice

Example 25. *The problem of optimal choice*
There are n objects of different quality, they appear in random order. We would like to choose the best object. The procedure of choice has a constraint: if we did'nt chosen an object it disappears forever.

▲

The strategy of choice: at the beginning the number $m,\ 0 \leqslant m \leqslant n$ is fixed and the decision taken not to choose any of the first m objects, only evaluate their "quality". Thereafter we choose the first object which is better than all m objects observed.

Optimal choice

What is the probability to choose the best among all n objects? What is the probability to choose the second, the third, the fourth ... in the "quality row"? What is the probability not to choose anything?

Optimal choice

Denote the probability to choose the best object by p_m, and let $p_m(m+1), p_m(m+2), \ldots, p_m(n)$ be the probabilities that the best object appears in the row at $m+1, m+2, ..., n$ place, respectively. Hence

$$p_m = p_m(m+1) + p_m(m+2) + \cdots + p_m(n).$$

Optimal choice

$$\begin{aligned}
p_m(m+1) &= \frac{(n-1)!}{n!} = \frac{1}{n} \\
p_m(m+2) &= \frac{m \cdot C_{n-1}^{m+1} \cdot m! \cdot (n-m-2)!}{n!} = \frac{m}{n} \cdot \frac{1}{m+1} \\
p_m(m+j) &= \frac{m}{n} \cdot \frac{1}{m+j-1}
\end{aligned}$$

The probability to get the best object with the strategy parameter m is

$$p_m = \frac{m}{n}\Big(\frac{1}{m} + \frac{1}{m+1} + \cdots + \frac{1}{n-1}\Big) = \frac{m}{n}\sum_{j=m}^{n-1}\frac{1}{j}.$$

The practical question

If we have to choose one from, say, $n = 10, 20$ objects, which value of m guarantees the largest probability to get the best one?
Let m_n is the value of m, which maximizes p_m.
Computational results:

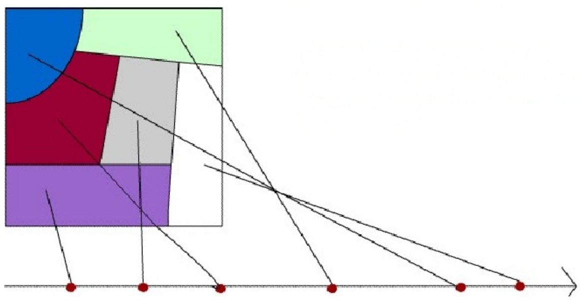
$n =$	5	10	15	20	25	30
$m_n =$	2	4	5	7	9	11
$p_{m_n} =$	$0,433$	$0,398$	$0,389$	$0,384$	$0,381$	$0,379$

$$\frac{m_n}{n} \to \frac{1}{\mathrm{e}}, \quad p_{m_n} \to \frac{1}{\mathrm{e}}, \quad n \to \infty.$$

1.8. Simple random variables and probabilities 77 / 369

Simple random variables

A random variable is a function assigning numbers to outcomes



Definition 7. *A function $\xi : \Omega \to \mathbb{R}$ is called simple random variable, if ξ takes the values from the finite set and for each value x*

$$\xi^{-1}(x) = \{\omega : \xi(\omega) = x\} \in \mathcal{A}.$$

It follows from the definition that the probabilities $P(\xi = x)$ can be found!

Indicators

Let $A \in \mathcal{A}$ be some random event, then the function

$$I_A(\omega) = \begin{cases} 1, & \text{if } \omega \in A, \\ 0, & \text{if } \omega \notin A. \end{cases}$$

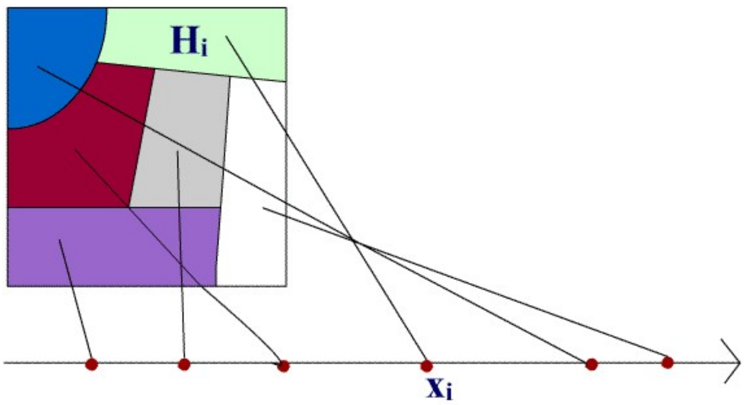
is a simple random variable. We call this variable indicator of event A. It indicates whether the event A occured or not.

$$I_{A \cup B} = I_A + I_B - I_{A \cap B}, \; I_A \cdot I_B = I_{A \cap B}.$$

79 / 369

Simple random variables

Simple random variables via the indicators



If ξ is a simple random variable with the values x_i, let $H_i = \xi^{-1}(x_i)$. Then

$$\xi(\omega) = \sum_i x_i I_{H_i}(\omega).$$

80 / 369

Sum of variables

Theorem 7. Let ξ, η be the simple random variables and a, b some real numbers. Then $\zeta = a\xi + b\eta$ is also a simple random variable.

The statement is true in the general case: a linear combination of finite set of simple random variables is a simple random variable, too.

Expected value of random variable

Definition 8. *Let ξ be a simple random variable with the values x_i. Let $H_i = \xi^{-1}(x_i)$ $(i = 1, \ldots, n)$. The expected value of the random variable ξ is defined by*

$$\mathbf{E}[\xi] = \sum_{i=1}^{n} x_i P(H_i) = \sum_{i=1}^{n} x_i P(\xi = x_i).$$

$$\mathbf{E}[I_A] = 0 \cdot P(I_A = 0) + 1 \cdot P(I_A = 1) = P(A).$$

Additivity of the expected value

Theorem 8. Let ξ, η be simple random variables and a, b real numbers. If $\zeta = a\xi + b\eta$, then

$$\mathbf{E}[\zeta] = a\mathbf{E}[\xi] + b\mathbf{E}[\eta].$$

Theorem 9. Let $\xi_1, \dots, \xi_n$ be simple random variables and $a_1, \dots, a_n$ some real numbers. Then

$$\mathbf{E}[a_1\xi_1 + \cdots + a_n\xi_n] = a_1\mathbf{E}[\xi_1] + \cdots + a_n\mathbf{E}[\xi_n].$$

83 / 369

Probability of the union of events

Method of proof: start with

$$I_{A\cup B} = I_A + I_B - I_{A\cap B},$$

apply the additivity property of expected value and use the equality $P(C) = \mathbf{E}[I_C]$ which holds for every event C. We get then

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

84 / 369

Probability of the union of events

The method of proof just outlined works in the general setting.

Theorem 10. For arbitrary random events A_i $(i = 1, \ldots, n)$ we have

$$P(\bigcup_{i=1}^{n} A_i) = \sum_{r=1}^{n} (-1)^{r-1} S_r,$$

where

$$S_r = \sum_{1 \le i_1 < \ldots < i_r \le n} P(A_{i_1} \cap \ldots \cap A_{i_r}).$$

A different formula

$$P(\bigcup_{i=1}^{n} A_i) = 1 - P(\overline{A_1} \cap \overline{A_2} \cap \ldots \overline{A_n}).$$

Exactly k events

Theorem 11. Let A_i $(i = 1, \ldots, n)$ be arbitrary random events and B_k $(k = 0, 1, \ldots, n)$ is the event which means that exactly k events from the system A_i occured. Then

$$P(B_k) = \sum_{r=k}^{n} (-1)^{r-k} C_r^k S_r,$$

where the quantities S_r are defined in the previous theorem.

Example

Example 26. *Pack of cards*
Let the pack having N different cards is thoroughly shuffled. What is the probability that in the shuffled pack atl least one card will remain in the same place? Exactly m $(m = 0, 1, \ldots, N)$ cards will be in their places?

▲

Let $q(N), p_m(N)$ be the probabilities of events we are interested in. Let A_i is the event, that the ith card of the pack remains in its place after shuffling. We have

$$P(A_{i_1} \cap \ldots \cap A_{i_r}) = \frac{(N-r)!}{N!}, \quad S_r = \frac{C_N^r (N-r)!}{N!} = \frac{1}{r!}.$$

Example

$$\begin{aligned} q(N) &= P(\bigcup_{i=1}^{N} A_i) = \sum_{n=1}^{N} (-1)^{n-1} \frac{1}{n!} = 1 - \sum_{j=0}^{N} (-1)^j \frac{1}{j!}, \\ p_m(N) &= \frac{1}{m!} \sum_{n=0}^{N-m} (-1)^n \frac{1}{n!}. \end{aligned}$$

$q(N) \to 1 - \mathrm{e}^{-1}, p_m(N) \to \mathrm{e}^{-1}/m!$, as $N \to \infty$, here e is the basis of the natural logarithm.
We can interpret this problem as problem of letters dropped randomly into the boxes. What is the probability that at least one letter will be dropped correctly?

Growing events

Theorem 12. Let A_i $(i = 1, \ldots)$ be a sequence of growing events:

$$A_1 \subset A_2 \subset \ldots, \quad A = \bigcup_{i=1}^{\infty} A_n.$$

Then $P(A_n) \to P(A),$ *as* $n \to \infty$.

Decreasing events

Theorem 13. Let A_i $(i = 1, \ldots)$ be a sequence of decreasing events:

$$A_1 \supset A_2 \supset \ldots, \quad A = \bigcap_{i=1}^{\infty} A_n.$$

Then $P(A_n) \to P(A),$ *as* $n \to \infty$.

Example

Example 27.
Suppose that you and and two friends of yours have to draw lots. There are three balls in the urn: the black, the blue and the white one. Each person draws one ball. The winner wil be who gets the white ball.

The probability to win for the first player is $\frac{1}{3}$. Are the second and the third in the worse situation?

▲

Example

Example 28.
Let there are in the urn $n = 5$ black and $m = 4$ white balls. Three balls are drawn randomly without replacement. What are the probabilities of the events $A_i = \{i\text{th drawn ball is white}\},\ i = 1, 2, 3,$?

▲

Recalculation of probability

Suppose you decided to draw the second ball and the trial started. If the first ball is white, then you can recalculate the probability to get the white ball. It is equal to $\frac{3}{8}$.

Initially the probability was calculated before the trial. The second probability was calculated using the knowledge supplied by the event A_1.

We say, that the first probability is **unconditional**, and the second one – **conditional** with the condition that the event A_1 occurred.
We shall denote the probabilities by

$$P(A_2) = \frac{4}{9}, \quad P(A_2|A_1) = \frac{3}{8}.$$

95 / 369

The conditional probability

Definition 9. *Let A, B are two random events related to the same experiment, $P(B) > 0$. The conditional probability of the event A with the condition that B occurred is the number*

$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

96 / 369

The same properties

Theorem 14. Let A, B, C be the random events, $P(C) > 0$. Then

1. $P(\Omega|C) = 1, \; P(\emptyset|C) = 1;$
2. $P(A|C) = 1 - P(\overline{A}|C);$
3. if A, B are disjoint, then

$$P(A \cup B|C) = P(A|C) + P(B|C).$$

Example

Rewrite the equality of definition as

$$P(A \cap B) = P(B)P(A|B)$$

Example 29.
Let there are $n = 5$ blacks and $m = 4$ white balls in the urn. Two balls are drawn randomly without replacement. What is the probability that both balls will be white? That at least one ball will be white?

▲

Example

Example 30.
Let there are $n = 5$ black and $m = 4$ white balls in the urn. Two balls are drawn randomly. After the first ball is drawn three balls of the same color are put back in the urn. What is the probability that two balls drawn randomly from the urn will be white? That at least one will be white?

▲

Multiplication rule

Theorem 15. Let $A_1, A_2, \ldots, A_n$ be arbitrary random events, $P(A_1 \cap A_2 \cap \ldots \cap A_n) > 0$. The following equality holds

$$P(A_1 \cap \ldots \cap A_n) = P(A_1)P(A_2|A_1)\cdots P(A_n|A_1 \cap \ldots \cap A_{n-1}).$$

Example

Example 31.
The game is as in the previous example, but three balls are randomly drawn. What is the probability that all three will be white? At least one of them will be white?

▲

Interrupted game

Example 32. *Interrupted game*
Two gamblers paid to game's bank 5 euros each and start playing with the symmetrical coin. If the tossed coin falls on the head, the first gambler gets a point. If the coin falls on the tail, the second gambler gets a point. The gambler who gets three points first gets all the bank. Let A_1 and A_2, denote the events that the first and the second wins respectively.
Because both have the same chances, $P(A_1) = P(A_2) = \frac{1}{2}$.

▲

Interrupted game

Suppose in the first toss the first gambler has got a point and is ahead by $1:0$.
If the game will be interrupted because of *force majore* how to share the bank fairly?

1.11. Properties of the conditional probabability

Examples

Example 33. *Doors and keys*
We have to unlock the doors. In one pocket there are three identical keys suitable to the lock, in the second one unsuitable key, in the third – two unsuitable keys. The pocket and the key from that pocket are chosen randomly. What is the probability that we unlock the door in the first trial?

▲

Examples

Example 34. *Doors and keys*
Suppose now that in the first pocket there is one suitable key only, in the second – one suitable and one unsuitable key, in the third – one suitable and two unsuitable keys. What is the probability that we unlock the door in the first trial?

▲

Partition rule

Theorem 16. Let $H_1, H_2, \ldots$ be the disjoint events with positive probabilities and

$$\Omega = H_1 \cup H_2 \cup \ldots.$$

Then for an arbitrary event A

$$P(A) = P(H_1)P(A|H_1) + P(H_2)P(A|H_2) + \cdots$$

Note. The sequence of events H_i may be finite or infinite.

Example

Example 35. *Exam*
There are four questions and four answers on the test sheet. It is required to join the questions and answers correctly. There were prepared 25 test sheets. A student prepared for the exam: he found all the correct pairs of the two first sheets, he found the correct answers to the first two questions on the 3-5 test sheets, found how to answer correctly to the first question of the 6-19 test sheets, and left the remaining test sheets without correct pairs. The exam is passed if all four question-answer pairs will be correct. What is the probability that this student will pass the exam?

▲

Bayes' rule

Theorem 17. Let the random events $H_1, H_2, \ldots$ be disjoint, their probabilities positive, $\Omega = H_1 \cup H_2 \cup \ldots$, A be an arbitrary event, $P(A) > 0$. Then for each H_i we have

$$P(H_i|A) = \frac{P(H_i)P(A|H_i)}{P(A)},$$

here $P(A) = P(H_1)P(A|H_1) + P(H_2)P(A|H_2) + \cdots$

Example

Example 36. *Monty Hall problem*
There are three closed envelopes before you. Two envelopes are empty, one of them contains a prize. You are allowed to choose one. After you choose the closed envelope the organizer opens that one of the remaining envelopes which is empty. Now there are two closed envelopes. If you want, you can change your envelope to the one on the table. Makes this change any sense? Find the probability to win the prize if you don't change and if you change.

▲

Example

Example 37.
There are four closed envelopes, three of them are empty, one contains a prize. You are allowed to choose two of them. After you make a choice the organizer opens the empty envelope. You can change your envelope for one on the table. What is the probability that you get a prize if you don't change and if you change?

▲

Example

Example 38. *Coffee automats*
There are three automatic coffee machines in the row. It is known that one of them does'nt operate at all, one sells coffee in approximately 50% cases, and one always operates correctly. Suppose that the randomly chosen machine gave to you coffee two times in succession. What is the probability that you have chosen an automatic coffee machine which always operates perfect?

▲

Gamblers ruin

Example 39. *Game with the coin*
Two gamblers play with the symmetric coin. If the flipped coin is a head, the second gambler pays to the first one euro, if the flipped coin is a tail, the first one pays to the second one euro. At the beginning the first gambler has x euro, the second is infinitely wealthy. The first gambler will stop playing if he loses all his money or reaches the amount equal to a euros. What is the probability that the first gambler will be ruined?

▲

Example

Example 40.
There are two white balls and one black ball in the urn. Two players draw randomly one ball each. The white ball means a prize. Let $A_1 = \{$the first player won a prize$\}$ and $A_2 = \{$the second one won a prize$\}$. The events are not disjoint. Before the trial we compute

$$P(A_1) = P(A_2) = \frac{2}{3}.$$

Suppose that the event A_1 occurred. Then

$$P(A_2|A_1) = \frac{1}{2}.$$

▲

Consequently the event are dependent. How the computation will change if the first player drops his ball into the urn back?

Independent events

Definition 10. *The random events A_1, A_2 are called independent, if*

$$P(A_1 \cap A_2) = P(A_1)P(A_2).$$

The certain and the impossible events do not depend on any other event.

Example 41.
Are the random events

$$\begin{aligned} A &= \{\textit{in the family with three children there are daughters and sons }\}, \\ B &= \{\textit{in the family with three children there is at least two sons}\} \end{aligned}$$

independent?

▲

Independent events

Theorem 18. If the events A_1 and A_2 are independent, then the events in the pairs $\overline{A_1}$ and A_2, A_1 and $\overline{A_2}$, $\overline{A_1}$ and $\overline{A_2}$ are independent as well.

How to define a larger system of independent events?

There are four balls in the urn with the numbers $0, 1, 2, 3$. There are three players but only one ball is randomly drawn from the urn. If the number on this ball is 0, all three players win a prize. If the number is $1, 2$ or 3 only the first, the second and the third player wins, respectively. Let us denote the events $A_i = \{$the ith player won$\}$. Evidently, if $i \neq j$, then

$$P(A_1) = P(A_2) = P(A_3) = \frac{1}{2},\ P(A_i \cap A_j) = \frac{1}{4},\ P(A_i \cap A_j) = P(A_i)P(A_j).$$

Hence A_i and A_j are independent. For example, A_1 does not depend either on A_2, or A_3. But what about the dependence on both events simultaneously, i. e. on $A_2 \cap A_3$?

Sequence of independent events

Definition 11. *We say that the events in the sequence $A_1, A_2, \ldots, A_n$ are independent, if with all values of $i_1, i_2, \ldots, i_n \in \{0, 1\}$ the equalities*

$$P(A_1^{i_1} \cap A_2^{i_2} \cap \ldots \cap A_n^{i_n}) = P(A_1^{i_1}) \cdot P(A_2^{i_2}) \cdots P(A_n^{i_n}),$$

holds; here $A_i^0 = A_i$, $A_i^1 = \overline{A_i}$.
Note, that not all equalities in the definition are independent.
The independence of the system of events can be also defined using an equivalent requirement: each event A_i and any intersection of finite subsystem (not including A_i) are independent events.

Quick exercise

Example 42. *Two dices*
Experiment – rolling two symmetrical six-sided dices. Are the events

$$\begin{aligned} A &= \{\text{the sum of the points is even}\}, \\ B &= \{\text{the sum of the points exceeds } 6\}. \end{aligned}$$

independent?

▲

Quick exercise

Example 43.
Experiment – rolling two symmetrical six-sided dices. Consider two events: $A = \{\text{the first dice rolled 3 or 4 points}\}$,
$B = \{\text{the sum of the points is less than 8}\}$. Prove, that these events are independent.

▲

System of independent events

Definition 12. *We say that the events of the infinite system $\mathcal{S} = \{A_\lambda : \lambda \in \Lambda\}$ are independent, if events in any finite subsystem are independent.*

Borel-Cantelli lemma

Theorem 19. Let the events in the system $A_1, A_2, \ldots$ be independent, $p_n = P(A_n)$,

$$\sum_n p_n = \infty.$$

Then the probability that an infinite number of events of the system A_n occur is equal to 1.

1.13. Independent trials

Bernoulli trials

Consider the experiments with two outcomes denoted by 0 (failure) and 1 (success). We call them Bernoulli trials. Let the probability of success be p, and probability of failure be $q = 1 - p$.
We have a simple probabilistic space

$$\Omega = \{0, 1\}, \quad P(1) = p, \quad P(0) = q = 1 - p.$$

Independent Bernoulli trials

Suppose the homogeneous Bernoulli trials are repeated n times and the outcomes of the trials are independent. If $n = 3$, then the set of outcomes of the series of trials is

$$\Omega_3 = \{000, 001, 010, 100, 011, 101, 110, 111\}.$$

The probabilities:

$$\begin{aligned}
&P(000) = q \cdot q \cdot q = q^3,\ P(001) = P(010) = P(100) = q \cdot q \cdot p = pq^2, \\
&P(011) = P(101) = P(110) = p^2 q,\ P(111) = p^3, \\
&P(\omega_1 \omega_2 \omega_3) = p^m q^{3-m}, \quad m = 0, 1, 2, 3,
\end{aligned}$$

here $m = \omega_1 + \omega_2 + \omega_3$ is the number of successes in the series.

126 / 369

Bernoulli model

The sequence consists of n homogeneous and independent Bernoulli trials. Let the set of all possible outcomes of the sequence be Ω_n,

$$\Omega_n = \{\omega_1 \omega_2 \ldots \omega_n : \omega_i = 0, 1\}.$$

The probability of the outcome $\omega = \omega_1 \omega_2 \ldots \omega_n$ is defined by

$$\begin{aligned}
P(\omega) &= p^m q^{n-m}, \quad m = \text{the number of successes } = \omega_1 + \ldots + \omega_n. \\
P(A) &= \sum_{\omega \in A} P(\omega).
\end{aligned}$$

We call this probabilistic space the Bernoulli model.

127 / 369

Probability of m successes

Theorem 20. Let the probability of success in the Bernoulli model be p $(0 < p < 1)$, and n the number of trials. Let S_n be the number of successes. Then

$$P(S_n = m) = C_n^m p^m q^{n-m}, \quad q = 1 - p, \quad m = 0, 1, \ldots, n.$$

Examples

Simulations
Bernoulli model
Bernulli model
Random walk
The Galton experiment

Most likely number of successes

Let us denote

$$P_n(m) = C_n^m p^m q^{n-m}, m = 0, 1, \ldots, n.$$

Which probability is the largest? The number of successes $m,$ maximizing $P_n(m),$ is called the most likely number of successes.

Most likely number of successes

$$\begin{aligned} \frac{P_n(m)}{P_n(m-1)} &= \frac{C_n^m p^m q^{n-m}}{C_n^{m-1} p^{m-1} q^{n-m+1}} = \frac{p}{q} \cdot \frac{n-m+1}{m} \\ &= 1 + \frac{p(n+1)-m}{mq} = 1 + \lambda_m \end{aligned}$$

Theorem 21. The largest integer m with the condition $m < (n+1)p$ is the most likely number of successes. If $(n+1)p$ is an integer, then $P_n(m)$ is maximal as $m = (n+1)p$ and $m = (n+1)p - 1.$

Estimation of probability

Let $m > (n+1)p$, we estimate $P(S_n \geq m)$ from above

$$\begin{aligned}
&1 > 1+\lambda_{m+1} > 1+\lambda_{m+2} > \dots\\
&P_n(m+1) = (1+\lambda_{m+1})P_n(m),\\
&P_n(m+2) = (1+\lambda_{m+2})P_n(m+1) < (1+\lambda_{m+1})^2 P_n(m),\\
&P(S_n \geqslant m) = P_n(m) + P_n(m+1) + \dots\\
&P(S_n \geqslant m) < P_n(m)(1+\rho+\rho^2+\dots),\ \rho = 1+\lambda_{m+1},\\
&P(S_n \geqslant m) \leq P_n(m)\frac{1}{1-\rho} = P_n(m)\frac{mq}{m-(n+1)p}
\end{aligned}$$

1.14. The polynomial model

The polynomial model

Let the sample space of the experiment consists of r outcomes. The outcomes will be denoted by numbers $1, 2, \dots, r$. The probabilities of the outcomes are given:

$$\Omega = \{1, 2, \dots, r\}, \quad p_i = P(i), \quad 0 < p_i < 1, \quad p_1 + p_2 + \dots + p_r = 1.$$

The sample space of the series of n independent homogeneous experiments is

$$\Omega_n = \{\omega_1\omega_2\dots\omega_n : \omega_i = 1, 2, \dots, r\}.$$

If the numbers of outcomes $1, 2, \dots, r$ in the sequence $\omega = \omega_1\omega_2\dots\omega_n$ are $m_1, \dots, m_r$ ($m_1 + m_2 + \dots + m_r = n$), then

$$P(\omega) = p_1^{m_1} p_2^{m_2} \cdots p_r^{m_r}.$$

The probabilities of other events are defined by sums of the probabilities of appropriate sequencies ω.

The probabilities in the polynomial model

Theorem 22. Let $\Omega = \{1, 2, \ldots, r\}$ be a sample space of the experiment, $P(i) = p_i, \ i = 1, 2, \ldots, r$ be the probabilities of outcomes, S_n^i – the number of the outcome i in the series of n independent trials, m_i some non-negative integers, $m_1 + m_2 + \ldots + m_r = n$. Then

$$P(S_n^1 = m_1, S_n^2 = m_2, \ldots, S_n^r = m_r) = \frac{n!}{m_1!m_2!\ldots m_r!} p_1^{m_1} p_2^{m_2} \cdots p_r^{m_r}.$$

Quick exercises

Example 44.
Two coins are tossed. The probabilities of the head are equal to 0.4, and 0.6, respectively. What is the probability that in $n = 6$ tosses we shall get exactly twice two heads and exactly twice two tails?

▲

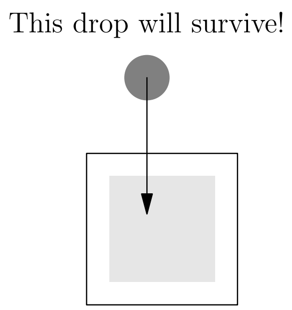
Example 45.
In the rectangle $ABCD$ $n = 5$ points are chosen randomly. What is the probability that the side AB will be nearest one to the three of them, BC and CD to exactly one point.

▲

The first example

Example 46. *Small drops, the large net*
Suppose the stream consists of $n = 1000$ drops falling on the net made up of 100cm $\times$ 100cm squares. A drop is a ball with radius $r = 0,1$ cm. If a drop touches a side of the net, it falls to pieces. What is the probability that exactly $m = 5$ drops will disappear?

▲



The second example

Example 47. *Large drops, the small net*
Let now suppose that the drops are balls with the radius $r = 0,5$ cm, and there are $n = 10000$ drops in the stream. The net is made up of 10cm $\times$ 10cm squares. A lot of drops vanish flying through the net. What is the probability that the number of drops falling into pieces will be at least 1900 but will not exceed 2000?

▲

The Poisson theorem

Theorem 23. Let n be the number of trials in the Bernoulli model, the probability of success depends on the number of trials, let us denote it by p_n. Suppose that n growing unboundedly, p_n vanishes, but there exist a number $\lambda > 0$, such that $np_n \to \lambda$. Then for an arbitrary m we have

$$P(S_n = m) = C_n^m p_n^m (1 - p_n)^{n-m} \to \frac{\lambda^m}{m!}\, \mathrm{e}^{-\lambda}, \quad n \to \infty,$$

here $\mathrm{e} \approx 2,71828$ is the basis of the natural logarithm.

140 / 369

The Moivre-Laplace theorem

Theorem 24. Let p be the probability of success in a Bernoulli trial, n be the number of experiments, and S_n – the number of successes in the sequence of n independent trials. Then as $n \to \infty$ we have for any numbers $a < b$

$$P\Big(a < \frac{S_n - np}{\sqrt{np(1-p)}} < b\Big) \to \Phi(b) - \Phi(a),$$
$$\Phi(v) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{v} \mathrm{e}^{-x^2/2} \mathrm{d}x.$$

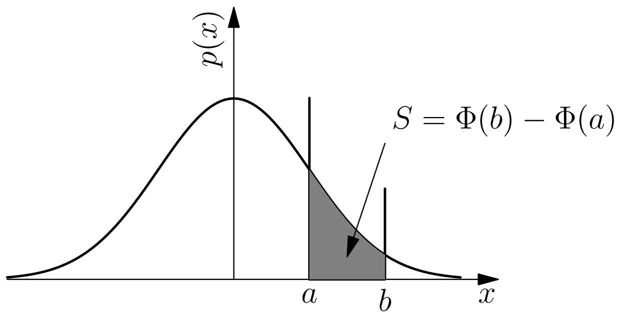
141 / 369

The function $\Phi(x)$

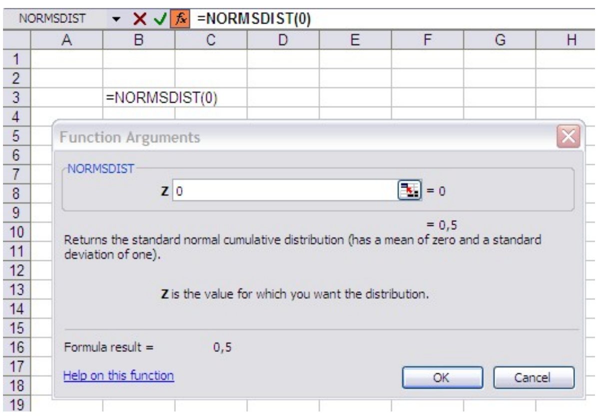
Geometrically the difference $\Phi(b) - \Phi(a)$ is equal to the area under the graph of function

$$p(x) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-x^2/2}$$

restricted by the lines $y = 0, x = a, x = b,$.



Computing the function in Excel



Examples

Example 48.
The bear factory announces the lottery: two specially marked corks from the bottles are exchanged for a prize. There were 4% of the bottles were marked at the factory. What is the probability to win at least one prize buying $n = 50$ bottles?

▲

Example

Example 49.
The test of the exam consists of 10 questions, the possible answers are yes and no. The grade is equal to the number of correct answers. The students use the "probabilistic method": they choose the answers randomly. What is the probability that among $n = 700$ students using this method exactly three of them will get grade "8"?

▲

Example

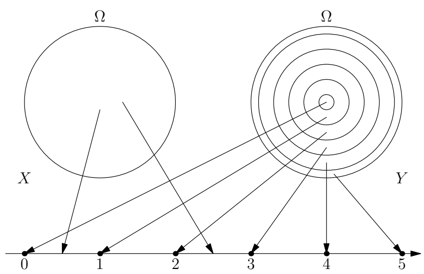
Example 50.
The probability that the student will finish the studies successfully is equal to $0,6$. What the smallest numbers of students should be such that with the probability equal to $0,8$ at least 200 students will be successful?
▲

Example 51.
What is the smallest number of rolls of the symmetrical dice, such that the probability to get at least 100 sixies should be larger than $0,7$?
▲

2.1. Random variables

Examples

Let X be the distance from the point chosen randomly in the circle with radius $r = 5$ to its center, and Y – the distance rounded to the integer.



$$P(X < 3) = \frac{\pi \cdot 3^2}{\pi \cdot 5^2} = 0,36, \quad P(Y < 3) = \frac{\pi \cdot 2,5^2}{\pi \cdot 5^2} = 0,25,$$

$$P(X = 3) = 0, \quad P(Y = 3) = \frac{\pi \cdot (3,5^2 - 2,5^2)}{\pi \cdot 5^2} = 0,24.$$

Definition

Definition 13. *A function $\xi : \Omega \to \mathbb{R}$ is called random variable if for every Borelian set $B \in \mathcal{B}$*

$$\{\omega : \xi(\omega) \in B\} = \xi^{-1}(B) \in \mathcal{A}.$$

The random vectors

Definition 14. *A function $\xi : \Omega \to \mathbb{R}^n$*

$$\xi(\omega) = \langle \xi_1(\omega), \xi_2(\omega), \dots, \xi_n(\omega) \rangle,$$

where ξ_i are random variables is called a random vector.

Without Borelian sets

Theorem 25. A function $\xi : \Omega \to \mathbb{R}$ is a random variable if and only if for every number x

$$\{\omega : \xi(\omega) < x\} \in \mathcal{A}.$$

151 / 369

Functions of random variables

Definition 15. *A function $f : \mathbb{R} \to \mathbb{R}$ is called Borelian if for any Borelian set $B \in \mathcal{B}$*

$$f^{-1}(B) \in \mathcal{B}.$$

Theorem 26. *Let $\xi : \Omega \to \mathbb{R}$ be some random variable and $f : \mathbb{R} \to \mathbb{R}$ some Borelian function. Then $\eta = f(\xi)$ is a random variable too.*

152 / 369

Algebraic operations with random variables

Theorem 27. Let $\xi, \eta : \Omega \to \mathbb{R}$ be the random variables. Then $\xi \pm \eta,\ \xi \cdot \eta,\ \xi/\eta$ (if $\eta \neq 0)$ are random variables too.

153 / 369

Analytic operations with random variables

Theorem 28. Let $\xi_n : \Omega \to \mathbb{R}$ $(n = 1, 2, \dots)$ be the random variables. Then the functions

$$\begin{aligned} \eta_1 &= \inf_n \xi_n, \quad \eta_2 = \sup_n \xi_n, \\ \eta_3 &= \liminf_n \xi_n, \quad \eta_4 = \limsup_n \xi_n \end{aligned}$$

are also the random variables.

154 / 369

Distribution function

Definition 16. *Let $\xi : \Omega \to \mathbb{R}$ – be a random variable. The function $F_\xi : \mathbb{R} \to [0,1]$,*

$$F_\xi(x) = P(\xi < x)$$

is called (cumulative) distribution function of the random variable ξ.

Definition 17. *Let $\xi : \Omega \to \mathbb{R}^m$ – be a random vector, $\xi = \langle \xi_1, \dots, \xi_m \rangle$. The function $F_\xi : \mathbb{R}^m \to [0,1]$,*

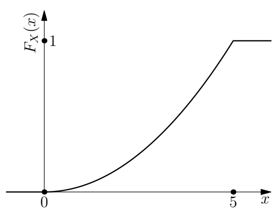
$$F_\xi(x_1, \dots, x_m) = P(\xi_1 < x_1, \dots, \xi_m < x_m).$$

is called the distribution function of the random vector.

Example

Let the radius of the circle be $r = 5$. A point of the circle is chosen randomly. Let X be a distance from the point chosen to the center of the circle. The random variable X takes the values in the interval $[0; 5]$.

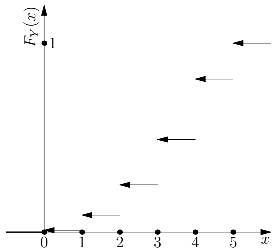
$$F_X(x) = \begin{cases} 0, & \text{if } x \le 0, \\ \frac{x^2}{25}, & \text{if } 0 < x < 5, \\ 1, & \text{if } x \ge 5. \end{cases}$$



The second example

Suppose now that we measure the distance between the point randomly chosen in the circle and the the center and round the result to the whole number according to custom rule of rounding. We get this way a random variable Y, which can take six different values.

$y =$	0	1	2	3	4	5
$P(Y = y) =$	$0,01$	$0,08$	$0,16$	$0,24$	$0,32$	$0,19$



158 / 369

The properties of distribution function

Theorem 29. Let F_ξ be the distribution function of the random variable ξ. Then :

- F_ξ is non-decreasing function;
- F_ξ is left-continuous;
- $\lim\limits_{x\to-\infty} F_\xi(x) = 0, \lim\limits_{x\to\infty} F_\xi(x) = 1.$

159 / 369

Discrete random variables

Definition 18. *If a random variable (vector) ξ takes the values from the finite or infinite but countable set, it is called discrete random variable (vector).*

Let X takes the values from the finite set. The information on the probabilities can be displayed in the table:

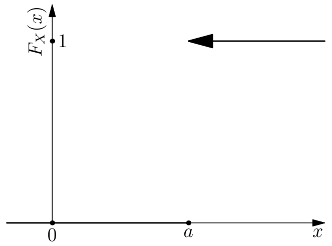
$$\begin{array}{c|c|c|c|c} x = & x_1 & x_2 & x_3 & \dots \\ \hline P(X=x) = & p_1 & p_2 & p_3 & \dots \end{array}, \quad p_i = P(X = x_i).$$

Discrete random variable

Theorem 30. The distribution function of the random variable ξ has jumps at points corresponding to the values of random variable.
If x is a value of the discrete random variable, then at this point the distribution function has a jump equal to $F_\xi(x+0) - F_\xi(x) = P(\xi = x)$.

Degenerated random variable

Definition 19. *If there exist a value a, such that for a random variable X we have $P(X = a) = 1$, then X is called degenerated random variable.*



We may wonder whether such a variable is not "random" at all; but there are some subtleties here: the variable can take some values with zero probability. Example: integer part of the distance from the point randomly chosen in the unit circle to its center. The value will always be zero, except as we choose the point on the circumference.

Binomial random variable

Let n be the number of trials in the Bernoulli model, p – the probability of success, ξ_n – number of successes in the series of n trials. The variable takes the values in the set $\{0, 1, \ldots, n\}$,

$$P(\xi_n = m) = C_n^m p^m (1-p)^{n-m}.$$

We shall denote $\xi_n \sim \mathcal{B}(n, p)$.

Binomial variable

Let us associate with the ith trial of Bernoulli model the random variable:

$$X_i = \begin{cases} 1, & \text{if there will be success in the } i\text{th trial,} \\ 0, & \text{if the } i\text{th trial is unsuccesful.} \end{cases}$$

Then

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

165 / 369

Geometrical random variable

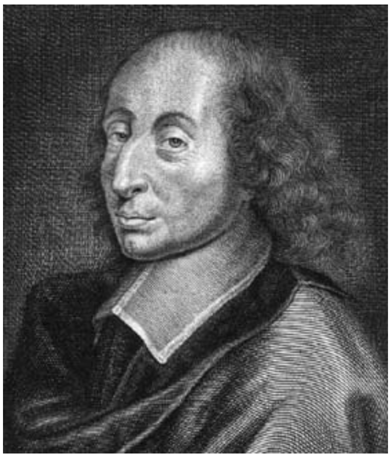
Definition 20. *A random variable X taking the values $m = 1, 2, \ldots$ with the probabilities*

$$P(X = m) = q^{m-1}p, \quad m = 1, 2, \ldots, \quad 0 < p < 1, \quad q = 1 - p,$$

is called geometrical. We shall write $X \sim \mathcal{G}(p)$.

166 / 369

Pascal's random variable



The Bernoulli trials are repeated until n successes are obtained. Let θ_n be the number of trials, $\eta_n = \theta_n - n$. Then η_n – is a number of failures in the series of trials.

Values and probabilities:

$$P(\eta_n = s) = C^s_{n+s-1} p^n (1-p)^s.$$

We say that η_n is a Pascal's random variable and write $\eta_n \sim \mathcal{B}^-(n, p)$.

167 / 369

Poisson's random variable

Let a random variable ξ takes the non-negative integer values with the probabilities

$$P(\xi = m) = \frac{\lambda^m}{m!} e^{-\lambda},$$

here $\lambda > 0$. The random variable ξ is called Poisson's random variable with the parameter λ. We write $\xi \sim \mathcal{P}(\lambda)$.

168 / 369

Poisson's theorem

Theorem 31. Let n denote the number of Bernoulli trials with the probability of success equal to p_n and S_n be the number of successes in the series of trials. If there exists a positive number λ, such that $np_n \to \lambda$, as $n \to \infty$, then for every $m = 0, 1, \ldots$

$$P(\xi_n = m) = C_n^m p_n^m q_n^{n-m} \to P(X = m), \quad X \sim \mathcal{P}(\lambda).$$

here $q_n = 1 - p_n$.

Hypergeometrical random variables

Let there are n white and m black balls in the urn and u $(u < n + m)$ balls are drawn randomly. Denote by X the number of white balls in the drawn set. Then X is a discrete random variable taking values with the probabilities

$$P(X = v) = \frac{C_n^v C_m^{u-v}}{C_{n+m}^u},$$
$$\max(u - m, 0) \leq v \leq \max(u, m).$$

The random variable X is called **hypergeometrical.**

Hypergeometrical random variables

Suppose the balls are drawn in succession and

$$X_i = \begin{cases} 1, & \text{if the } i\text{th ball is white,} \\ 0, & \text{if the } i\text{th ball is black,} \end{cases} \qquad i = 1, 2, \ldots, u.$$

Then

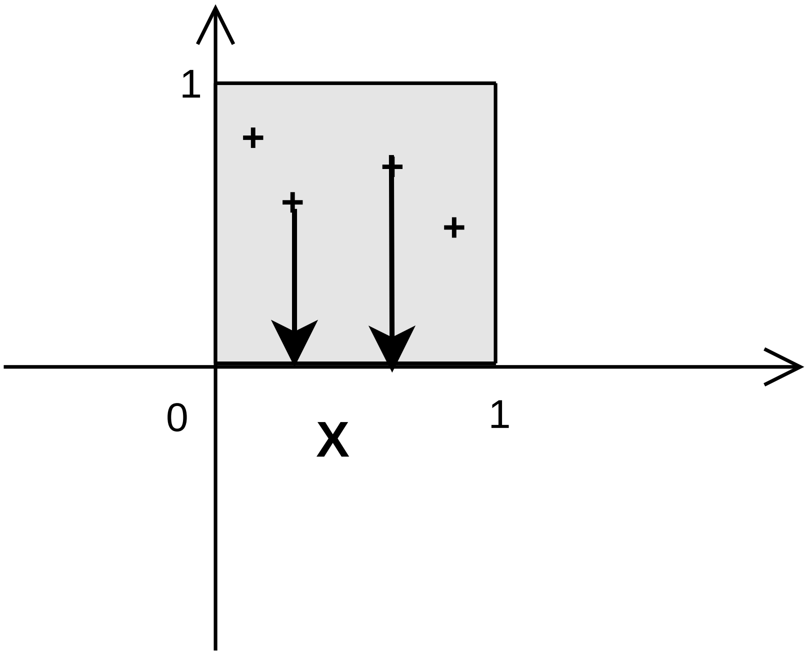
$$X = X_1 + X_2 + \ldots + X_u.$$

2.3. Absolutely continuous random variables

Continuous random variables

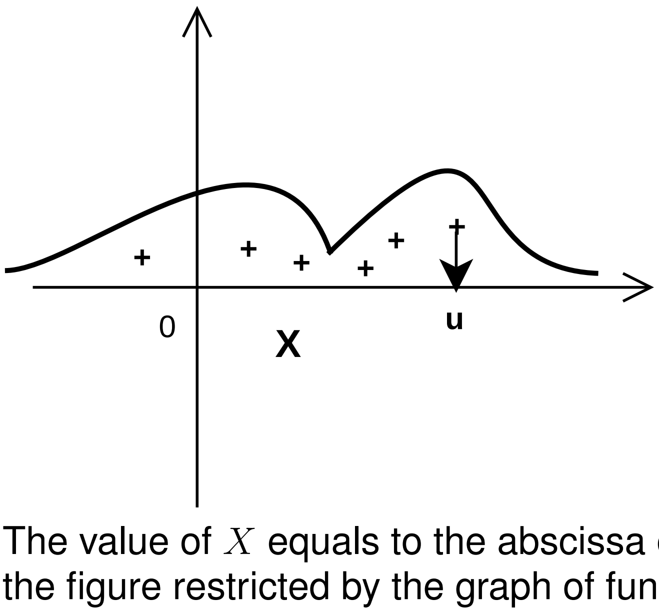
Definition 21. *If the distribution function F_ξ of the random variable (random vector) ξ is continuous, we call the random variable (random vector) continuous.*

Uniformly distributed random variable



The value of X equals to the abscissa of the point randomly chosen in the unite square. It takes values in the interval $[0; 1]$. All values are „equally likely".

Absolutely continuous random variables



The value of X equals to the abscissa of the point randomly chosen in the figure restricted by the graph of function $p(x)$ and the abscissa axis. The area of the figure is equal to 1.

$$P(X < u) = \int_{-\infty}^{u} p(x)\mathrm{d}x$$

Absolutely continuous random variables

Definition 22. *The random variable ξ is called absolutely continuous if there exists an integrable non-negative function $p_\xi(s)$, such that*

$$F_\xi(x) = \int_{-\infty}^{x} p_\xi(s)ds.$$

The function $p_\xi(s)$ is called density function of random variable ξ.

176 / 369

Absolutely continuous random vectors

Definition 23. *The random vector $\xi : \Omega \to \mathbb{R}^m$ is called absolutely continuous if there exists an integrable non-negative function $p_\xi(s_1, \dots, s_m)$, such that*

$$F_\xi(x_1, \dots, x_m) = \int_{\{s_1<x_1,\dots,s_m<x_m\}} p_\xi(s_1, \dots, s_m)ds_1 \cdot \ldots \cdot ds_m.$$

The function $p_\xi(s_1, \dots, s_m)$ is called density function of the random vector ξ.

177 / 369

Properties of density functions

- $p_\xi(x) \geq 0.$
- At almost all points $F'_\xi(x) = p_\xi(x).$
- $$\int_{-\infty}^{\infty} p_\xi(u)\mathrm{d}u = 1.$$

178 / 369

Uniformly distributed random variables

Definition 24. *We say that the random variable ξ is uniformly distributed in the interval $[a; b]$, if it has the density function $p_\xi(x)$ equal to 0 if $x \notin [a; b]$, and for all $x \in [a; b]$ $p_\xi(x)$ takes the same value, i.e. $p_\xi(x) = c$.*

Because of the integral of the density function is equal to 1, we get

$$1 = \int_{-\infty}^{\infty} p_\xi(x)dx = c\int_a^b dx = c(b-a), \quad c = \frac{1}{b-a}.$$

179 / 369

Exponential random variables

Example 52.
Suppose a fly flew in the room and seeks desperately for the open window to exite. Denote by X the time necessary to find the way. What is the probability that at least t seconds a fly will stay in the room, i.e. what is the value of $P(X \geqslant t)$?

▲

Exponential random variables

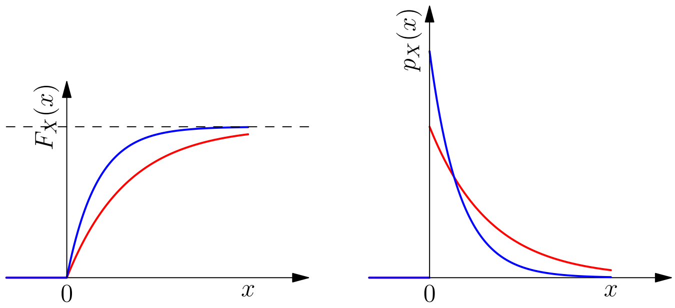
Definition 25. *If the density function of the random variable X is*

$$p_X(x) = \begin{cases} 0, & \text{if } x < 0, \\ \lambda \mathrm{e}^{-\lambda x}, & \text{if } x > 0, \end{cases}$$

where $\lambda > 0$, we say that X is an exponential random variable.
We shall write $X \sim \mathcal{E}(\lambda)$.

Exponential random variables

Graphs of distribution and density functions of exponential random variables



Example

Example 53.
The surviving time of a soap bubble is an exponential random variable. There were created 1000 soup bubbles and after one minute only 450 were not blown up. What is the probability that a soap bubble survives two minutes? What the smallest number of soap bubbles should be that with probability $0,9$ after three minutes we would have at least 300 bubbles survived?

▲

The Poisson process

Suppose you turned on your phone and wait for the first message. Let T_1 be the waiting time. What is the probability $P(T_1 \geqslant t)$ i.e. what is the probability that you should wait at least t seconds?

Let us divide the time interval $[0; t]$ into the small and equal pieces and make a conjecture that in a short interval of length $1/n$ we can receive only one message with the probability p_n, where $p_n,\ np_n \to \lambda$, as $n \to \infty$. We derive then that

$$P(T_1 \geqslant t) = \mathrm{e}^{-\lambda t}, \quad \text{i.e.} \quad T_1 \sim \mathcal{E}(\lambda).$$

The Poisson process

Let us denote by X_t the number of messages received in the time interval $[0; t]$. Then X_t is a discrete random variable and we know only the probability:

$$P(X_t = 0) = P(T_1 \geqslant t) = \mathrm{e}^{-\lambda t}.$$

We can find also

$$P(X_t = m) = \frac{(\lambda t)^m}{m!}\mathrm{e}^{-\lambda t}, \quad m = 0, 1, 2, \ldots.$$

$X_t \sim \mathcal{P}(\lambda t)$. We have an infinite system of Poisson's random variables. It is called the Poisson process.

The Poisson process

We considered the random value T_1 – the moment when the first message arrives. Let now for $k \geqslant 1,$ the random value T_k means the moment, when the kth message is received.

$$T_1 \leqslant T_2 \leqslant T_3 \leqslant \ldots \leqslant T_{k-1} \leq T_k.$$

$$\begin{aligned}
&P(T_k \geqslant t) = P(X_t < k) \\
&= P(X_t = 0) + \ldots + P(X_t = k-1) \\
&= \mathrm{e}^{-\lambda t} + \frac{(\lambda t)^1}{1!}\mathrm{e}^{-\lambda t} + \ldots + \frac{(\lambda t)^{k-1}}{(k-1)!}\mathrm{e}^{-\lambda t}.
\end{aligned}$$

186 / 369

Poisson process

The random variable T_k has the distribution function

$$F_{T_k}(t) = 1 - \mathrm{e}^{-\lambda t} - \frac{(\lambda t)^1}{1!}\mathrm{e}^{-\lambda t} + \cdots + (-1)^{k-1}\frac{(\lambda t)^{k-1}}{(k-1)!}\mathrm{e}^{-\lambda t}, \quad t \geqslant 0.$$

The density function

$$p_{T_k}(t) = F'_{T_k}(t) = \lambda \cdot \frac{(\lambda t)^{k-1}}{(k-1)!}\mathrm{e}^{-\lambda t} = \frac{\lambda^k t^{k-1}}{(k-1)!}\mathrm{e}^{-\lambda t}.$$

187 / 369

Gamma random variable

Definition 26. *If the density function of the random variable X is*

$$p_X(t) = \begin{cases} 0, & \text{if } t < 0, \\ \dfrac{\lambda^k t^{k-1}}{(k-1)!} e^{-\lambda t}, & \text{if } t > 0, \end{cases}$$

here $\lambda > 0, k \geq 1,$ then X is called gamma random variable.
We shall write $X \sim \Gamma(k, \lambda)$.

Gamma random variable

The gamma random variable T_k can be expressed via more simple random variables.
Let us denote by $T_{0|1}$ the waiting time of the first message, $T_{1|2}$ – the waiting time of the second message an so forth. Then

$$T_k = T_{0|1} + T_{1|2} + \ldots + T_{k-1|k}.$$
$$T_{0|1} \sim \mathcal{E}(\lambda), \ \ldots, \ T_{k-1|k} \sim \mathcal{E}(\lambda).$$

Pareto random variables

Definition 27. *Let the density function of the random variable X be*

$$p_X(x) = \begin{cases} 0, & \text{if } x < 1, \\ \frac{\alpha}{x^{\alpha+1}}, & \text{if } x \geq 1, \end{cases}$$

here $\alpha > 0$. The random variable X is called Pareto random variable. We shall write $X \sim \mathcal{P}ar(\alpha)$.

Standard normal variable

Definition 28. *The random variable X with the density function*

$$p(x) = \frac{1}{\sqrt{2\pi}} \, \mathrm{e}^{-x^2/2}$$

is called standard normal variable. We shall write $X \sim \mathcal{N}(0, 1)$.

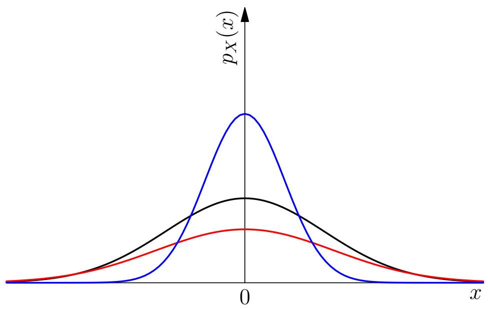
The normal random variables

Definition 29. *The random variable X with the density function*

$$p_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\,\mathrm{e}^{-(x-\mu)^2/2\sigma^2},$$

is callled normal random variable. We shall denote $X \sim \mathcal{N}(\mu, \sigma^2)$.

The normal variables



The graphs of density functions of normal random variables

The Moivre-Laplace theorem

Theorem 32. Let the probability of success in the Bernoulli model be p, and S_n denote the number of successes in the sequence of n trials. Then for every x, as $n \to \infty$, we have

$$P\left(\omega : \frac{S_n(\omega) - np}{\sqrt{np(1-p)}} < x\right) \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt.$$

This statement is called the integral Moivre-Laplace theorem.

The normal random vector

Definition 30. *We say that the random vector $\xi : \Omega \to \mathbb{R}^n$ is standard normal, if it has the density function*

$$p_\xi(x_1, \ldots, x_n) = \frac{1}{(2\pi)^{n/2}} \exp\left\{ -\frac{1}{2} \sum_{k=1}^{n} x_k^2 \right\}.$$

Quantiles and the critical values

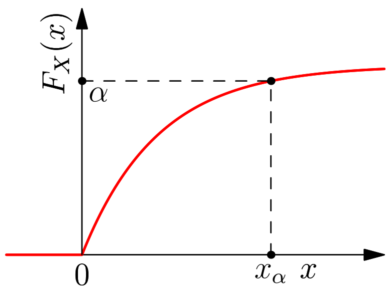
Definition 31. *Let the random variable X be continuous and $0 < \alpha < 1$. The smallest solution of the equation*

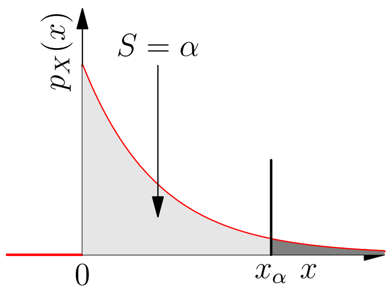
$$F_X(x) = \alpha$$

is called α-quantile of the random variable X; we denote it by u_α. The quantity u_α is also called $(1-\alpha)$-critical value.

Quantiles

If the random variable X has the density function $p(x)$, and u_α is the α-quantile, then the area below the graph of the density function (and above the abscissa axis) to the left of the line $x = u_\alpha$ is α, and to the right is $1-\alpha$.





Quick exercises

Example 54.
The random variable X is exponential, $X \sim \mathcal{E}(\lambda)$. Its $0,5$-quantile is equal to 3. Find the value of λ.

▲

Example 55.
Find the 0.5-quantile and $0,2$-critical value of the random variable $X \sim \mathcal{P}ar(3)$.

▲

Example 56.
The random variable X is standard normal, $F_X(0,5244) = \Phi(0,5244) = 0,7$. What should be the value of a, such that the $0,7$-quantile of the random variable $Y = 2X + a$ would be equal to 3?

▲

2.4. Probabilities and the density functions

Computing probabilities

Let ξ be an absolutely continuous random variable having the density function p_ξ. If B is a Borelian set, then

$$P(\xi \in B) = \int_B p_\xi(u)du.$$

Let $\xi = \langle \xi_1, \ldots, \xi_m \rangle, \xi : \Omega \to I\mathbb{R}^m$, be an absolutely continuous random vector with the density function p_ξ. Then for any Borelian set $B \in \mathcal{B}^m$

$$P(\xi \in B) = \int_B p_\xi(u_1, u_2, \ldots, u_m)du_1 du_2 \ldots du_m.$$

Density function of a random vector

Theorem 33. Let the absolutely continuous random vector $\xi = \langle \xi_1, \ldots, \xi_m \rangle$ has the density function $p_\xi(u_1, \ldots u_m)$. Then the random vector $\xi' = \langle \xi_1, \ldots, \xi_{m-1} \rangle$ is also absolutely continuous and its density function is

$$p_{\xi'}(u_1, \ldots, u_{m-1}) = \int_{-\infty}^{\infty} p_\xi(u_1, u_2, \ldots, u_m) \mathrm{d}u_m.$$

The density function of the random variable $\eta = f(\xi)$

Theorem 34. Let $\xi : \Omega \to \mathbb{R}$ be an absolutely continuous random variable and $f : \mathbb{R} \to \mathbb{R}$ some monotone and differentiable function. Then the random variable $\eta = f(\xi)$ has the density function

$$p_\eta(t) = p_\xi(f^{-1}(t)) \cdot |f'(\phi^{-1}(t))|^{-1}.$$

Note that in practice we can find the density function computing $F_\eta(x)$ and differentiating this function.

The normal variables

Let $\xi \sim \mathcal{N}(0,1)$, and $\eta = a + \sigma\xi$, $\sigma \neq 0$. Then the density function of the random variable η is

$$p_\eta(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big\{ -\frac{1}{2\sigma^2}(x-a)^2 \Big\}.$$

We say that η is the normal random variable with parameters a, σ^2 and write $\eta \sim \mathcal{N}(a, \sigma^2)$.

203 / 369

2.5. Independent random variables

204 / 369

Independent random variables

Definition 32. *The random variables ξ_1, ξ_2 are independent if with any Borelian sets B_1, B_2*

$$P(\xi_1 \in B_1, \xi_2 \in B_2) = P(\xi_1 \in B_1)P(\xi_2 \in B_2).$$

How to define the system of independent random variables?

205 / 369

Independent random vectors

Definition 33. *The random vectors* $\xi_1, \xi_2 : \Omega \to \mathbb{R}^n$ *are independent if with any Borelian sets* $B_1, B_2 \in \mathcal{B}(\mathbb{R}^n)$

$$P(\xi_1 \in B_1, \xi_2 \in B_2) = P(\xi_1 \in B_1)P(\xi_2 \in B_2).$$

Distribution functions

Theorem 35. The random variables $\xi, \eta : \Omega \to \mathbb{R}$ are independent if and only if

$$P(\xi < u, \eta < v) = P(\xi < u)P(\eta < v).$$

Theorem 36. The random variables $\xi, \eta : \Omega \to \mathbb{R}$ are independent if and only if the distribution function of the random vector $\zeta = \langle \xi, \eta \rangle$ satisfies

$$F_\zeta(u, v) = F_\xi(u)F_\eta(v),$$

here $\zeta = \langle \xi, \eta \rangle$.

The discrete random variables

Theorem 37. The discrete random variables $\xi, \eta : \Omega \to \mathbb{R}$ are independent if and only for all values of x, y

$$P(\xi = x, \eta = y) = P(\xi = x)P(\eta = y).$$

Absolutely continuous random variables

Theorem 38. Let $\xi, \eta : \Omega \to \mathbb{R}$ be the absolute continuous independent random variables with the density functions p_ξ, p_η. Then the random vector $\zeta = \langle \xi, \eta \rangle$ is also absolutely continuous and

$$p_\zeta(u_1, u_2) = p_\xi(u_1)p_\eta(u_2).$$

Let the random vector $\zeta = \langle \xi, \eta \rangle$ be absolutely continuous with the density function p_ζ.
If for densities of components p_ξ, p_η the equality $p_\zeta(u_1, u_2) = p_\xi(u_1)p_\eta(u_2)$ is satisfied, then the random variables ξ, η are independent.

Independent random variables and Borelian functions

Theorem 39. Let $\xi_1, \xi_2 : \Omega \to \mathbb{R}^n$ be independent random vectors and $f_1, f_2 : \mathbb{R}^n \to \mathbb{R}^m$ two Borelian functions. Then the random vectors $\eta_1 = f_1(\xi_1), \eta_2 = f_2(\xi_2)$ are independent, too.

Sum of independent random variables

Theorem 40. Let ξ_1, ξ_2 be continuous independent random variables with the density function p_{ξ_1}, p_{ξ_2}. Then the random variable $\eta = \xi_1 + \xi_2$ is also independent with the density function

$$\begin{aligned} p_{\xi_1+\xi_2}(u) &= \int_{-\infty}^{\infty} p_{\xi_1}(v) p_{\xi_2}(u-v) dv \\ &= \int_{-\infty}^{\infty} p_{\xi_2}(v) p_{\xi_1}(u-v) dv. \end{aligned}$$

2.6. Expected value of the discrete random variables

Definition

Definition 34. *Let ξ be a discrete random variable and the series*

$$\sum_x xP(\xi = x)$$

converges absolutely. We denote the sum of the series by $\mathbf{E}[\xi]$ and call this number expected value (expectation) of the random variable ξ.
Note. *If the discrete random variable ξ has the expected value, then the expected value of the random variable $|\xi|$ is also defined.*

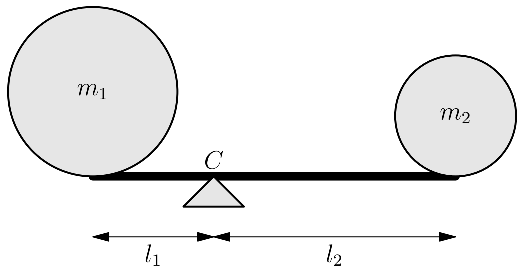
Expectation of the bounded variable

If the values of the discrete random variable ξ are bounded by some constants from above and below, then the expected value of the random variable exists.

The following statement can also be proved.

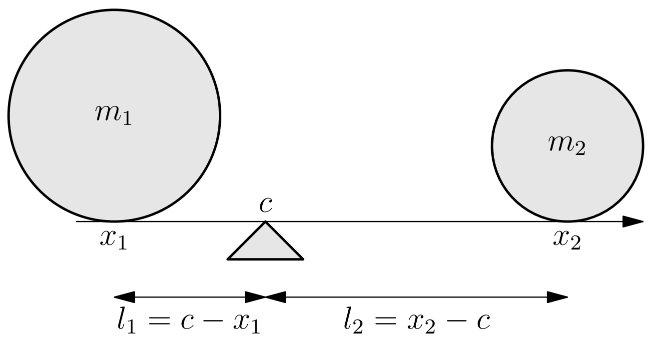
Theorem 41. Let ξ and η be two discrete random variables, $|\xi| \leq \eta$ and for the random variable η the expected value exists. Then the expected value exists also for the random variable ξ.

Expected value as center of gravity



The point C is the center of gravity of the system consisting of two bodies. Suppose the bodies are put on the real line at the points with the coordinates x_1, x_2. Let c be the coordinate of the center of gravity.

Expected value as center of gravity



$$m_1(c - x_1) = m_2(x_2 - c), \quad (x_1 - c)m_1 + (x_2 - c)m_2 = 0,$$
$$c = \frac{x_1 m_1 + x_2 m_2}{m_1 + m_2}.$$

Properties of the expectation

Theorem 42. Let X be a discrete random variables with the values $x_1, x_2, \ldots,$ and $Y = f(X)$ is the new random variable. If the expected value of Y exists, then

$$\mathbf{E}[Y] = \sum_i f(x_i) P(X = x_i).$$

Theorem 43. Let X be a discrete random variable having the expected value and a is an arbitrary number. Then for the random variable $Y = aX$ we have:

$$\mathbf{E}[a \cdot X] = a \cdot \mathbf{E}[X].$$

Additivity of expectation

Theorem 44. Let X and Y be the discrete random variables having the expected values. Then for the sum $X + Y$ we have:

$$\mathbf{E}[X + Y] = \mathbf{E}[X] + \mathbf{E}[Y].$$

Theorem 45. Let $X_1, X_2, \ldots, X_n$ be the discrete random variables having the expected values. Then

$$\mathbf{E}[X_1 + X_2 + \ldots + X_n] = \mathbf{E}[X_1] + \mathbf{E}[X_2] + \ldots + \mathbf{E}[X_n].$$

Independence and expectation

Theorem 46. Let X and Y be independent discret random value having the expected values. Then for the product $X \cdot Y$ we have:

$$\mathbf{E}[X \cdot Y] = \mathbf{E}[X] \cdot \mathbf{E}[Y].$$

Expected value of the binomial random variable

If $X \sim \mathcal{B}(n, p)$, then $\mathbf{E}[X] = np$.

Expected value of the Poisson random variable

If $X \sim \mathcal{P}(\lambda)$, then $\mathbf{E}[X] = \lambda$.

Expected value of the geometrical random variable

$$\text{If } X \sim \mathcal{G}(p), \text{ then } \mathbf{E}[X] = \frac{1}{p}.$$

Expected value of the Pascal random variable

If $X \sim \mathcal{B}^{-}(n, p)$, then

$$P(X = m) = C_{n+m-1}^{m} p^n q^m,$$

$$\mathbf{E}[X] = n\frac{q}{p}.$$

Expected value of the hipergeometrical random variable

Let there are m whites and n black balls in the urn. We draw randomly and without replacement u $(u \leq m+n)$ balls. The value of the variable X is the number of white balls in the set of balls drawn from the urn.
The expected value of the random variable is:

$$\mathbf{E}[X] = \sum_{v} v \cdot \frac{C_m^v C_n^{u-v}}{C_{m+n}^u} = u \cdot \frac{m}{m+n}.$$

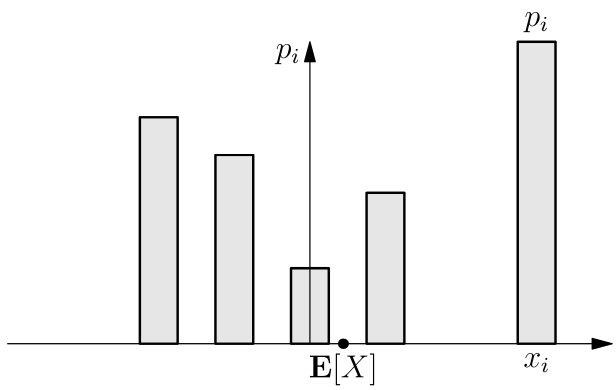
Note. The simplest way to compute the expectation X: use the expression and additivity property

$$X = X_1 + X_2 + \ldots + X_u,$$

where the random variable X_i takes the value $1,$ if the ith ball is white and 0 otherwise.

Expectation as center of gravity

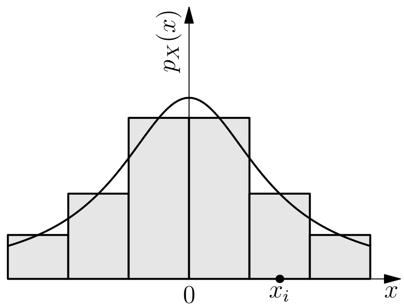
The expected value of the discrete random variable X can be interpreted as the coordinate of the center of gravity.



Expectation as center of gravity

Let X be a continuous random variable with the density function $p_X(x)$. Suppose that on the real line was put not the single weights but a figure, restricted from above by the graph of density function.
Where should be placed a support point for to keep the figure in balance?

Expectation as center of gravity



$$\mathbf{E}[X] \approx \sum_i x_i p_X(x_i)(x_{i+1} - x_i) \to \int_{-\infty}^{\infty} x p_X(x)\mathrm{d}x.$$

228 / 369

Definition

Definition 35. *Let the absolutely continuous random variable X has the density function $p_X(x)$. The expected value of the random variable X (if it exists), is the number*

$$\mathbf{E}[X] = \int_{-\infty}^{\infty} x p_X(x)\mathrm{d}x.$$

Note. For the existence of the expectation it is necessary that the integral converges absolutely.

229 / 369

Expected value of the uniformly distributed variable

Let the random variable X is uniformly distributed in the interval $[a;b]$, i.e. $X \sim \mathcal{T}([a,b])$. Then

$$p_X(x) = \begin{cases} \frac{1}{b-a}, & \text{if } x \in [a;b]; \\ 0, & \text{if } x \notin [a,b]. \end{cases}$$

$$\text{If } X \sim \mathcal{T}([a,b]), \text{ then } \mathbf{E}[X] = \frac{a+b}{2}.$$

Expected value of the exponential random variable

The exponential random variable $X \sim \mathcal{E}(\lambda)$ has the density function

$$p_X(x) = \begin{cases} \lambda \mathrm{e}^{-\lambda x}, & \text{jei } x \geq 0, \\ 0, & \text{jei } x < 0. \end{cases}$$

$$\text{If } X \sim \mathcal{E}(\lambda), \text{ then } \mathbf{E}[X] = \frac{1}{\lambda}.$$

Gamma random variables

Definition 36. *If the random variable X has the density function*

$$p_X(t) = \begin{cases} 0, & \textit{if } t < 0, \\ \dfrac{\lambda^k t^{k-1}}{(k-1)!} \mathrm{e}^{-\lambda t}, & \textit{if } t > 0, \end{cases}$$

here $\lambda > 0, k \geq 1,$ then X is called gamma random variable. We denote it by $X \sim \Gamma(k, \lambda)$.

Expected value of the gamma variable

Let T_k be the gamma random variable, $T_k \sim \Gamma(k, \lambda)$. It can be interpreted as the waiting time for k phone calls or messages.
Let $T_{0|1}$ be the waiting time of the first message, $T_{1|2}$ – the waiting time of the second message, and so forth. Then

$$T_k = T_{0|1} + T_{1|2} + \cdots + T_{k-1|k}.$$
$$T_{0|1} \sim \mathcal{E}(\lambda), \; \ldots, \; T_{k-1|k} \sim \mathcal{E}(\lambda).$$

Expected value of the normal random variables

Let $X \sim \mathcal{N}(0,1)$, then the density function and expected value is:

$$p_X(x) = \frac{1}{\sqrt{2\pi}}\,\mathrm{e}^{-x^2/2}, \quad \mathbf{E}[X] = 0.$$

Let $Y = \sigma X + \mu$, then Y is a normal random variable too: $Y \sim \mathcal{N}(\mu, \sigma^2)$. The expected value is

$$\mathbf{E}[Y] = \mathbf{E}[\sigma X + \mu] = \mathbf{E}[\sigma X] + \mathbf{E}[\mu] = \sigma \mathbf{E}[X] + \mu = \mu.$$

234 / 369

Computing the expected value

Theorem 47. Let the random variable X has the density function $p_X(x)$, and let $f(x)$ be a Borelian function. The random variable $Y = f(X)$ has the expected value if and only if the integral

$$\int_{-\infty}^{\infty} |f(x)| p_X(x) \mathrm{d}x$$

converges. Then

$$\mathbf{E}[Y] = \int_{-\infty}^{\infty} f(x) p_X(x) \mathrm{d}x.$$

235 / 369

Generalized theorem

Theorem 48. Let $X = \langle X_1, X_2, \dots, X_m \rangle$ be a random vector with the density function $p_X(x_1, x_2, \dots, x_m)$, and $f(x_1, x_2, \dots, x_m)$ – a Borelian function with real values. The random variable $Y = f(X)$ has the expected value if and oly if the integral

$$\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} |f(x_1, x_2, \dots, x_m)| p_X(x_1, x_2, \dots, x_m) \mathrm{d}x_1 \dots \mathrm{d}x_m$$

converges. Then

$$\mathbf{E}[Y] = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(x_1, \dots, x_m) p_X(x_1, \dots, x_m) \mathrm{d}x_1 \dots \mathrm{d}x_m.$$

236 / 369

Example

Example 57.
The durations of the phone conversations of two people are the random variables $X_1 \sim \mathcal{T}([0; a]),\ X_2 \sim \mathcal{T}([0; b].)$ Find the expected values of the random variables

$$\mathbf{E}[\min(X_1, X_2)], \mathbf{E}[\max(X_1, X_2)],$$
$$\mathbf{E}[\min(X_1, X_2) \cdot \max(X_1, X_2)].$$

▲

237 / 369

Uniform convergence

Definition 37. *Let $\xi_n, \xi : \Omega \to \mathbb{R}$ be the random variables. If for every $\delta > 0$ there exists some number $n(\delta)$, such that the inequality*

$$|\xi_n(\omega) - \xi(\omega)| < \delta$$

holds as $n > n(\delta)$ for all $\omega \in \Omega$, then we say that ξ_n converges uniformly to ξ.

Sequences of converging random variables

Let ξ be a random variable and $\epsilon > 0$. Define the discrete random variable $\xi^\epsilon(\omega)$ as follows:

$$\text{if } \xi(\omega) \in [n\epsilon, n\epsilon + \epsilon), \quad \text{then} \quad \xi^\epsilon(\omega) = n\epsilon.$$

Then

$$\xi(\omega) - \epsilon \le \xi^\epsilon(\omega) \le \xi,$$
$$P(\xi^\epsilon = n\epsilon) = F_\xi(n\epsilon + \epsilon) - F_\xi(n\epsilon).$$

Definition

If the discrete random variables ξ_n have their expected values and converge uniformly, then it can be proved that the limit $\lim_{n\to\infty} \mathbf{E}[\xi_n]$ exists.

Definition 38. *If the discrete random variables ξ_n have the expected values and converge uniformly to the random variable ξ, then the limit value of the expected values is called the expected value of the random variable ξ :*

$$\mathbf{E}[\xi] = \lim_{n\to\infty} \mathbf{E}[\xi_n.]$$

Expectation via the integral

If the random variable ξ has the expected value, then

$$\begin{aligned}\mathbf{E}[\xi] &= \lim_{\epsilon\to 0+} \mathbf{E}[\xi^\epsilon] = \lim_{\epsilon\to 0+} \sum_x xP(\xi^\epsilon = x)\\ &= \lim_{n\to\infty} \sum_{m=-n}^{n} (m\epsilon)(F_\xi(m\epsilon+\epsilon) - F_\xi(m\epsilon)).\end{aligned}$$

The notation:

$$\mathbf{E}[\xi] = \int_{-\infty}^{\infty} x\mathrm{d}F_\xi(x).$$

The properties are the same!

Theorem 49. Let ξ_1, ξ_2 be arbitrary random variables having the expected values. Then

1. with arbitrary numbers c_1, c_2 we have $\mathbf{E}[c_1\xi_1 + c_2\xi_2] = c_1\mathbf{E}[\xi_1] + c_2\mathbf{E}[\xi_2]$;
2. if $\xi_1 \le \xi_2$, then $\mathbf{E}[\xi_1] \le \mathbf{E}[\xi_2]$;
3. if ξ_1, ξ_2 are independent random variables, then $\mathbf{E}[\xi_1 \cdot \xi_2] = \mathbf{E}[\xi_1] \cdot \mathbf{E}[\xi_2]$.

2.9. The variance

Three random variables

$$
\begin{aligned}
&P(X_0 = 0) = 1;\\
&P(X_1 = x) = \frac{1}{2}, \quad x = \pm 1;\\
&P(X_2 = x) = \frac{1}{4}, \quad x = \pm\frac{1}{2},\ \pm 1;\\
&P(X_3 = x) = \frac{1}{6}, \quad x = \pm\frac{1}{4},\ \pm\frac{1}{2}\ , \pm 1.
\end{aligned}
$$

Which random variable has the largest dispersion of its values?

Definition

Definition 39. *Let X be a random variable having the expected value. The number*

$$\mathbf{D}[X] = \mathbf{E}[(X - \mathbf{E}[X])^2]$$

(if it exists) is called the variance of the random variable X. The number $\sigma(X) = \sqrt{\mathbf{D}[X]}$ is called standard deviation (from the expected value) of X.

The properties of variance

Theorem The following statements are true:

1. if for the random variable X the variance exists then $\mathbf{D}[X] \geq 0$; $\mathbf{D}[X] = 0$ if and ony if $P(X = \mathbf{E}[X]) = 1$;
2. $\mathbf{D}[X] = \mathbf{E}[X^2] - \mathbf{E}[X]^2$;
3. with any number c we have $\mathbf{D}[cX] = c^2\mathbf{D}[X]$;
4. if X, Y are independent random variables having the variances, then

$$\mathbf{D}[X + Y] = \mathbf{D}[X] + \mathbf{D}[Y].$$

.

The additivity property of the variance

Additivity property
Let $X_1, X_2, \ldots, X_n$ be the independent random variables having the variances. Then

$$\mathbf{D}[X_1 + X_2 + \cdots + X_n] = \mathbf{D}[X_1] + \mathbf{D}[X_2] + \cdots + \mathbf{D}[X_n].$$

Binomial random variable

If $X \sim \mathcal{B}(n, p)$, then

$$P(X = m) = C_n^m p^m (1-p)^{n-m},\ m = 0, 1, \ldots, n,$$
$$\mathbf{E}[X] = np, \quad \mathbf{D}[X] = np(1-p).$$

Poisson random variable

If $X \sim \mathcal{P}(\lambda)$, then

$$P(X = m) = \frac{\lambda^m}{m!}\,\mathrm{e}^{-\lambda}, \quad m = 0, 1, 2, \dots,$$
$$\mathbf{E}[X] = \mathbf{D}[X] = \lambda.$$

250 / 369

Geometrical random variable

If $X \sim \mathcal{G}(p)$, then $P(X = m) = q^{m-1}p$, $m = 1, \dots,\ q = 1 - p$,

$$\mathbf{E}[X] = \frac{1}{p},\ \mathbf{D}[X] = \frac{q}{p^2}.$$

251 / 369

Geometrical random variables

If we succeed in the first trial (this happens with probability p), then $X^2 = 1$. If we get failure in the first trial, then we start from the beginning, and $X^2 = (1+Y)^2$, here $Y \sim \mathcal{G}(p)$ is a geometrical random variable again. Hence, we may suppose that

$$\mathbf{E}[X^2] = p \cdot 1^2 + q \cdot \mathbf{E}[(1+Y)^2].$$

Let us denote $a = \mathbf{E}[X^2] = \mathbf{E}[Y^2]$, $\mathbf{E}[Y] = 1/p$:

$$a = p + q\mathbf{E}[1 + 2Y + Y^2] = p + q + 2q\mathbf{E}[Y] + q\mathbf{E}[Y^2] =$$
$$1 + \frac{2q}{p} + qa,$$
$$(1-q)a = 1 + \frac{2q}{p}, \quad a = \mathbf{E}[X^2] = \frac{1}{p} + \frac{2q}{p^2},$$
$$\mathbf{D}[X] = \mathbf{E}[X^2] - \mathbf{E}[X]^2 = \frac{1}{p} + \frac{2q}{p^2} - \frac{1}{p^2} = \frac{q}{p^2}.$$

252 / 369

Pascal's random variable

If $X \sim \mathcal{B}^-(n,p)$, then

$$P(X = m) = C_{n+m-1}^{m} p^n q^m,$$
$$\mathbf{E}[X] = n\frac{q}{p}, \quad \mathbf{D}[X] = \frac{nq}{p^2}.$$

253 / 369

The uniformly distributed random variable

If $X \sim \mathcal{T}([a,b])$, then

$$p_X(x) = \begin{cases} \frac{1}{b-a}, & \text{if } x \in [a;b] \\ 0, & \text{if } x \notin [a,b], \end{cases}$$

$$\mathbf{E}[X] = \frac{a+b}{2}, \quad \mathbf{D}[X] = \frac{(b-a)^2}{12}.$$

Exponential random variable

If $X \sim \mathcal{E}(\lambda)$, then

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

$$\mathbf{E}[X] = \frac{1}{\lambda}, \quad \mathbf{D}[X] = \frac{1}{\lambda^2}.$$

Gamma random variable

If $X \sim \Gamma(k, \lambda)$, then

$$p_X(t) = \begin{cases} 0, & \text{if } t < 0, \\ \dfrac{\lambda^k t^{k-1}}{(k-1)!} \mathrm{e}^{-\lambda t}, & \text{if } t > 0, \end{cases}$$

$$\mathbf{E}[X] = \frac{k}{\lambda}, \quad \mathbf{D}[X] = \frac{k}{\lambda^2}.$$

256 / 369

Normal random variable

If $X \sim \mathcal{N}(\mu, \sigma^2)$, then

$$p_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \, \mathrm{e}^{-(x-\mu)^2/(2\sigma^2)},$$

$$\mathbf{E}[X] = \mu, \quad \mathbf{D}[X] = \sigma^2.$$

257 / 369

Quick exercises

Example 58.

There are two symmetrical dices. The sides of the first one are marked by numbers $1, 1, 3, 4, 5, 6,$ and of the second $1, 2, 3, 4, 6, 6.$ The values of the random variables X_1, X_2 are the numbers on the sides of rolled dices. Which of two random variables has the larger dispersion of the values, i.e. which has the bigger variance?

▲

Example 59.

The value of the random variable X is equal to the point on the side of rolled conventional dice, the value of random variable Y equals to the number chosen from the interval $[0; a]$ at random. What should be the value of constant a, such that the variances of both random variables be equal?

▲

2.10. The law of large numbers

Why we repeat the measurements?

Trying to get more exact value of the quantity, we measure it repeatedly and having the values $x_1, x_2, \ldots, x_n$ take the arithmetical mean value

$$y_n = \frac{x_1 + x_2 + \ldots + x_n}{n}.$$

We suppose that $x_n \approx a.$

What is this opinion based on?

The Chebyshev inequality

Theorem 50. Let X be a random variable having the expected value and variance. Then for any $\epsilon > 0$ we have

$$P(|X - \mathbf{E}[X]| \geqslant \epsilon) \leqslant \frac{\mathbf{D}[X]}{\epsilon^2}.$$

The law of large numbers

Theorem 51. Let $X_1, X_2, X_3, \ldots$ be the independent random variables having the same expected value $\mathbf{E}[X_j] = a$ and the same variance. Then for each $\epsilon > 0$

$$P\Big(\Big|\frac{X_1 + X_2 + \ldots + X_n}{n} - a\Big| > \epsilon\Big) \to 0, \quad n \to \infty.$$

The Monte-Carlo method

Let K be a unite square and $S \subset K$ some region with the complicated boundary. We have to compute the area of this region.
Our experiment: we choose the points $A_1, A_2, \ldots, A_n$ in the square randomly and define the random variables

$$X_i = \begin{cases} 1, & \text{if } x \in A_i \in S; \\ 0, & \text{if } x \notin S. \end{cases}$$

Let $x_1, x_2, \ldots, x_n$ be the values of X_i obtained from the experiment. Then

$$area(S) \approx (x_1 + x_2 + \ldots + x_n)/n.$$

Generalization of the law of large numbers

Theorem 52. Let $X_1, X_2, X_3, \ldots$ be independent random variables having the variances and satisfying the condition

$$\frac{1}{n^2} \sum_{m=1}^{n} \mathbf{D}[X_m] \to 0, \quad n \to \infty.$$

Let

$$S_n = X_1 + \cdots + X_n, \quad E_n = (\mathbf{E}[X_1] + \cdots + \mathbf{E}[X_n])/n.$$

Then for any $\epsilon > 0$

$$P\big(\big|S_n/n - E_n\big| > \epsilon\big) \to 0, \quad n \to \infty.$$

2.11. The moments of random variables

Definition

Definition 40. *Let ξ be a random variable, $k > 0$. If the expected values $\mathbf{E}[\xi^k], \mathbf{E}[|\xi|^k]$ exist, then these values are called the moments of kth order; the second one - the absolute moment of kth order of the random variable ξ.*

Note. *If $\mathbf{E}[|\xi|^k]$ exists, then $\mathbf{E}[\xi^k]$ exists too. If the moment of random variable ξ of kth order exists, then for every $0 < r < k$ the expected value $\mathbf{E}[|\xi|^r]$ exists too.*
If there exists the moment of ξ of the kth order, then for any a the expected value $\mathbf{E}[|\xi - a|^k]$ exists as well.

The central moments

Definition 41. *Let for the random variable ξ the moments of the kth order exist $(k > 1)$. Then the expexted values*

$$\mathbf{E}[(\xi - \mathbf{E}[\xi])^k], \quad \mathbf{E}[|\xi - \mathbf{E}[\xi]|^k]$$

are called the central moments of kth order; the second one is called the absolute central moment of order k.

Note. *The variance is the central moment of the second order.*

Mixed moments

Theorem 53. Let the random variables ξ_1, ξ_2 have the moments of second order. Then the expected value $\mathbf{E}[\xi_1\xi_2]$ exists and

$$\mathbf{E}[|\xi_1 \cdot \xi_2|] \leq \sqrt{\mathbf{E}[\xi_1^2] \cdot \mathbf{E}[\xi_2^2]}.$$

Note. If the random variables ξ_1, ξ_2 have their variances, then the expected value $\mathbf{E}[(\xi_1 - \mathbf{E}[\xi_1]) \cdot (\xi_1 - \mathbf{E}[\xi_1])]$ exists and

$$\mathbf{E}[|(\xi_1 - \mathbf{E}[\xi_1])(\xi_1 - \mathbf{E}[\xi_1])|] \leq \sqrt{\mathbf{D}[\xi_1]\mathbf{D}[\xi_2]}.$$

The variance of the sum

Let ξ_1, ξ_2 be independent random variables having their variances. Then

$$\mathbf{D}[\xi_1 + \xi_2] = \mathbf{D}[\xi_1] + \mathbf{D}[\xi_2].$$

Suppose that ξ_1, ξ_2 can be dependent. Then for the sum we have

$$\mathbf{D}[\xi_1 + \xi_2] = \mathbf{E}[(\xi_1 + \xi_2 - \mathbf{E}[\xi_1] - \mathbf{E}[\xi_2])^2] =$$
$$\mathbf{E}[(\xi_1 - \mathbf{E}[\xi_1])^2 + 2(\xi_1 - \mathbf{E}[\xi_1])(\xi_2 - \mathbf{E}[\xi_2]) + (\xi_2 - \mathbf{E}[\xi_2])^2] =$$
$$\mathbf{D}[\xi_1] + \mathbf{D}[\xi_2] + 2\mathbf{E}[(\xi_1 - \mathbf{E}[\xi_1])(\xi_2 - \mathbf{E}[\xi_2])].$$

The covariance between the random variables

Definition 42. *Let ξ_1, ξ_2 be the random variables having the variances. The number*

$$cov(\xi_1, \xi_2) = \mathbf{E}[(\xi_1 - \mathbf{E}[\xi_1]) \cdot (\xi_2 - \mathbf{E}[\xi_2])]$$

is called the covariance between the random variables ξ_1, ξ_2.

Theorem 54. *For the random variable ξ_1, ξ_2 having the variances we have*

$$cov(\xi_1, \xi_2) = \mathbf{E}[\xi_1 \cdot \xi_2] - \mathbf{E}[\xi_1] \cdot \mathbf{E}[\xi_2].$$

Example

Three white balls in the urn are marked with numbers $1, 0, 0$ and two black ones with number $1, 1$. Two balls are drawn randomly without replacement. The value of X is equal to the number of white balls, the value of Y is equal to the sum of numbers on the balls drawn from the urn. Compute the covariance between X, Y.
The table of probabilities $P(X = x, Y = y)$ is very helpful:

	$X = 0$	$X = 1$	$X = 2$	
$Y = 0$	0	0	$0,1$	$0,1$
$Y = 1$	0	$0,4$	$0,2$	$0,6$
$Y = 2$	$0,1$	$0,2$	0	$0,3$
	$0,1$	$0,6$	$0,3$	

Correlated random variables

Definition 43. *Let ξ_1, ξ_2 be the random variables. If $cov(\xi_1, \xi_2) > 0,$ then the random variables are called positively correlated, if $cov(\xi_1, \xi_2) < 0,$ we say that the random variables are negatively correlated. If $cov(\xi_1, \xi_2) = 0,$ the random variables are uncorrelated.*

Note. *Independent random variables having the variances are uncorrelated.*

If the random variables are uncorrelated, they are not necessarily independent.

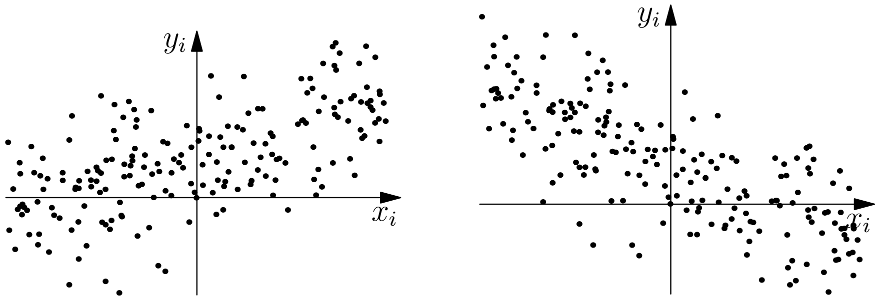
272 / 369

The correlated random variables

Let an experiment was repeated n times and the pairs of values of random variables X, Y were obtained

$$(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n).$$

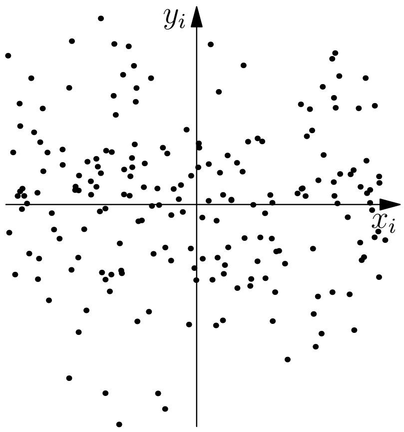
The data can be represented by the points on the plain.



273 / 369

Uncorrelated random variables

If $cov(X, Y) = 0$, the points representing the data resemble a cloud, without any grouping tendency along some line.



The correlation coefficient

Definition 44. *Let ξ_1, ξ_2 be the random variables and $\mathbf{D}[\xi_1] > 0, \mathbf{D}[\xi_2] > 0$. The number*

$$\rho(\xi_1, \xi_2) = \frac{cov(\xi_1, \xi_2)}{\sqrt{\mathbf{D}[\xi_1]\mathbf{D}[\xi_2]}}$$

is called the correlation coefficient of the random variables ξ_1, ξ_2. If for at least one random variable $\mathbf{D}[\xi_i] = 0$, we set $\rho(\xi_1, \xi_2) = 0$.

The correlation coefficient

Theorem 55. Let ξ_1, ξ_2 be the random variables and $\mathbf{D}[\xi_1] > 0, \mathbf{D}[\xi_2] > 0$. Let a_1, a_2, b_1, b_2 be arbitrary numbers, $a_1, a_2 \neq 0$. Then

$$\rho(a_1\xi_1 + b_1, a_2\xi_2 + b_2) = \begin{cases} \rho(\xi_1, \xi_2), & \text{if } a_1a_2 > 0, \\ -\rho(\xi_1, \xi_2), & \text{if } a_1a_2 < 0. \end{cases}$$

The properties of the correlation coefficient

Theorem 56. Let ξ_1, ξ_2 be the random variables with non-zero variances. The following statements are true

1. $-1 \leq \rho(\xi_1, \xi_2) \leq 1$;
2. if $\xi_2 = a\xi_1 + b$, here a, b are some numbers, then $\rho(\xi_1, \xi_2) = 1$, as $a > 0$ and $\rho(\xi_1, \xi_2) = -1$, as $a < 0$;
3. if $\rho(\xi_1, \xi_2) = \pm 1$, then there exist some numbers $a \neq 0, b$, such that

$$P(\xi_2 = a\xi_1 + b) = 1.$$

Convergence almost surely

Definition 45. *Let ξ and $\xi_1, \xi_2, \ldots$ be the random variables, defined on the same probabilistic spece. We say that ξ_n converges almost surely (or with probability 1) to ξ, if*

$$P(\omega : \xi_n(\omega) \xrightarrow[n\to\infty]{} \xi(\omega)) = 1.$$

The notation: $\xi_n \xrightarrow{1} \xi$.

Example

$\Omega = [0; 1]$, P is geometrical length, $X_n(\omega) = 1/n$, $X(\omega) = 0$, then

$$X_n \xrightarrow{1} X$$

The limit random variable is not uniquely defined. Let $X^*(\omega) = 0$, if ω is an irrational number and $X(\omega) = \omega$, if ω is a rational number. Then

$$X_n \xrightarrow{1} X^*.$$

Convergence in probability

Definition 46. *Let* $\xi, \xi_n : \Omega \to \mathbb{R}$ *be the random variables,* $n = 1, 2, \ldots$ *We say that the sequence of random variables* ξ_n *converges in probability to the random variable* ξ *, if for any* $\epsilon > 0$

$$P(\omega : |\xi_n(\omega) - \xi(\omega)| > \epsilon) \to 0, \quad n \to \infty.$$

The notation: $\xi_n \xrightarrow[n\to\infty]{P} \xi$.

Example

A sequence of random variables X_n can be constructed such that

$$X_n \xrightarrow[n\to\infty]{P} X$$

but all sequences of numbers $X_n(\omega)$ diverge!

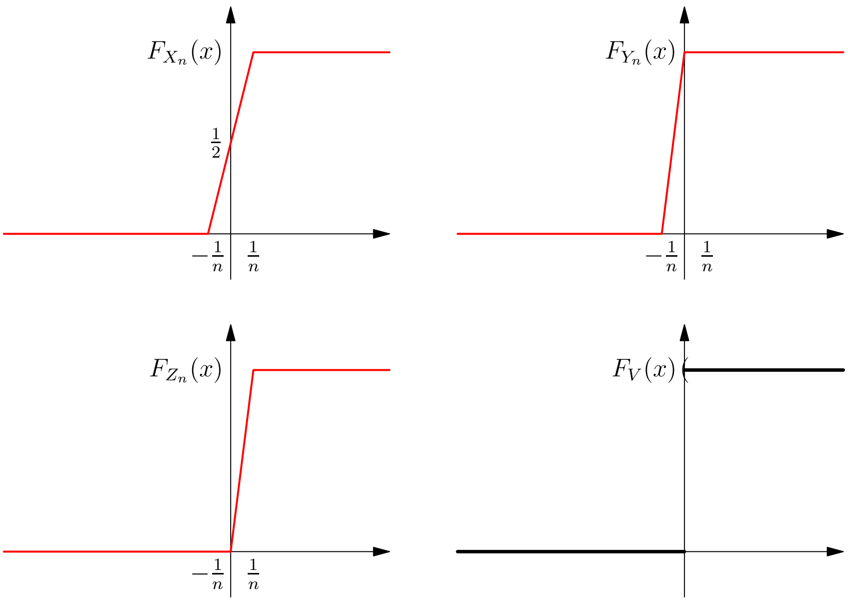
The weak convergence

Definition 47. *Let ξ_n, ξ be the random variables and F_n, F the distribution functions of the random variables, respectively. We say that the sequence ξ_n converges weakly to the random variable ξ, if for any continuity point x of $F(x)$ we have*

$$F_n(x) \to F(x), \quad n \to \infty.$$

The notation for the weak convergence: $\xi_n \Rightarrow \xi$. We can also say, that the sequence of distribution functions F_n converges to the limit distribution function F.

The weak convergence



The sequences of distribution functions $F_{X_n}, F_{V_n}, F_{Z_n}$ converge weakly to the distribution function F_V.

Converges but not always to the distribution function

Theorem 57. Let F_n be a sequence of distribution functions of some random variables. Then there exists a subsequence F_{n_m} and a function $G(x)$ which is non-decreasing and continuous from the left, such that we have $F_{n_m} \to G(x)$, as $m \to \infty$ in all continuity points of $G(x)$.

Question

How to investigate the weak convergence of given sequence of distribution functions?

The complex random variables

Definition 48. *Let $\langle \Omega, \mathcal{A}, P \rangle$ be a probabilistic spece, $\mathbb{C}$ the set of complex numbers. A function $\xi : \Omega \to \mathbb{C}$ is called complex random variable, if*

$$\xi = \xi_1 + i\xi_2,$$

where $\xi_1 = \operatorname{Re} \xi, \xi_2 = \operatorname{Im} \xi$ are real valued random variables.

Independent random variables

Definition 49. *The complex random variables*

$$\xi = \xi_1 + i\xi_2, \qquad \eta = \eta_1 + i\eta_2$$

are called independent, if the random variables in all pairs $\{\xi_i, \eta_j\}$ are mutually independent.

Expectation

Definition 50. *Let the real valued random variables ξ_1, ξ_2 have their expected values. Then the expected value of the complex random variable*

$$\xi = \xi_1 + i\xi_2$$

is the complex number

$$\mathbf{E}[\xi] = \mathbf{E}[\xi_1] + i\mathbf{E}[\xi_2].$$

Properties of expectation

Theorem 58.

1. Let ξ_1, ξ_2 be the complex random variables having the expected values, and a, b some complex numbers. Then the random variable $\xi = a\xi_1 + b\xi_2$ has its expected value and $\mathbf{E}[\xi] = a\mathbf{E}[\xi_1] + b\mathbf{E}[\xi_2]$.
2. Let the complex random variable ξ has the expected value, then

$$|\mathbf{E}[\xi]| \leq \mathbf{E}[|\xi|].$$

3. Let ξ_1, ξ_2 be independent complex random variables having the expected values. Then the random variable $\xi = \xi_1 \cdot \xi_2$ has the expected value and $\mathbf{E}[\xi] = \mathbf{E}[\xi_1] \cdot \mathbf{E}[\xi_2]$.

Characteristic function

Definition 51. *Let ξ be a real-valued random variable. The function defined on the real line by*

$$\phi_\xi(t) = \mathbf{E}[\mathrm{e}^{it\xi}] = \mathbf{E}[\cos(t\xi)] + i\mathbf{E}[\sin(t\xi)].$$

is called the characteristic function of the random variable ξ.
Let $\xi = \langle \xi_1, \dots, \xi_n \rangle$ be a random vector. The function

$$\phi_\xi(t_1, \dots, t_n) = \mathbf{E}[\mathrm{e}^{it_1\xi_1 + \dots + it_n\xi_n}].$$

is called the characteristic function of random vector ξ.

The properties of characteristic function

Theorem 59. The following statements are true:

1. The characteristic function is continuous in all points of definition.
2. Let ξ be a random variable and a, b some fixed numbers, $\eta = a\xi + b$. Then
$$\phi_\eta(t) = \mathrm{e}^{itb}\phi_\xi(at).$$
3. Let ξ_1, ξ_2 be two independent random variables, $\xi = \xi_1 + \xi_2$. Then
$$\phi_\xi(t) = \phi_{\xi_1}(t)\phi_{\xi_2}(t).$$

Important examples

If $\xi \sim \mathcal{B}(n, p)$, then $\phi_\xi(t) = \left(pe^{it} + q\right)^n$.

If $\xi \sim \mathcal{P}(\lambda)$, then $\phi_\xi(t) = \exp\{\lambda(e^{it} - 1)\}$.

If $\xi \sim \mathcal{N}(0, 1)$, then $\phi_\xi(t) = e^{-t^2/2}$.

Characteristic function and the moments

Theorem 60. Let for the random variable ξ the moment of the mth order exists. Then at every point t there exists the mth derivative of characteristic function $\phi_\xi(t)$ and

$$\phi_\xi(t)^{(m)} = \mathbf{E}[(i\xi)^m e^{it\xi}].$$

For the characteristic function the following asymptotic expansion holds:

$$\phi_\xi(t) = \sum_{l=0}^{m} \mathbf{E}[\xi^l] \frac{(it)^l}{l!} + r_m(t) \frac{(it)^m}{m!};$$

here $r_m(t) \to 0, t \to 0$.

Uniqueness theorem

Theorem 61. If the distribution functions of two random variables are different, then the characteristic functions are different too.

Applications: The sum of independent Poisson random variables is a Poisson random variable.
The sum of independent normal random variables is a normal random variable.

Continuity theorem

Theorem 62. The sequence of distribution functions F_n converge weakly to some limit distribution function F, if and only if the sequence of corresponding characteristic functions $\phi_n(t)$ converge at every point to some continuous at the point $t = 0$ function $\phi(t)$.
The function $\phi(t)$ is the characteristic function corresponding to the distribution function F.

The Poisson theorem

Theorem 63. Let ξ_n be the random variables, $\xi_n \sim \mathcal{B}(n, p_n)$ and $np_n \to \lambda$, as $n \to \infty$, here $\lambda > 0$. Then the distribution functions of the random variables ξ_n converge weakly to the distribution function of the Poisson random variable $\xi \sim \mathcal{P}(\lambda)$.

The law of large numbers

Theorem 64. Let $\xi_1, \xi_2, \ldots$ be independent random variables having the same distribution function and the expected value a. Then for an arbitrary $\epsilon > 0$

$$P\Big(\Big|\frac{\xi_1 + \xi_2 + \ldots + \xi_n}{n} - a\Big| > \epsilon\Big) \to 0, \quad n \to \infty.$$

The central limit theorem

Theorem 65. Let ξ_m be independent random variables having the same distribution function, the expected value $\mathbf{E}[\xi_m] = a$ and the variance $\mathbf{D}[\xi_m] = \sigma^2$. Then the distribution functions

$$F_n(x) = P\Big(\sum_{m=1}^{n} \frac{\xi_m - a}{\sigma\sqrt{n}} < x\Big)$$

converge weakly to the distribution function of the standard normal variable $\Phi(x)$, i. e. for all x

$$F_n(x) \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \mathrm{e}^{-u^2/2} du, \quad n \to \infty.$$

301 / 369

III. The mathematical statistics 302 / 369

3.1. Descriptive statistics 303 / 369

Example

Example 60.
You have to share the dewy matches with your friend. A dewy match strikes with the probability $p = 0,6$. How many dewy matches you should give to your friend for he could put on fire with the probability no less then $0{,}9$?

▲

Important question: from where we know the value of probability p?

304 / 369

Gathering data

Some subset of objects are chosen from the general set for investigation. On the basis of data obtained the conclusions are made about the general set of objects.

How this practise should by described mathematically?

305 / 369

The population

Suppose that the property we are interested in are expressed using the random variable X.

The values of X are numbers or symbolic strings. The experiment consists of choosing the object and obtaining the value of X. Suppose that the objects are chosen independently.

The random variable associated with the first choice we denote by X_1, with the second choice – X_2 and so forth.

306 / 369

The population

The mathematical concept, corresponding to the model of gathering data by random choice is
the sequence of independent identically distributed random variables

$$X_1, \ldots, X_n.$$

Random sample

Definition 52. *The sequence of independent identically distributed random variables*

$$\langle X_1, X_2, \ldots, X_n \rangle$$

is called the random sample.

Random sample and its realization

From the actual measurements we get the values of the elements of random sample.

Definition 53. *Let $\langle X_1, X_2, \ldots, X_n\rangle$ be the random sample. A sequence of values of the elements $\langle x_1, x_2, \ldots, x_n\rangle$ is called realization of the random sample, or simply the sample.*

309 / 369

The first problem

The aim of descriptive statistics – to arrange, systematize and visualize the data collected from the experiments.

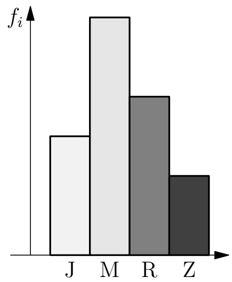
310 / 369

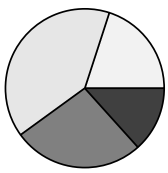
Frequencies

Data	x_1	x_2	x_3	$\ldots$	x_m
Frequencies n_i	n_1	n_2	n_3	$\ldots$	n_m
Relative frequencies f_i	n_1/n	n_2/n	n_3/n	$\ldots$	n_m/n
Cummulative frequencies $\sum_{j<i} f_j$	0	f_1	$f_1 + f_2$	$\ldots$	$f_1 + \ldots + f_{m-1}$

Tables and diagrams

	J	M	R	Ž
n_i	3	6	4	2
f_i	3/15	6/15	4/15	2/15
$\sum_{j<i} f_j$	0	3/15	9/15	13/15



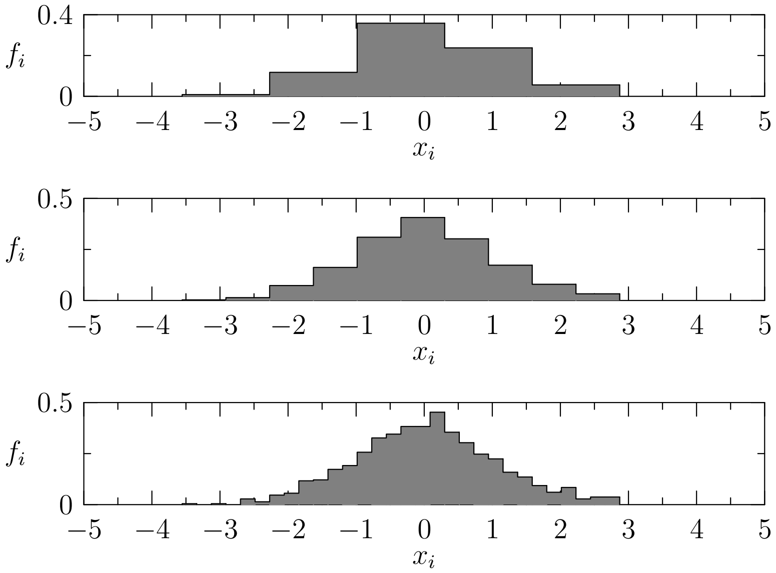


The grouped data and histogram

Data intervals	I_1	I_2	I_3	$\ldots$	I_N
Numbers of data in the intervalsI_j	n_1	n_1	n_2	$\ldots$	n_N
Relative frequencies p_j	n_1/nd	n_2/nd	n_3/nd	$\ldots$	n_N/nd

313 / 369

Histograms



314 / 369

Histograms

Three histograms of the same data set with $n = 1000$ data: $N = 5, 10, 30$.
Recommended parameter for the number of intervals

$$N \approx 1 + 3,3 \lg n.$$

Ordered list

Let $x_1, x_2, \ldots, x_n$ be the numerical dataset. Let us put them in order. We call this new arranged dataset

$$x_{(1)} \leqslant x_{(2)} \leqslant \ldots \leqslant x_{(n)}$$

the ordered list of the sample. The elements $x_{(k)}$ are called the order statistics.
For example, the orderd list of the sample

$$2; 1{,}5; 3, 1{,}5; 2, 3; 1{,}7; 2$$

is $1{,}5;\ 1{,}5;\ 1{,}7;\ 2;\ 2;\ 2;\ 3;\ 3,\ x_{(1)} = x_{(2)} = 1{,}5; x_{(8)} = 3.$

The empirical distribution function

Let the observed random variable X takes the numerical values, and $x_1, x_2, \dots, x_n$ is a sample of this random variable. We construct the empirical distribution function based on this dataset.
Let $n(x)$ be the number of data in the dataset less than $x.$

Then the empirical distribution function is defined by

$$F_X^*(x) = \frac{n(x)}{n}.$$

The empirical distribution function

The empirical distribution function based on the sample $2; 1{,}5; 3, 1{,}5; 2{,}3; 1{,}7; 2$ is

$$F_X^*(x) = \begin{cases} 0, & \text{if } x \le 1,5; \\ \frac{2}{8}, & \text{if } 1{,}5 < x \le 1{,}7; \\ \frac{3}{8}, & \text{if } 1{,}7 < x \le 2; \\ \frac{6}{8}, & \text{if } 2 < x \le 3; \\ 1, & \text{if } x > 3; \end{cases}$$

The quantiles of the sample

Let us denote by $\underline{n}(x)$ the number of data in the sample $x_1, x_2, \ldots, x_n$ not exceeding x (i. e. satisfying $x_i \leqslant x$), and by $\overline{n}(x)$ the number of data no less than x (i. e. satisfying $x_i \geqslant x$). Then $\underline{n}(x) + \overline{n}(x) \geqslant n$.

We want to define the qth empirical v_q in a way that the following property should be satisfied:

$$q \leqslant \frac{\underline{n}(v_q)}{n}, \quad \frac{\overline{n}(v_q)}{n} \geqslant 1 - q.$$

319 / 369

The quantiles of the sample

Definition 54. *The qth quantile of the sample $x_1, x_2, \ldots, x_n$ is a number v_q, defined by:*

$$v_q = \begin{cases} x_{([qn]+1)}, & \text{if } qn \text{ is not an integer}, \\ (x_{(qn)} + x_{(qn+1)})/2, & \text{if } qn \text{ is an integer}, \end{cases}$$

here $0 < q < 1$, and $[qn]$ means the integer part, $x_{(i)}$ is the ith order statistics of the sample.

320 / 369

The quartiles and median

The mostly used qth quantiles are for $q = \frac{1}{4}, \frac{2}{4}, \frac{3}{4}$. They are called quartiles and denoted by Q_1, Q_2, Q_3.

The quartile Q_2 is called median.

The expectation and variance of the sample

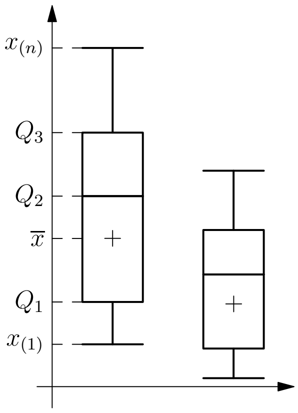
Definition 55. *Let $\langle x_1, x_2, \dots, x_n \rangle$ be a sample corresponding to the random variable X. The expectation and variance of the sample are defined by*

$$\overline{x} = \frac{x_1 + x_2 + \cdots + x_n}{n},$$

and

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x})^2.$$

Visualization of the numerical characteristics



Several numerical characteristics of the sample in one diagram. Useful for comparing different samples.

Quick exercises

Problem 9. *The expenses of the buyers at the check are:*

$$27,\ 20,\ 12,\ 20,\ 15,\ 20,\ 45,\ 10,\ ,15,\ 10,\ 15,\ 30,\ 25,\ 20,\ 12.$$

Find the order statistics of the sample $x_{(5)}, x_{(11)}$. *Find the quartiles.*

Quick exercises

Problem 10. *The score of mid-term exam of the students are presented in the table of frequencies:*

$x_i =$	1	2	3	4	5	6	7	8	9	10
$n_i =$	0	1	2	4	3	5	3	5	5	4

Find the median and the quartiles of the sample. Find the expectation and variance.

Quick exercises

Problem 11. *The grades of nine students are*

$$7,\ 6,\ 7,\ 10,\ 6,\ 5,\ 5,\ ,9,\ 8.$$

Find the median and expectation of the sample. The grade of the tenth student is unknown, but it is known that it will be at least 1. *What is the largest possible decrease of the expectation? What is the largest possible increase? Can the value of expectation remain the same as the tenth grade will be available? When is it possible? If the expectation will remain the same, how the variance will change? How the median depends on the tenth grade: when it will change, when not?*

Quick exercises

Problem 12. *The durations of N phone talks of some person are:*

$$3,\ 5,\ 4,\ 3,\ 4,\ 2,\ 4,\ 7.$$

How many jumps the empirical distribution function of this sample will have? What is the value of the largest jump? Draw the graph of the empirical distribution function.

327 / 369

3.2. Estimators

328 / 369

The estimators

Suppose that the distribution of values of the random variable observed depends on some parameter θ.
How we could find (approximate) value of θ?

Having a sample we compute the value of some function

$$\theta^* = h(x_1, x_2, \ldots, x_n),$$

which (as we believe on the basis of some arguments) give the value of θ.

329 / 369

Statistics of the random sample

Definition 56. *Let $\langle X_1, X_2, \ldots, X_n \rangle$ be a random sample. The random variable*

$$T = h(X_1, X_2, \ldots, X_n)$$

is called statistics of the sample, here h is some function.

330 / 369

Unbiased estimator

Definition 57. *We say that statistics $\theta^* = h(X_1, X_2, \ldots, X_n)$ is an unbiased estimator of the parameter θ, if*

$$\mathbf{E}[\theta^*] = \mathbf{E}[h(X_1, X_2, \ldots, X_n)] = \theta.$$

331 / 369

Expectation and variance

Theorem 66. Let the expectation of the random variable X is a parameter a and the variance σ^2. Let $\langle X_1, X_2, \dots, X_n \rangle$ be the random sample of X. Then the statistics

$$\overline{X} = \frac{X_1 + X_2 + \ldots + X_n}{n}, \quad S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2$$

are unbiased estimators of the parameters a and σ^2.

Empirical moments

Definition 58. *Let the kth moment of the random variable X exists, $\alpha_k = \mathbf{E}[X^k]$. The estimator*

$$a_k = \frac{X_1^k + X_2^k + \cdots + X_n^k}{n}$$

is called the kth empirical moment of the random variable; here $\langle X_1, X_2, \dots, X_n \rangle$ is the random sample of X.

The method of moments

Suppose the distribution of the random variable depends on the parameters $\theta_1, \theta_2, \ldots, \theta_r$ and the first r moments of the random variable exist.
Solve the system of equations

$$
\begin{aligned}
&\alpha_1(\theta_1, \theta_2, \ldots, \theta_r) = a_1, \\
&\alpha_2(\theta_1, \theta_2, \ldots, \theta_r) = a_2, \\
&\ldots \\
&\alpha_r(\theta_1, \theta_2, \ldots, \theta_r) = a_r.
\end{aligned}
$$

and find the estimators of $\theta_1, \theta_2, \ldots, \theta_r$.

Example

Example 61.
How many shots and what the accuracy?
The number of targets is n, to each of them k shots were fired. We know the number of hits of each target $x_1, x_2, \ldots, x_n$. Find the estimator of the number of shots n and the probability to hit the target by one shot.

▲

Example

Example 62. *Being late from the school*
The time to come back from the school is 15 minutes, but usually the pupil is late. The time of being late is the random variable X, which is made up from two independent components:

$$X = X_1 + X_2,$$

here $X_1 \sim \mathcal{T}([0, a])$ is the additional time of delay on the road, and $X_2 \sim \mathcal{P}(\lambda)$ the time of talk with the schoolfriend before saying farewell. The times of being late are known for n days: $\langle x_1, x_2, \ldots, x_n \rangle$. We need to compute the estimated values of a and λ.

▲

Numerical data:

$$8,07;\ 16,53;\ 12,63;\ 11,57;\ 12.16;\ 4,49;\ 7,39;\ 13,73;\ 13,78;\ 16,83.$$

3.3. The confidence intervals

The confidence intervals

Definition 59. *Let $\langle X_1, X_2, \ldots, X_n \rangle$ be the random sample of the variable X, and θ some parameter related to the distribution of X. Let*

$$\underline{\theta}(X_1, X_2, \ldots, X_n) \leqslant \overline{\theta}(X_1, X_2, \ldots, X_n)$$

be two estimators of the parameter θ and Q is some number $0 < Q < 1$. The interval

$$I = (\underline{\theta}(X_1, X_2, \ldots, X_n), \overline{\theta}(X_1, X_2, \ldots, X_n))$$

is called a Q confidence interval for θ, if

$$P(\theta \in (\underline{\theta}(X_1, X_2, \ldots, X_n), \overline{\theta}(X_1, X_2, \ldots, X_n)) \geqslant Q.$$

Wormy mushrooms

Example 63.
Let p be the probability that a mushroom found is wormy. It is unknown. There were 1000 mushrooms found, 470 of them were wormy. Construct the Q confidence interval for the unknown probability if $Q = 0,9$.

▲

The normal random variables

The random variable X is called standard normal ($X \sim \mathcal{N}(0,1)$), if its density function is

$$p_X(x) = \frac{1}{\sqrt{2\pi}}\,\mathrm{e}^{-x^2/2}.$$

We know that $\mathbf{E}[X] = 0, \mathbf{D}[X] = 1$. If $X \sim \mathcal{N}(0,1)$, then with the arbitrary numbers $\sigma \neq 0, \mu$, the random variable $Y = \sigma X + \mu$ also is a normal random variable

$$Y \sim \mathcal{N}(\mu, \sigma^2),\ \mathbf{E}[Y] = \mu,\ \mathbf{D}[Y] = \sigma^2.$$

Let us choose arbitrary numbers $a \neq 0$ and b and define the new random variable $Z = aY + b$. This random variable will be also normal.
Any linear transform of the normal variable gives the new normal variable again.

The normal random variables

Theorem 67. Let X_1, X_2 be two independent normal random variables. Then the sum $X = X_1 + X_2$ will be the normal random variable, too.

Theorem 68. Let $X_1, X_2, \ldots, X_n$ be the independent normal random variables and $a_1, a_2, \ldots, a_n, b$ arbitrary numbers not all equal to zero. Then the random variable

$$Y = a_1X_1 + a_2X_2 + \cdots + a_nX_n + b$$

will also be normal.

341 / 369

The normal random variables

Theorem 69. Let $X_i \sim \mathcal{N}(\mu, \sigma^2)$ $(i = 1, 2, \ldots, n)$ be indepenent normal random variables. Then the random variable

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}, \quad \overline{X} = X_1 + X_2 + \cdots + X_n,$$

is standard normal, i. e. $Z \sim \mathcal{N}(0, 1)$.

342 / 369

The confidence interval for expectation

The confidence interval for the expectation of $X \sim \mathcal{N}(\mu, \sigma^2)$, if σ^2 is known

The Q confidence interval for the expectation is

$$\left(\overline{X} - z_{(1+Q)/2}\frac{\sigma}{\sqrt{n}}; \overline{X} + z_{(1+Q)/2}\frac{\sigma}{\sqrt{n}}\right),$$

here $0 < Q < 1$, and $z_{(1+Q)/2}$ is the solution of the equation $\Phi(z) = (1+Q)/2$.

Example

Example 64.
The sample of the random variable $X \sim \mathcal{N}(\mu, 1)$ consists of 10 entries:

$$5.26,\ 4.80,\ 4.91,\ 4.98,\ 4.79,\ 4.99,\ 3.81,\ 5.29,\ 6.15,\ 4.21.$$

The expected value of the sample is $\overline{X} = 4.919$. Construct the confidence intervals for the expectations for some values of Q. Compute and compare the lengths of the intervals.

▲

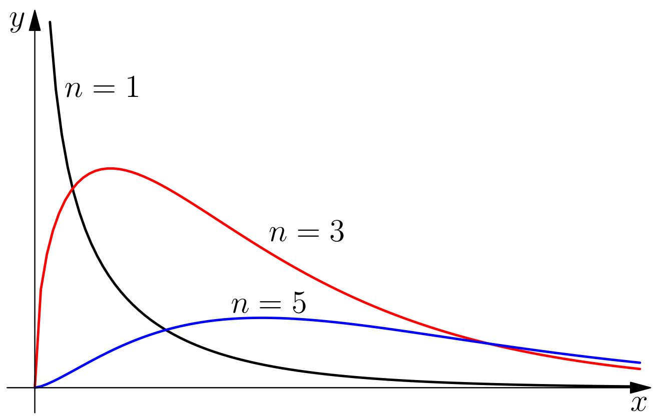
New random variables

Definition 60. *Let $X_0, X_1, X_2, \dots, X_n$ be independent standard normal random variables. Let us define*

$$\chi_n^2 = X_1^2 + X_2^2 + \cdots + X_n^2, \quad T_n = \frac{X_0}{\sqrt{\chi_n^2/n}}.$$

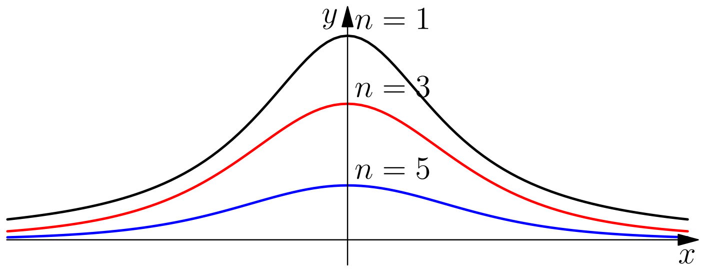
We say that the random variable χ_n^2 has the chi-square distribution with n degrees of freedom and denote $\chi_n^2 \sim \chi^2(n)$. The random variable T_n has t-distribution (or Student distribution) with n degrees of freedom, the notation is $T_n \sim \mathcal{S}t(n)$.

The graphs of density functions



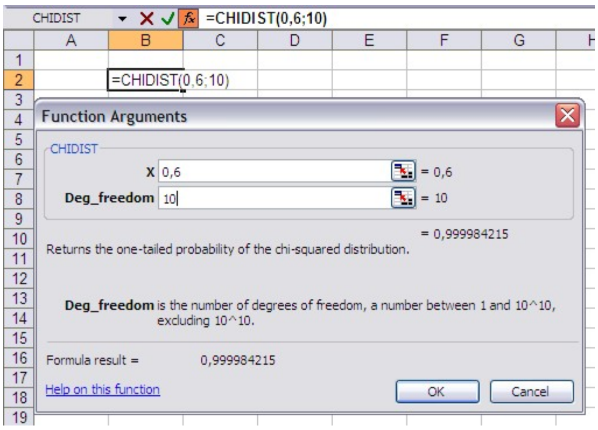
The graphs of density functions of the random variables χ_n^2.
If $n = 1$, the density function grows unboundedly in the zero neighbourhood.

The graphs of density functions



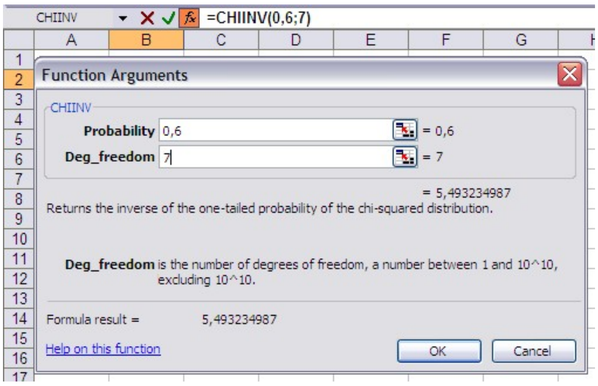
The graphs of density functions of random variables $T_n \sim \mathcal{S}t(n)$.

Computing the values of distribution function



The values of χ-square distribution function using Excel.

Computing the quantiles



The values of χ-square distribution function and quantiles using Excel.

349 / 369

The main theorem

Theorem 70. Let $X \sim \mathcal{N}(\mu, \sigma^2)$, and let $\langle X_1, X_2, \ldots, X_n \rangle$ be the random sample of this random variable. Then

$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}} \sim \mathcal{S}t(n-1).$$

350 / 369

The confidence interval

The confidence interval for expectation of $X \sim \mathcal{N}(\mu, \sigma^2)$ as σ^2 is unknown

The Q confidence interval is

$$\left(\overline{X} - t\frac{S}{\sqrt{n}}; \overline{X} + t\frac{S}{\sqrt{n}}\right),$$

here $0 < Q < 1, t = t_{(1+Q)/2}(n-1)$ is the solution of the equation $F_{T_{n-1}}(t) = (1+Q)/2, T_{n-1} \sim \mathcal{S}t(n-1)$,

$$\overline{X} = \frac{X_1 + \cdots + X_n}{n}, \quad S = \sqrt{S^2} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2}.$$

Example

Example 65. *A sample with ten entries*

The sample of $X \sim \mathcal{N}(\mu, \sigma^2)$ consists of the following numbers

$$4.19,\ 4.20,\ 5.12,\ 6.11,\ 4.37,\ 5.50,\ 4.81,\ 4.44,\ 4.17,\ 5.91.$$

The expectation of the sample is $\overline{X} = 4.88$, the variance and standard deviation equal to $s^2 = 0.544, s = 0.738$ respectively. The Q confidence intervals are given in the table.

$Q =$	$0,6$	$0,7$	$0,8$	$0,9$	$0,95$
$z =$	0.883	1.10	1.38	1.83	2.26
$\underline{\mu} =$	4.67	4.62	4.56	4.45	4.35
$\overline{\mu} =$	5.09	5.14	5.20	5.31	5.41
$length =$	0.411	0.512	0.645	0.853	1.06

▲

The confidence interval for the probability

Example 66. *The wormy mushrooms*
Among $n = 1000$ mushrooms $m = 470$ were wormy. Construct the confidence intervals for the probability that a mushroom is wormy using the Chebyshev inequality and the central limit theorem with $Q = 0,6; 0,7; 0,8.$

▲

The confidence interval for variance

The confidence interval for the variance if the expectation is known
Let $X \sim \mathcal{N}(\mu, \sigma^2)$, and let $\langle X_1, X_2, \ldots, X_n \rangle$ be the random sample with the known expectation μ. The Q confidence interval for the variance σ^2 is

$$\left(\frac{nS_0^2}{v}; \frac{nS_0^2}{u}\right)$$

here u, v are the quantiles of the random variable $\chi_n^2 \sim \chi^2(n)$ of orders $(1-Q)/2, (1+Q)/2$ respectively, and

$$S_0^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2.$$

The confidence interval for the variance

The confidence interval for variance if the expectation is unknown
Let $X \sim \mathcal{N}(\mu, \sigma^2)$, and let $\langle X_1, X_2, \ldots, X_n \rangle$ be the random sample, the expectation μ is unknown. The Q confidence interval for the variance σ^2 is

$$\Big(\frac{(n-1)S^2}{v}; \frac{(n-1)S^2}{u}\Big);$$

here u, v are the quantiles of the random variable $\chi_n^2 \sim \chi^2(n-1)$ of order $(1-Q)/2$ and $(1+Q)/2$ respectively, and

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2, \quad \overline{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}.$$

The confidence interval for the probability of success

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

The expected value is $\mathbf{E}[X] = p$, and the sample is a sequence of zeroes and ones $\langle x_1, x_2, \ldots, x_n \rangle$. The estimator of probability is the first empirical moment

$$\overline{X} = \frac{X_1 + X_2 + \cdots + X_n}{n} = \frac{\text{number of successes}}{n}.$$

The confidence interval for the probability of success

The sense of central limit theorem informally can be formulated as follows: let $\langle X_1, \ldots, X_n \rangle$ be the random sample of random variable of interest. Then for large values of n the values of statistics

$$Z = \frac{X_1 + X_2 + \cdots + X_n - np}{\sqrt{np(1-p)}} = \frac{\overline{X} - p}{\sqrt{p(1-p)/n}}$$

are distributed approximately as the values of standard normal variable.

The confidence interval for the probability of success

Let $0 < Q < 1$ and $z_{(1+Q)/2}$ is the $(1+Q)/2$ quantile of the standard normal variable. Then we believe (on the basis of central limit theorem) that

$$P(-z_{\frac{1+Q}{2}} < Z < z_{\frac{1+Q}{2}}) = P\Big(- z_{\frac{1+Q}{2}} < \frac{\overline{X} - p}{\sqrt{p(1-p)/n}} < z_{\frac{1+Q}{2}}\Big) \approx Q.$$

This equality can be rewritten as:

$$P\Big(\overline{X} - z\frac{\sqrt{p(1-p)}}{\sqrt{n}} < p < \overline{X} + z\frac{\sqrt{p(1-p)}}{\sqrt{n}}\Big) \approx Q, \quad z = z_{\frac{1+Q}{2}}.$$

Statistical hypothesis

Two hypothesis related to the distribution of the random variable observed are formulated : the null hypothesis $\mathbf{H}_0$ against the alternative hypothesis $\mathbf{H}_1$. Using the data of the sample we decide which one should be accepted. There are two possibilities: $\mathbf{H}_0$ is true or wrong and there are two possible decisions: to accept $\mathbf{H}_0$ or reject it. Then we have four cases

	$\mathbf{H}_0$ true	$\mathbf{H}_0$ wrong
$\mathbf{H}_0$ accepted	true decision	error of II type
$\mathbf{H}_0$ rejected	error of I type	true decision

The statistical hypothesis

Usually the model is formulated in a way that the error of I type is more important, i.e. should be controlled. For to control it let us choose some small number $0 < \alpha < 1$ and create the decision procedure which ensures that

$$P(\text{error of I type}) = P(\mathbf{H}_0 \text{ rejected}|\mathbf{H}_0 \text{ true}) \leqslant \alpha.$$

The number α is called significance level of testing. There are however many criteria with the same significance level. If we always accept the hypothesis H_0, then the inequality is also satisfied. In the set of all criteria with the same significance level one tries to choose that one which minimizes the probability

$$P(\text{error of II type}) = P(\mathbf{H}_0 \text{ accepted}|\mathbf{H}_0 \text{ false}).$$

Hypothesis about the expectation of the normal random variable

Hypothesis about the expectation of the normal random variable, if the variance is known

The random variable observed is $X \sim \mathcal{N}(\mu, \sigma^2)$, the variance σ^2 is known, $\langle X_1, X_2, \ldots, X_n \rangle$ is the random sample of X.

Hypotheses:

$$\mathbf{H}_0 : \mu = \mu_0,$$
$$\mathbf{H}_1 : \mu \neq \mu_0,$$

α is the significance level, z – solution of the equation $\Phi(z) = 1 - \alpha/2$, i.e. the $\alpha/2$ critical value of the standard normal random variable,

$$Z = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}.$$

Decision: if $|Z| > z$, the hypothesis $\mathbf{H}_0$ is rejected, if $|Z| \leq z$, hypothesis $\mathbf{H}_0$ is accepted.

Hypothesis about the expectation of the normal random variable

Hypothesis about the expectation of the normal random variable, if the variance is unknown

The random variable observed is $X \sim \mathcal{N}(\mu, \sigma^2)$, the variance σ^2 is unknown, $\langle X_1, X_2, \ldots, X_n \rangle$ is the random sample.

Hypotheses:

$$\mathbf{H}_0 : \mu = \mu_0,$$
$$\mathbf{H}_1 : \mu \neq \mu_0,$$

α is the significance level, t – solution of the equation

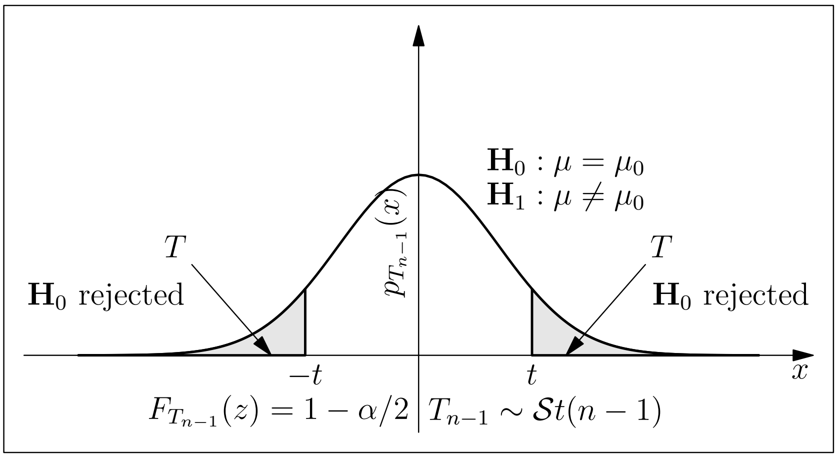
$$F_{T_{n-1}}(t) = 1 - \alpha/2, \quad T_{n-1} \sim \mathcal{S}t(n-1),$$

i.e. $\alpha/2$ critical value of t-distribution with $n-1$ degrees of freedom,

$$T = \frac{\overline{X} - \mu_0}{S/\sqrt{n}}.$$

Hypothesis about the expectation of the normal random variable

Decision: if $|T| > t$, hypothesis $\mathbf{H}_0$ is rejected, if $|T| \leqslant t$, hypothesis $\mathbf{H}_0$ is accepted.



364 / 369

Hypothesis about the probability of success

The random variable X takes the value 1 if the trial is successful, and the value 0 if one gets a failure. Let $\langle X_1, X_2, \ldots, X_n \rangle$ be the random sample, the number of trials n is large, α is the significance level,

$$P(X = 1) = p, \quad P(X = 0) = 1 - p, \quad Z = \frac{\overline{X} - p_0}{\sqrt{p_0(1 - p_0)/n}}.$$

Hypotheses about the probability of success p:

$$\mathbf{H}_0 : p = p_0,$$
$$\mathbf{H}_1 : p \neq p_0.$$

If $|Z| \geqslant z$, where z is the $\alpha/2$ critical value of the standard normal variable, the null hypothesis H_0 is rejected, if $|Z| < z$, the null hypothesis is accepted.
If the alternative hypothesis is $\mathbf{H}_1 : p > p_0$ or $\mathbf{H}_1 : p < p_0$, the α critical value of the standard normal random variable is used. In the first case the null hypothesis is rejected if $Z \geqslant z$, in the second case – if $Z \leqslant -z$.

365 / 369

Hypothesis about the probability of success

The random variable X takes the value $1,$ if the trial is successful and the value $0,$ if one gets a failure. Let $\langle X_1, X_2, \ldots, X_n\rangle$ be the random sample, the number of trials n is small, α – the significance level,

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

Hypotheses about the probability of success p:

$$\mathbf{H}_0 : p = p_0,$$
$$\mathbf{H}_1 : p \neq p_0$$

.

Hypothesis about the probability of success

Let the value of statistics $S_n = X_1 + X_2 + \cdots + X_n$ is u. Compute the probabilities

$$t_1 = P(S_n \geq u|\mathbf{H}_0) = \sum_{i=u}^{n} C_n^i p_0^i (1-p_0)^{n-i},$$
$$t_2 = P(S_n \leq u|\mathbf{H}_0) = \sum_{i=0}^{u} C_n^i p_0^i (1-p_0)^{n-i}.$$

If the alternative hypothesis is $\mathbf{H}_1 : p \neq p_0$, then it is accepted (the null hypothesis rejected), if at least one of probabilities t_1, t_2 is less than $\alpha/2.$
If the alternative hypothesis is $\mathbf{H}_1 : p > p_0$, then it is accepted, if $t_1 < \alpha.$
If the alternative hypothesis is $\mathbf{H}_1 : p < p_0$, then it is accepted, if $t_2 < \alpha.$

Example

Problem 13. *Two sacks contain the barley grain with admixture of oat. The percentage of out in the sacks is* 20%, *and* 30%. *We have to choose the sack with smaller percentage of oat. For to decide wich one of the sacks should be chosen, we took from one sack* $n = 758$ *grains and found among them* $m = 166$ *grains of oat. Test the hypothesis than the sack with smaller percentage of oat was chosen, if the significance level is* $\alpha = 0,2$

Example

Problem 14. *If the coin is symmetrical anybody can guess approximately on half of results of tosses. The magician claims that he can guess more than a half of results of tossing a symmetrical coin. We want to test the hypothesis that his capability to guess is as of ordinary people against the claim that he can guess better. Let the significance level be* $\alpha = 0,1$. *If the number of tosses is* $n = 15$, *how many times the magician should guess correctly for to convince the audience to accept the alternative hypothesis?*