Elements of Probability Theory and Mathematical Statistics

By

Gogi Pantsulaia, Zurab Kvatadze and Givi Giorgadze

Georgian Technical University
Tbilisi 2013

Contents

Introduction

The modern probability theory is an interesting and most important part of mathematics, which has great achievements and close connections both with classical parts of mathematics (geometry, mathematical analysis, functional analysis), and its various branches(theory of random processes, theory of ergodicity, theory of dynamical system, mathematical statistics and so on). The development of these branches of mathematics is mainly connected with the problems of statistical mechanics, statistical physics, statistical radio engineering and also with the problems of complicated systems which consider the random and the chaotic influence. At the origin of the probability theory were standing such famous mathematicians as I.Bernoulli, P.Laplace, S.Poisson, A.Cauchy, G.Cantor, F.Borel, A.Lebesgue and others. A very controversial problem connected with the relation between the probability theory and mathematics was entered in the list of unsolved mathematical problems raised by D.Gilbert in 1900. This problem has been solved by Russian mathematician A.Kolmogorov in 1933 who gave us a strict axiomatic basis of the probability theory. A.Kolmogorov conception to the basis of the probability theory is applied in the present book. Giving a strong system of axioms (according to A.Kolmogorov) the general probability spaces and their c$\acute{o}$omposite comp$\acute{o}$nents are described in the present book. The main purpose of the present book is to help students to acquire such skills that are necessary to construct mathematical models (i.e., probability spaces) of various (social, economical, biological, mechanical, physical, etc) processes and to calculate their numerical characteristics. In this sense the last chapters (in particular, chapters 14-15) are of interest, where some applications of various mathematical models(Markov chains, Brownian motion, etc) are presented. The present book consists of twenty one chapters. More of chapters are equipped with exercises (i.e. tests), the solutions of which will help the student in deep comprehend and assimilation of experience of the presented elements of probability theory and mathematical statistics.

Chapter 1

Set-Theoretical Operations. Kolmogorov Axioms

Let Ω be a non-empty set and let $\mathcal{P}(\Omega)$ be a class of all subsets of Ω. ($\mathcal{P}(\Omega)$ is called a powerset of Ω).

Definition 1.1 Let, $A_k \in \mathcal{P}(\Omega)$ $(1 \le k \le n)$. An union of the finite family of subsets $(A_k)_{1 \le k \le n}$ is denoted by $\cup_{k=1}^{n} A_k$ and is defined by

$$\cup_{k=1}^{n} A_k = \{x | x \in A_1 \bigvee \cdots \bigvee x \in A_n\},$$

where $\bigvee$ denotes the symbol of conjunction.

Definition 1.2 Let $A_k \in \mathcal{P}(\Omega)$ $(k \in N)$. An union of the countable family of subsets $(A_k)_{k \in N}$ is denoted by $\cup_{k \in N} A_k$ and is defined by

$$\cup_{k \in N} A_k = \{x | x \in A_1 \bigvee x \in A_2 \bigvee \cdots\}.$$

Definition 1.3 Let $A_k \in \mathcal{P}(\Omega)$ $(1 \le k \le n)$. An intersection of the finite family of subsets

$(A_k)_{1 \le k \le n}$ is denoted by the symbol $\cap_{k=1}^{n} A_k$ and is defined by

$$\cap_{k=1}^{n} A_k = \{x | x \in A_1 \bigwedge \cdots \bigwedge x \in A_n\},$$

where $\bigwedge$ denotes the symbol of disjunction.

Definition 1.4. Let $A_k \in \mathcal{P}(\Omega)$ $(k \in N)$. An intersection of the countable family of subsets $(A_k)_{k \in N}$ is denoted by the symbol $\cap_{k \in N} A_k$ and is defined by

$$\cap_{k \in N} A_k = \{x | x \in A_1 \bigwedge x \in A_2 \bigwedge \cdots\}.$$

Definition 1.5. Let $A, B \in \mathcal{P}(\Omega)$. A difference of subsets A and B is denoted by the symbol

$A \setminus B$ and is defined by

$$A \setminus B = \{x | x \in A \bigwedge x \notin B\}.$$

Remark 1.1 De-Morgan's formulas are central for the theory of probability :

1) $\Omega \setminus \cup_{k=1}^{n} A_k = \cap_{k=1}^{n} (\Omega \setminus A_k)$;
2) $\Omega \setminus \cup_{k \in N} A_k = \cap_{k \in N} (\Omega \setminus A_k)$;
3) $\Omega \setminus \cap_{k=1}^{n} A_k = \cup_{k=1}^{n} (\Omega \setminus A_k)$;
4) $\Omega \setminus \cap_{k \in N} A_k = \cup_{k \in N} (\Omega \setminus A_k)$.

Definition 1.6. A class $\mathcal{A}$ of subsets Ω is called an algebra if the following conditions are satisfying :

1) $\Omega \in \mathcal{A}$;
2) If $A, B \in \mathcal{A}$, then $A \cup B \in \mathcal{A}$ and $A \cap B \in \mathcal{A}$;
3) If $A \in \mathcal{A}$, then $\Omega \setminus A \in \mathcal{A}$.

Remark 1.2. In the condition 2) it is sufficient to require only the validity $A \cup B \in \mathcal{A}$ or $A \cap B \in \mathcal{A}$, because applying Remark 1.1, the following set-theoretical equalities are true:

$$A \cup B = \Omega \setminus ((\Omega \setminus A) \cap (\Omega \setminus B)),$$

$$A \cap B = \Omega \setminus ((\Omega \setminus A) \cup (\Omega \setminus B)).$$

Remark 1.3 The algebra is such class of subsets of Ω which is closed under finite number of set-theoretical operations " $\cap, \cup, \setminus$ " .

Definition 1.7. A class $\mathcal{F}$ of subsets of Ω is called σ-algebra if :

1) $\Omega \in \mathcal{F}$;
2) If $A_k \in \mathcal{F}$ $(k \in N)$, then $\cup_{k \in N} A_k \in \mathcal{F}$ and $\cap_{k \in N} A_k \in \mathcal{F}$;
3) If $A \in \mathcal{F}$, then $\Omega \setminus A \in \mathcal{F}$.

Remark 1.4 The σ-algebra is such class of subsets of Ω which is closed under countable number of set-theoretical operations " $\cap, \cup, \setminus$ " .

Definition 1. 8. A real-valued function P defined on the σ-algebra $\mathcal{F}$ of subsets of Ω is called a probability, if:

1) For arbitrary $A \in \mathcal{F}$ we have $P(A) \geq 0$ (The property of the non-negativity);
2) $P(\Omega) = 1$ (The property of the normality);
3) If $(A_k)_{k \in N}$ is pairwise disjoint family of elements $\mathcal{F}$ then $P(\cup_{k \in N} A_k) = \sum_{k \in N} P(A_k)$ (The property of countable-additivity).

Kolmogorov[1] **axioms.** The triplet $(\Omega, \mathcal{F}, P)$, where

[1] Andrey Kolmogorov [12(25).4.1903 Tambov-25.10.1987 Moscow] Russian mathematician, Academician of the Academy Sciences of the USSR (1939), Professor of the Moscow State University. He has firstly considered a mathematical conception of the axiomatical foundation of the probability theory in 1933.

1) Ω is a non-empty set,
2) $\mathcal{F}$ is a σ-algebra of subsets of Ω,
3) P is a probability defined on $\mathcal{F}$, is called a probability space.

Ω is called a space of all elementary events; An arbitrary point $\omega \in \Omega$ is called elementary event; An arbitrary element of $\mathcal{F}$ is called an event; $\emptyset$ is called an impossible event; Ω is called a necessary event; For arbitrary event A an event $\overline{A} = \Omega \setminus A$ is called its complementary event ; The product of events A and B is denoted by AB and is defined by $A \cap B$; The events A and B are called non-consistent if the event AB is an impossible event; A sum of two non-consistent events A and B is denoted by $A+B$ and is defined by $A \cup B$; For arbitrary event A the number $P(A)$ is called a probability of the event A .

Definition 1.9 A sum of pairwise disjoint events $(A_k)_{k\in N}$ is denoted by the symbol $\sum_{k\in N} A_k$ and is defined by

$$\sum_{k\in N} A_k = \cup_{k\in N} A_k.$$

Remark 1.4 Like the numerical operations of sums and product, the set theoretical operations have the following properties:
1) $A+B = B+A,\ AB = BA,$
2) $(A+B)+C = A+(B+C), (AB)C = A(BC),$
3) $(A+B)C = AC+BC,\ C(A+B) = CA+CB,$
4) $C(\sum_{k\in N} A_k) = \sum_{k\in N} CA_k,$
5) $(\sum_{k\in N} A_k)C = \sum_{k\in N} A_kC.$

Tests

1.1.Assume that $A_k = [\frac{k+1}{k+2}, 1]\ (k \in N)$. Then

1) $\cap_{4\le k\le 10} A_k$ coincides with
a) $[\frac{1}{2}, 1]$, b) $]\frac{11}{12}, 1]$, c) $[\frac{11}{12}, 1]$, d) $[\frac{1}{2}, 1]$;

2) $\cup_{3\le k\le 10} A_k$ coincides with
a) $[\frac{4}{5}, 1]$, b) $[\frac{3}{4}, 1]$, c) $[\frac{2}{3}, 1]$, d) $[\frac{5}{6}, 1]$;

3) $\cup_{2\le k\le 10} A_k \setminus \cap_{1\le k\le 10} A_k$ coincides with
a) $[\frac{3}{4}, \frac{11}{12}[$, b) $[\frac{4}{5}, \frac{12}{13}[$, c) $[\frac{4}{5}, 1[$, d) $[\frac{5}{6}, 1[$;

4) $\cap_{k\in N} A_k$ coincides with
a) $\{1\}$, b) $\{0\}$, c) $\{\emptyset\}$, d) $[0, 1]$;

5) $\cup_{k\in N} A_k$ coincides with
a) $[\frac{4}{5}, 1]$, b) $[\frac{3}{4}, 1]$, c) $[\frac{2}{3}, 1]$, d) $[\frac{5}{6}, 1]$;

6) $\cup_{k\in N} A_k \setminus \cap_{k\in N} A_k$ coincides with
a) $[\frac{3}{4}, 1[$, b) $[\frac{2}{3}, 1[$, c) $[\frac{4}{5}, 1[$, d) $[\frac{5}{6}, 1[$.

1.2. Assume that $A_k = [\frac{k-3}{3k}, \frac{2k+3}{3k}]\ (k \in N)$. Then

1) $\cap_{5\le k\le 10} A_k$ coincides with
a) $[\frac{8}{33}, \frac{25}{33}]$, b) $[\frac{7}{30}, \frac{23}{30}]$, c) $[\frac{2}{15}, \frac{13}{15}]$, d) $[\frac{1}{12}, \frac{11}{12}]$;

2) $\cup_{10\le k\le 20} A_k$ coincides with

a) $[\frac{8}{33},\frac{25}{33}]$, b) $[\frac{7}{30},\frac{23}{30}]$, c) $[\frac{2}{15},\frac{13}{15}]$, d) $[\frac{1}{12},\frac{11}{12}]$;

3) $\cap_{k\in N}A_k$ coincides with

a) $[\frac{8}{33},\frac{25}{33}]$, b) $[\frac{1}{3},\frac{3}{4}]$, c) $[\frac{1}{3},\frac{2}{3}]$, d) $[\frac{1}{12},\frac{11}{12}]$;

4) $[0,1]\setminus\cap_{k\in N}A_k$ coincides with

a) $[0,1]\setminus[0,\frac{1}{3}]\cup]\frac{3}{4};1[$, b) $[0,\frac{1}{3}\cup]\frac{3}{4}\cup]\frac{3}{4};1[$, c) $[\frac{1}{3},\frac{2}{3}]$, d) $[\frac{1}{12},\frac{11}{12}]$.

1.3*. Let θ be a positive number such that $\frac{\theta}{\pi}$ is an irrational number. We set

$$\Delta=\{(x,y)|-1<x<1,\ -1<y<1\}.$$

Let denote by A_n a set obtained by counterclockwise rotation of the set Δ about the origin of the plane for the angle $n\theta$. Then

1) $\cap_{k\in N}A_k$ coincides with

a) $\{(x,y)|x^2+y^2\le 1\}$, b) $\{(x,y)|x^2+y^2\le 2\}$,
c) $\{(x,y)|x^2+y^2<1\}$, d) $\{(x,y)|x^2+y^2<2\}$;

2) $\cup_{k\in N}A_k$ coincides with

a) $\{(x,y)|x^2+y^2\le 1\}$, b) $\{(x,y)|x^2+y^2\le 2\}$,
c) $\{(x,y)|x^2+y^2<1\}$, d) $\{(x,y)|x^2+y^2<2\}$.

1.4. Suppose that $\Omega=\{0;1\}$.

1) The algebra of subsets of Ω is

a) $\{\{0\},\{0;1\}\}$, b) $\{\{0\};\{0;1\};\emptyset\}$,
c) $\{\{0\};\{1\};\{0;1\};\emptyset\}$; d) $\{\{1\},\{0;1\}\}$;

2) The σ-algebra of subsets of Ω is

a) $\{\{0\},\{0;1\}\}$, b) $\{\{0\};\{0;1\};\emptyset\}$,
c) $\{\{0\};\{1\};\{0;1\};\emptyset\}$; d) $\{\{1\},\{0;1\}\}$.

1.5. Assume that $\Omega=[0,1[$.

Then

1) the algebra of subsets of Ω is

a) $\{X|X\subset[0,1[, X$ is presented by the finite union of intervals open from the right and closed from the right$\}$,

b) $\{X|X\subset[0,1[,\ X$ is presented by the finite union of intervals closed from the right and open from the right$\}$,

c) $\{X|X\subset[0,1[,\ X$ is presented by the finite union of closed from both side intervals $\}$,

d) $\{X|X\subset[0,1[,\ X$ is presented by the finite union of open from both side intervals $\}$;

2) Suppose that $\mathcal{A}=\{X|X\subset[0,1[$ and X be presented as the finite union of intervals open from the right and closed from the left $\}$. Then $\mathcal{A}$

a) is not the algebra,

b) is the σ-algebra,

c) is the σ-algebra, but is not the algebra,

d) is the algebra, but is not the σ-algebra.

Chapter 2

Properties of Probabilities

Let $(\Omega, \mathcal{F}, P)$ be a probability space. Then the probability P has the following properties.

Property 2.1 $P(\emptyset) = 0$.

Proof. We have $\emptyset = \emptyset \cup \emptyset \cup \cdots$. From the property of countable-additivity of the probability P, we have

$$P(\emptyset) = \lim_{n \to \infty} nP(\emptyset).$$

Since P is finite, $P(\emptyset) \in R$. Hence, above-mentioned equality is possible if and only if $P(\emptyset) = 0$. □

Property 2.2 (The property of the finite-additivity). *If $(A_k)_{1 \le k \le n}$ is a finite family of pairwise disjoint events, then*

$$P(\cup_{k=1}^{n} A_k) = \sum_{k=1}^{n} P(A_k).$$

Proof. For arbitrary natural number $k > n$ we set $A_k = \emptyset$. Following Property 2.1 and the property of the countable-additivity of P we have

$$P(\cup_{k=1}^{n} A_k) = P(\cup_{k=1}^{\infty} A_k) = \sum_{k=1}^{\infty} P(A_k) = \sum_{k=1}^{n} P(A_k) + \sum_{k=n+1}^{\infty} P(A_k) = \sum_{k=1}^{n} P(A_k).$$

□

Property 2.3. *For $A \in \mathcal{F}$ we have*

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

Proof. Since$\Omega = A + \overline{A}$ and $P(\Omega) = 1$, using the property of the finitely-additivity, we have

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

It follows that

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

□

Property 2.4 *Suppose that $A, B \in \mathcal{F}$ and $A \subseteq B$. Then $P(B) = P(A) + P(B \setminus A)$.*

Proof. Using the equality $B = A + (B \setminus A)$ and the property of countably additivity of P, we have $P(B) = P(A) + P(B \setminus A)$.

□

Property 2.5 *Suppose that $A, B \in \mathcal{F}$ and $A \subseteq B$. Then $P(A) \leq P(B)$.*

Proof. Following Property 2.4, we have $P(B) = P(A) + P(B \setminus A)$. Hence $P(A) = P(B) - P(B \setminus A) \leq P(B)$. □

Property 2.6. *Suppose that $A, B \in \mathcal{F}$. Then*

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

Proof. Using the representation $A \cup B = (A \setminus B) + AB + (B \setminus A)$ and the property of finitely-additivity of P, we have

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

□

Property 2.7 *Suppose that $A, B \in \mathcal{F}$. Then*

$$P(A \cup B) \leq P(A) + P(B).$$

Proof. Following Property 2.6, we have

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

It yields

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

□

Property 2.8 (Continuity from above) *Assume that* $(A_n)_{n\in N}$ *be an decreasing sequence of events, i.e.,*

$$(\forall n)(n \in N \to A_{n+1} \subseteq A_n).$$

Then the following

$$P(\cap_{n\in N} A_n) = \lim_{n\to\infty} P(A_n)$$

holds.

Proof. For $n \in N$ we have

$$A_n = \cap_{k\in N} A_k + (A_n \setminus A_{n+1}) + (A_{n+1} \setminus A_{n+2}) + \cdots .$$

Using the property of the countably-additivity of P, we obtain

$$P(A_n) - P(\cap_{k\in N} A_k) = \sum_{p=1}^{\infty} P(A_{n+p} \setminus A_{n+p+1}).$$

Note that the sum $\sum_{p=1}^{\infty} P(A_{n+p} \setminus A_{n+p+1})$ is the n-th residual series of absolutely convergent series $\sum_{n=1}^{\infty} P(A_n \setminus A_{n+1})$. From the necessary and sufficient condition of the convergence of the numerical series, we have

$$\lim_{n\to\infty} \sum_{p=1}^{\infty} P(A_{n+p} \setminus A_{n+p+1}) = 0.$$

It means that

$$\lim_{n\to\infty} (P(A_n) - P(\cap_{k\in N} A_k)) = \lim_{n\to\infty} P(A_n) - P(\cap_{k\in N} A_k) = 0,$$

i.e.,

$$\lim_{n\to\infty} P(A_n) = P(\cap_{k\in N} A_k).$$

□

Property 2.9 (Continuity from below) *Let* $(B_n)_{n\in N}$ *be an increasing sequence of events, i.e.,*

$$(\forall n)(n \in N \to B_n \subseteq B_{n+1}).$$

Then the following equality is valid

$$P(\cup_{n\in N} B_n) = \lim_{n\to\infty} P(B_n).$$

Proof. For $\cup_{n\in N}B_n$ we have the following representation

$$\cup_{n\in N}B_n = B_1 + (B_2 \setminus B_1) + \cdots + (B_{k+1} \setminus B_k) + \cdots .$$

Following the property of the countable-additivity of P, we get

$$P(\cup_{n\in N}B_n) = P(B_1) + P(B_2 \setminus B_1) + \cdots + P(B_{k+1} \setminus B_k) + \cdots .$$

From Property 2.4 we have

$$P(B_{k+1}) = P(B_k) + P(B_{k+1} \setminus B_k).$$

If we define $P(B_{k+1} \setminus B_k)$ from the above-mentioned equality and enter it in early considered equality we obtain

$$P(\cup_{n\in N}B_n) = P(B_1) + (P(B_2) - P(B_1)) + \cdots + (P(B_{k+1}) - P(B_k)) + \cdots .$$

Note that the series on the right is convergent. For the sum S_n of the first n members we have

$$S_n = P(B_n).$$

From the definition of the series sum, we obtain

$$P(\cup_{n\in N}B_n) = \lim_{n\to\infty} S_n = \lim_{n\to\infty} P(B_n).$$

□

Tests

Assume that $(\Omega, \mathcal{F}, P)$ be a probability space.

2.1. If $P(A) = 0,95$, then $P(A)$ is equal to

a) $0,56$, b) $0,55$, c) $0,05$, d) $0,03$.

2.2 Assume that$A, B \in \mathcal{F}$, $A \subset B$, $P(A) = 0,65$ and $P(B) = 0,68$. Then $P(B \setminus A)$ is equal to

a) $0,02$, b) $0,03$, c) $0,04$, d) $0,05$.

2.3 Assume that $A, B \in \mathcal{F}$, $P(A) = 0,35$, $P(B) = 0,45$ and $P(A \cup B) = 0,75$. Then $P(A \cap B)$ is equal to

a) $0,02$, b) $0,03$, c) $0,04$, d) $0,05$.

2.4 Let $(A_n)_{n\in N}$ be a decreasing sequence of events and $P(\cap_{n\in N}A_n) = 0,89$. Then $\lim_{n\to\infty} P(\overline{A_n})$ is equal to

a) $0,11$, b) $0,12$, c) $0,13$, d) $0,14$.

2.5 Let $(A_n)_{n\in N}$ be a decreasing sequence of events and $P(A_n) = \frac{n+1}{3n}$. Then $P(\cap_{n\in N}A_n)$ is equal to

a) $\frac{1}{2}$, b) $\frac{1}{3}$, c) $\frac{1}{4}$, d) $\frac{1}{5}$.

2.6. Let $(A_n)_{n\in N}$ be an increasing sequence of events and $P(\cup_{n\in N}A_n) = 0,89$. Then $\lim_{n\to\infty} P(\overline{A_n})$ is equal to

a) $0,11$, b) $0,12$, c) $0,13$, d) $0,14$.

2.7. Let $(A_n)_{n\in N}$ be an increasing sequence of events and $P(A_n) = \frac{n-1}{3n}$. Then

a) $\frac{1}{2}$, b) $\frac{1}{3}$, c) $\frac{1}{4}$, d) $\frac{1}{5}$.

Chapter 3

Examples of Probability Spaces

3.1. Classical probability space. Let Ω contains n points, i.e. $\Omega = \{\omega_1, \cdots, \omega_n\}$. We denote by $\mathcal{F}$ the class of all subsets of Ω. Let us define a real-valued function P on $\mathcal{F}$ by the following formula

$$(\forall A)(A \in \mathcal{F} \rightarrow P(A) = \frac{|A|}{|\Omega|}),$$

where $|\cdot|$ denotes a cardinality of the corresponding set. One can easily demonstrate that the triplet $(\Omega, \mathcal{F}, P)$ is a probability space. This probability space is called a classical probability space. The numerical function P is called a classical probability.

Definition 3.1 Let A be any event. We say that an elementary event $\omega \in \Omega$ is successful for the event A if $\omega \in A$. We obtain the following rule for calculation of the classical probability:

The classical probability of the event A is equal to the fraction a numerator of which is equal to the number of all successful elementary events for the event A and a denominator of which is equal to the number of all possible elementary events.

3.2 Geometric probability space. Let Ω be a Borel subset of the n- dimensional Euclidean space R^n with positive Borel[1] measure b_n(cf. Example 5. 3). Let denote by $\mathcal{F}$ the class of all Borel subsets of Ω. We set

$$(\forall A)(A \in \mathcal{F} \rightarrow P(A) = \frac{b_n(A)}{b_n(\Omega)}).$$

The triplet $(\Omega, \mathcal{F}, P)$ is called an n-dimensional geometrical probability space associated with the Borel set Ω. The function P is called n-dimensional geometrical probability defined on Ω. When a point is falling in the set $\Omega \subset R^n$ and the probability that this point

[1] Borel Felix Eduard Justion Emil (7.01. 1871 3.03.1956.)-French mathematician, member of the Paris Academy of Sciences (1921), professor of the Paris University (1909-1941).)

will fall in any Borel subset of $A \subset \Omega$ is proportional to its Borel b_n-measure, then we have the following rule for a calculation of a geometric probability:

The geometrical probability of the event that a point will fall in the Borel subset $A \subset \Omega$ is equal to the fraction with a numerator $b_n(A)$ and a denominator $b_n(\Omega)$.

Let us consider some examples demonstrating how we can model probability spaces describing random experiments.

Example 3.1

Experiment. We roll a six-sided dice.

Problem. What is the probability that we will roll an even number?

Modelling of the random experiment. Since the result of the random experiment is an elementary event, a space of all elementary events Ω has the following form

$$\Omega = \{1;2;3;4;5;6\}.$$

We denote by $\mathcal{F}$ a σ-algebra of all subsets of Ω(i.e. the powerset of Ω). It is clear, that

$$\mathcal{F} = \{\emptyset;\{1\};\cdots\{6\};\{1;2\};\cdots\{1;2;3;4;5;6\}\}.$$

Let denote by P a classical probability measure defined by

$$(\forall A)(A \in \mathcal{F} \to P(A) = \frac{|A|}{6}).$$

The triplet $(\Omega, \mathcal{F}, P)$ is the probability space (i.e. the stochastic mathematical model) which describes our experiment.

Solution of the problem. We must calculate the probability of the event B having the following form

$$B = \{2;4;6\}.$$

By definition of P, we have

$$P(B) = \frac{|B|}{6} = \frac{3}{6} = \frac{1}{2};$$

Conclusion. The probability that we roll an even number is equal to $\frac{1}{2}$.

Example 3.2

Experiment. We accidentally choose 3 cards from the complect of 36 cards.

Problem. What is the probability that in these 3 cards one will be ”ace”?

Modelling of the experiment. Since the result of the random experiment is an elementary event and it coincides with tree cards, the space of all elementary events would be the space of all possible different tree cards. It is clear that

$$|\Omega| = C_{36}^{3}.$$

We denote by $\mathcal{F}$ a σ-algebra of all subsets of Ω. Let define a probability P by the following formula

$$(\forall A)(A \in \mathcal{F} \to P(A) = \frac{|A|}{C_{36}^{3}}).$$

Note, that $(\Omega, \mathcal{F}, P)$ is the probability space describing our experiment.

Solution of the problem. If we will choose 1 card from the complect of aces and 2 cards from the other cards, then considering all their possible combinations, we will obtain the set A of all threes of cards where at least one card is ace. It is clear that number of A is equal to $C_4^1 \cdot C_{32}^2$. By the definition of P we have

$$P(A) = \frac{|A|}{C_{36}^{3}} = \frac{C_4^1 \cdot C_{32}^2}{C_{36}^{3}}.$$

Conclusion. If we choose accidentally 3 cards from the complect of 36 cards, then the probability of the event that between them at list one card will be "ace" is equal $\frac{C_4^1 \cdot C_{32}^2}{C_{36}^{3}}$.

Example 3.3

Experiment. There are passed parallel lines on the plane such that the distance between neighboring lines is equal to $2a$. A $2l$ $(2l < 2a)$-long needle is thrown accidentally on the plane.

Problem (Buffon)[2] What is the probability that the accidentally thrown on the plane needle intersects any of the above-mentioned parallel line?

Modelling of the experiment. The result of our experiment is an elementary event, which can be defined by x and φ, where x is the distance from the middle of the needle to the nearest line and φ is the angle between the needle and the above mentioned line. It is clear that x and φ satisfy the following conditions $0 \le x \le a, 0 \le \varphi \le \pi$. Hence, a space of all elementary events Ω is defined by

$$\Omega = [0;\pi] \times [0;a] = \{(\varphi;x) : 0 \le \varphi \le \pi,\ 0 \le x \le a\}.$$

[2]Buffon Georges Louis Leclerc (7.9.1707 -16.4.1788) French experimentalist, member of the Petersburgs Academy of Sciences (1776), member of Paris Academy of Sciences (1733). The first mathematician, who worked on the problems of geometrical probabilities.)

We denote by $\mathcal{F}$ a class of all Borel subsets of Ω. Let define a probability P by the following formula:

$$(\forall A)(A \in \mathcal{F} \rightarrow P(A) = \frac{b_2(A)}{b_2(\Omega)}).$$

Evidently, $(\Omega, \mathcal{F}, P)$ is the probability space which describes our experiment.

Solution of the problem. It is clear that to the event the needle accidentally thrown on the plane intersects any above-mentioned parallel line - corresponds a subset B_0, defined by

$$B_0 = \{(\varphi, x)|\ 0 \leq \varphi \leq \pi,\ 0 \leq x \leq l \sin\varphi\}.$$

By the definition of P we have

$$P(B_0) = \frac{b_2(B)}{a \cdot \pi} = \frac{\int_0^{\pi} l \sin\varphi d\varphi}{a \cdot \pi} = \frac{2l}{a\pi}.$$

Conclusion. The probability of the event that the needle accidentally thrown on the plane will intersect any parallel line is equal to $\frac{2l}{a\pi}$.

Tests

3.1. There are 5 white and 10 black balls in the box. The probability that the accidentally chosen ball would be black is equal to

a) $\frac{1}{3}$, b) $\frac{2}{3}$, c) $\frac{1}{5}$, d) $\frac{1}{6}$.

3.2. There are 7 white and 13 red balls in the box. The probability that between accidentally chosen 3 balls 2 balls would be red is equal to

a) $\frac{C_{13}^2 \cdot C_7^1}{C_{20}^3}$, b) $\frac{C_{13}^1 \cdot C_7^2}{C_{20}^3}$, c) $\frac{C_{13}^2 \cdot C_7^2}{C_{20}^3}$, d) $\frac{C_{13}^1 \cdot C_7^1}{C_{20}^3}$.

3.3. We roll two six-sided dices. The probability that the sum of dices's numbers is less than 8, is equal to

a) $\frac{13}{18}$, b) $\frac{5}{6}$, c) $\frac{1}{5}$, d) $\frac{1}{6}$.

3.4. There are 17 students in the group. 8 of them are boys. There are staged 7 tickets to be drawn. The probability that between owners of tickets are 4 boys, is equal to

a) $\frac{C_{13}^2 \cdot C_7^2}{C_{15}^7}$, b)$\frac{C_8^1 \cdot C_7^2}{C_{17}^7}$, c) $\frac{C_8^4 \cdot C_9^3}{C_{17}^7}$, d) $\frac{C_{13}^1 \cdot C_7^1}{C_{25}^7}$.

3.5.A cube, each side of which is painted, is divided on 1000 equal cubes. The obtained cubes are mixed. The classical probability, that an accidentally chosen cube

1)has 3 painted sides, is equal to

a) $\frac{1}{1000}$, b) $\frac{1}{125}$, c) $\frac{1}{250}$, d) $\frac{1}{400}$;

2)has 2 painted sides, is equal to

a) $\frac{12}{124}$, b) $\frac{11}{120}$, c) $\frac{12}{125}$, d) $\frac{9}{125}$;

3) has 1 painted side, is equal to

a) $\frac{54}{250}$, b) $\frac{43}{145}$, c) $\frac{48}{125}$, d) $\frac{243}{250}$;

4) has no painted side, is equal to

a) $\frac{8}{250}$, b) $\frac{64}{125}$, c) $\frac{4}{165}$, d) $\frac{23}{250}$.

3.6. A group of 10 girls and 10 boys is accidentally divided into two subgroups. The classical probability that in both subgroups the numbers of girls and boys will be equal, is

a) $\frac{(C_{10}^5)^2}{C_{20}^{10}}$, b) $\frac{C_{10}^5}{C_{20}^{10}}$, c) $\frac{(C_{10}^5)^3}{C_{20}^{10}}$, d) $\frac{C_{10}^5}{C_{20}^5}$.

3.7. We have 5 segments with lengths 1, 3, 4, 7 and 9. The classical probability that by accidently choosing 3 segments we can construct a triangle, is equal to

a) $\frac{3}{C_5^3}$, b) $\frac{2}{C_5^3}$, c) $\frac{4}{C_5^3}$, d) $\frac{5}{C_5^3}$.

3.8. When roll two six-sided dice, the classical probability that

1) the sum of cast numbers is less than 5, is equal to

a) $\frac{7}{36}$, b) $\frac{5}{18}$, c) $\frac{1}{4}$, d) $\frac{3}{9}$;

2) we roll 5 by any dice, is equal to

a) $\frac{7}{36}$, b) $\frac{8}{36}$, c) $\frac{11}{36}$, d) $\frac{3}{19}$;

3) we roll only one 5 , is equal to

a) $\frac{7}{36}$, b) $\frac{5}{18}$, c) $\frac{10}{36}$, d) $\frac{12}{19}$;

4) the sum of rolled numbers is divided by 3, is equal to

a) $\frac{1}{3}$, b) $\frac{2}{5}$, c) $\frac{1}{6}$, d) $\frac{2}{9}$;

5) the module of the difference of rolled numbers is equal to 3, is

a) $\frac{1}{6}$, b) $\frac{2}{5}$, c) $\frac{2}{6}$, d) $\frac{2}{5}$;

6) the product of rolled numbers is simple, is equal to

a) $\frac{5}{36}$, b) $\frac{7}{36}$, c) $\frac{11}{36}$, d) $\frac{2}{36}$.

3.9. We choose a point from a square with inscribed circle. The geometrical probability that the chosen point does not belong to the circle, is equal to

a) $1-\frac{\pi}{3}$, b) $1-\frac{\pi}{4}$, c) $1-\frac{\pi}{5}$, d) $1-\frac{\pi}{6}$.

3.10. The telephone line is damaged by storm between 160 and 290 kilometers. The probability that this line is damaged between 200 and 240 kilometers, is equal to

a) $\frac{1}{13}$, b) $\frac{2}{13}$, c) $\frac{4}{13}$, d) $\frac{5}{13}$.

3.11. The distance between point A and the center of the circle with radius R is equal to $d(d > R)$. Then:

1) the probability that an accidentally drawing line with the origin at point A, will intersect the circle, is equal to

a) $\frac{2\ \arcsin(\frac{R}{d})}{\pi}$, b) $\frac{3\ arcsin(\frac{R}{d})}{\pi}$, c) $\frac{arcsin(\frac{R}{d})}{\pi}$, d) $\frac{2\ arcsin(\frac{2R}{d})}{\pi}$;

2) the probability that an accidentally drawing ray with origin A, will intersect the circle, is equal to

a) $\frac{2arcsin(\frac{R}{d})}{\pi}$, b) $\frac{3arcsin(\frac{R}{d})}{\pi}$, c) $\frac{arcsin(\frac{R}{d})}{\pi}$, d) $\frac{2arcsin(\frac{2R}{d})}{\pi}$.

3.12. We accidentally choose a point in a cube, in which is inscribed a ball. The geometrical probability that an accidentally chosen point does not belong to the ball, is equal to

a) $1-\frac{\pi}{3}$, b) $1-\frac{\pi}{4}$, c) $1-\frac{\pi}{5}$, d) $1-\frac{\pi}{6}$.

3.13. We accidentally choose a point in a ball, in which is inscribed a cube. The geometrical probability that an accidentally chosen point does not belong to the cube, is equal to

a) $1-\frac{2\sqrt{3}}{3\pi}$, b) $1-\frac{\sqrt{3}}{4\pi}$, c) $1-\frac{\sqrt{3}}{5\pi}$, d) $1-\frac{\sqrt{3}}{6\pi}$.

3.14. We accidentally choose a point in a tetrahedron , in which is inscribed a ball. The geometrical probability that an accidentally chosen point does not belong to the ball, is equal to

a) $1-\frac{5\pi\sqrt{2}}{48}$, b) $\frac{5\pi\sqrt{2}}{45}$, c) $1-\frac{\sqrt{3}\pi}{18}$, d) $1-\frac{5\pi\sqrt{2}}{80}$.

3.15. We accidentally choose a point in a ball, in which is inscribed a tetrahedron. The geometrical probability that an accidentally chosen point does not belong to the tetrahedron, is equal to

a) $1-\frac{12\sqrt{3}}{45\pi}$, b) $1-\frac{1}{9\pi}$, c) $1-\frac{12\sqrt{3}}{47\pi}$, d) $1-\frac{12\sqrt{3}}{43\pi}$.

3.16. We accidentally choose a point M in a square Δ, which is defined by

$$\Delta = \{(x,y) : 0 \le x \le 1,\ 0 \le y \le 1\}.$$

The geometrical probability that coordinates (x,y) of the point M satisfy the following condition

$$x+y \ge \frac{1}{2},$$

is equal to

a) $\frac{7}{8}$, b) $\frac{7}{9}$, c) $\frac{3}{5}$, d) $\frac{4}{5}$.

3.17. We accidentally choose a point M in a square Δ , which is defined by

$$\Delta = \{(x,y) : 0 \le x \le \frac{\pi}{2},\ 0 \le y \le \frac{\pi}{2}\}.$$

The geometrical probability that coordinates (x,y) of the point M satisfy the following condition

$$\sin(x) \le y \le x,$$

is equal to a) $1+\frac{\pi^2}{4}$, b) $1+\frac{\pi^2}{8}$, c) $1+\frac{\pi^2}{12}$, d) $1+\frac{\pi^2}{16}$.

3.18. We accidentally choose a point M in a cube Δ, which is defined by

$$\Delta = \{(x,y,z) : 0 \le x \le 1,\ 0 \le y \le 1,\ 0 \le z \le 1\}.$$

The geometrical probability that coordinates (x,y,z) of the point M satisfy the following condition

$$x^2+y^2+z^2 \le \frac{1}{4},\ x+y+z \ge \frac{1}{2},$$

is equal to

a) $\frac{\pi-1}{48}$, b) $\frac{4\pi-1}{24}$, c) $\frac{\pi-2}{50}$, d) $\frac{\pi+2}{50}$.

3.19. Two friends must meet at the concrete place in the interval of time $[12,\ 13]$. The friend which arrived first waits no longer than 20 minutes. The probability that the meeting between friends will happen within the mentioned interval, is equal to

a) $\frac{5}{9}$, b) $\frac{5}{8}$, c) $\frac{5}{7}$, d) $\frac{6}{7}$.

3.20. A student has planned to take money out of the bank. It is possible that he comes to the bank in the interval of time 140015 140025. It is known also that the robbery of this bank is planned in the same interval of time and it will continue for 4 minutes. The probability that the student will be in the bank at moment of robbery is equal to

a) $\frac{1}{10}$, b) $\frac{1}{11}$, c) $\frac{1}{5}$, d) $\frac{1}{6}$.

3.21. We accidentally choose three points A, B and C on the circumference of a circle with radius R. The probability that ABC will be an acute triangle is equal to

a) $\frac{1}{3}$, b) $\frac{1}{4}$, c) $\frac{1}{5}$, d) $\frac{1}{6}$.

3.22. We accidentally choose two points C and D on an interval $[AB]$ with length l. The probability that we can construct a triangle by the obtained three intervals is equal to:

a) $\frac{1}{4}$, b) $\frac{1}{5}$, c) $\frac{1}{6}$, d) $\frac{1}{7}$.

3.23. We accidentally choose point $M = (p, q)$ in cube Δ which is defined by

$$\Delta = \{(p,q) : 0 \le p \le 1,\ 0 \le q \le 1\}.$$

The geometrical probability that the roots of the equation $x^2 + px + q = 0$ will be real numbers is equal to

a) $\frac{1}{12}$, b) $\frac{1}{13}$, c) $\frac{1}{5}$, d) $\frac{1}{6}$.

3.24. We accidentally choose point M from the sphere with radius R. The probability that distance ρ between M and the center of the above mentioned sphere satisfies condition $\frac{R}{2} < \rho < \frac{2R}{3}$, is equal to

a) $\frac{7}{27}$, b) $\frac{1}{4}$, c) $\frac{37}{216}$, d) $\frac{8}{29}$.

Chapter 4

Total Probability and Bayes' Formulas

Let $(\Omega, \mathcal{F}, P)$ be a probability space and let B be an event with positive probability (i.e., $P(B) > 0$). We denote with $P(\cdot \mid B)$ a real-valued function defined on the σ-algebra $\mathcal{F}$ by

$$(\forall X)(X \in \mathcal{F} \to P(X|B) = \frac{P(X \cap B)}{P(B)}).$$

The function $P(\cdot \mid B)$ is called a conditional probability relative to the hypothesis that the event B occurred. The number $P(X|B)$ is the probability of the event X relative to the hypothesis that the event B occurred.

Theorem 4.1. *If $B \in \mathcal{F}$ and $P(B) > 0$, then the conditional probability $P(\cdot \mid B)$ is the probability.*

Proof. We have to show :

1) $P(A|B) \geq 0$ for $A \in \mathcal{F}$;
2) $P(\Omega|B) = 1$;
3) If $(A_k)_{k \in N}$ is a pairwise-disjoint family of events, then

$$P(\cup_{k \in N} A_k|B) = \sum_{k \in N} P(A_k|B).$$

The validity of the item 1) follows from the definition of $P(\cdot \mid B)$ and from the non-negativity of the probability P. Indeed,

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \geq 0.$$

The validity of the item 2) follows from the following relations

$$P(\Omega|B) = \frac{P(\Omega \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1.$$

The validity of the item 3) follows from the countable-additivity of P and from the elementary fact that if $(A_n)_{n\in N}$ is a family of pairwise-disjoint events, then the family of events $(A_n \cap B)_{n\in N}$ also has the same property. Indeed,

$$P(\cup_{n\in N}A_n|B) = \frac{P((\cup_{n\in N}A_n)\cap B)}{P(B)} = \frac{P(\cup_{n\in N}(A_n\cap B))}{P(B)} =$$

$$= \frac{\sum_{n\in N}P(A_n\cap B)}{P(B)} = \sum_{n\in N}\frac{P(A_n\cap B)}{P(B)} = \sum_{n\in N}P(A_n \mid B)$$

This ends the proof of theorem.

□

Theorem 4.2 *If* $P(B) > 0$*, then* $P(\overline{B}|B) = 0$.

Proof.

$$P(\overline{B}|B) = \frac{P(\overline{B}\cap B)}{P(B)} = \frac{P(\emptyset)}{P(B)} = 0.$$

□

Definition 4.1 Two events A and B are called independent if

$$P(A\cap B) = P(A)\cdot P(B).$$

Example 4.1 Assume that $\Omega = \{(x,y) : x \in [0;1], y \in [0;1]\}$. Let $\mathcal{F}$ denotes a class of all Borel subsets of Ω. (cf. Chapter 5, Example 5.3). Let denote by P the classical Borel probability measure b_2 on Ω. Then two events

$$A = \{(x,y) : x \in [0;\frac{1}{2}], y \in [0;1]\},$$

$$B = \{(x,y) : x \in [0;1], y \in [\frac{1}{2};\frac{3}{4}]\}$$

are independent.

Indeed, on the one hand, we have

$$P(A\cap B) = b_2(A\cap B) = b_2(\{(x,y) : x \in [0;\frac{1}{2}], y \in [\frac{1}{2};\frac{3}{4}]\} = \frac{1}{2}\cdot\frac{1}{4} = \frac{1}{8}.$$

On the other hand, we have

$$P(A)\cdot P(B) = b_2(A)\cdot b_2(B) = \frac{1}{2}\cdot\frac{1}{4} = \frac{1}{8}.$$

Finally, we get

$$P(A\cap B) = P(A)\cdot P(B),$$

which shows us that two events A and B are independent.

Theorem 4.3 *If two events* A *and* B *are independent and* $P(B) > 0$*, then* $P(A|B) = P(A)$.

Proof. From the definition of the conditional probability, we have

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

The independence of events A and B gives $P(A \cap B) = P(A) \cdot P(B)$. Finally, we get

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

This ends the proof of theorem . □

Remark 4.1 Theorem 4.3 asserts that when events A and B are independent then any information concerned an occurrence or a non-occurrence of one of them does not influence on the probability of the occurrence of other.

Theorem 4.4 *If two events A and B are independent, then independent are also events $\overline{A}$ and B.*

Proof. We have

$$P(\overline{A} \cap B) = P((\Omega \setminus A) \cap B) = P((\Omega \cap B) \setminus (A \cap B)) =$$

$$= P(B \setminus (A \cap B)) = P(B) - P(A \cap B) = P(B) - P(A) \cdot P(B) =$$
$$= P(B)(1 - P(A)) = P(B) \cdot P(\overline{A}).$$

This ends the proof of theorem. □

Example 4.2

Experiment. We throw two six-sided dices.

Problem. What is the probability that the sum of thrown numbers is equal to 8 relative to the hypothesis that the sum of thrown numbers is even ?

Modelling of the experiment. A probability space Ω of all elementary events has the following form

$$\Omega = \{(x, y) : x \in N, y \in N,\ 1 \le x \le 6,\ 1 \le y \le 6\},$$

where x and y denote thrown numbers on the first and second dice, respectively.

We denote with $\mathcal{F}$ a class of all subsets of Ω. Let P denote a classical probability. Finally, a probability space $(\Omega, \mathcal{F}, P)$ describing our experiment is constructed.

Solution of the problem. Let denote with A a subset of Ω which corresponds to the event: ” The sum of thrown numbers is equal to 8 ”. Then, the event A has the following form:

$$A = \{\ (6;2);\ (5;3);\ (4;4);\ (3;5);\ (2;6)\ \}.$$

Let denote with B the following event: "The sum of thrown numbers is even". We have

$$B = \{(1;1);\ (1;3);\ (2;2);\ (3;1);\ (1;5);\ (2;4);\ (3;3);(4;4);\ (5;1);$$
$$(6;2);\ (5;3);(4;4);\ (3;5);\ (2;6);\ (6;4);\ (5;5);\ (4;6);(6;6)\ \}.$$

Note that $A \cap B = A$. By the definition of the classical probability we have

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{5}{36} : \frac{18}{36} = \frac{5}{18}.$$

Conclusion. If we throw two six-sided dices then the conditional probability that the sum of thrown numbers is equal to 8 concerning a hypothesis that the sum of thrown numbers is even, is equal to $\frac{5}{18}$.

Definition 4.2 Suppose that $J \subseteq N$. A family of event $(A_i)_{i \in J}$ is called a complete system of representatives if :

1) $A_i \cap A_j = \emptyset,\ i, j \in J, i \neq j,$
2) $(\forall j)(j \in J \to P(A_j) > 0),$
3) $\cup_{j \in J} A_j = \Omega.$

Theorem 4.5 *Suppose that $J \subseteq N$ and $(A_j)_{j \in J}$ be a complete system of representatives. For arbitrary event B the following formula*

$$P(B) = \sum_{j \in J} P(B|A_j) \cdot P(A_j)$$

is valid, which is called the formula of total probability.

Proof. We have

$$B = \cup_{j \in J}(B \cap A_k),$$

where $(B \cap A_j)_{j \in J}$ is a family of pairwise-disjoint events. Indeed, we have,

$$B = B \cap \Omega = B \cap (\cup_{j \in J} A_j) = \cup_{j \in J}(B \cap A_j).$$

From the countable-additivity of P we have

$$P(B) = \sum_{j \in J} P(B \cap A_j).$$

Note that for arbitrary natural number j $(j \in J)$ we have

$$P(B|A_j) = \frac{P(B \cap A_j)}{P(A_j)},$$

Hence,

$$P(B \cap A_j) = P(A_j) \cdot P(B|A_j).$$

Finally, we get

$$P(B) = \sum_{j \in J} P(B \cap A_j) = \sum_{j \in J} P(A_j) \cdot P(B|A_j).$$

This ends the proof of theorem. □

Example 4.3

Experiment. There are placed 3 white and 3 black balls in an urn I, 3 white and 4 black balls in an urn II and 4 white and 1 black balls in the urn III. We accidentally choose a box and further accidentally choose a ball from this urn.

Problem. What is the probability that accidentally chosen ball will be white if the probability of a choice of any urn is equal to $\frac{1}{3}$?.

Solution of the Problem. Let A_i denote an event that we have chosen i-th urn $(1 \leq i \leq 3)$. Then we obtain that $P(A_1) = P(A_2) = P(A_3) = \frac{1}{3}$. Let B denote an event which corresponds to hypothesis that we have chosen white ball. By the definition of the conditional probability, we have $P(B_j|A_1) = \frac{1}{2}; P(B_j|A_2) = \frac{3}{7}; P(B_j|A_2) = \frac{4}{5}$. Using the formula of total probability, we have

$$P(B) = \frac{1}{3} \cdot \frac{2}{5} + \frac{1}{3} \cdot \frac{3}{7} + \frac{1}{3} \cdot \frac{4}{5} = \frac{57}{105}.$$

Conclusion. The probability that accidentally chosen ball will be white in our experiment is equal to $\frac{57}{105}$.

Example 4.4

Experiment. The probability of formation of k bacteria $(k \in N)$ is equal to $\frac{\lambda^k}{k!} e^{-\lambda}$ $(\lambda > 0)$. The probability of adaptation with environment of the formed bacterium is equal to $p(0 < p < 1)$.

Problem. What is the probability that n bacteria $(n \in N)$ will pass the adaptation process ?

Solution of the problem. Let A_k be the event that k bacteria $(k \in N)$ pass adaptation process. Note that $(A_k)_{k \in N}$ is a complete system of representatives. Let B_n denotes the event that n baqteria pass the adaptation process $(n \in N)$. Note that $P(B_n|A_k) = 0$, when $k \leq n-1$. If $k \geq n$, then $P(B_n|A_k) = C_k^n p^n (1-p)^{k-n}$. We have

$$P(B_n) = \sum_{k \in N} P(A_k) P(B_n|A_k) = \sum_{k=n} \frac{\lambda^k}{k!} e^{-\lambda} C_k^n p^n (1-p)^{k-n} =$$

$$= \sum_{k=n} \frac{\lambda^k}{k!} e^{-\lambda} \frac{k!}{n! \cdot (k-n)!} p^n (1-p)^{k-n} =$$

$$= \frac{(p\lambda)^n}{n!} e^{-\lambda} \sum_{k=n} \frac{\lambda^{k-n}}{\cdot(k-n)!} (1-p)^{k-n} = \frac{(p\lambda)^n}{n!} e^{-\lambda} \sum_{k=n} \frac{(\lambda \cdot (1-p))^{k-n}}{\cdot(k-n)!} =$$

$$= \frac{(p\lambda)^n}{n!} e^{-\lambda} e^{\lambda \cdot (1-p)} = \frac{(p\lambda)^n}{n!} e^{-p \cdot \lambda}.$$

Conclusion. The probability that n bacteria pass the adaptation process $(n \in N)$ is equal to $\frac{(p\lambda)^n}{n!} e^{-p \cdot \lambda}$.

Theorem 4.6 *Assume that $J \subseteq N$ and $(A_j)_{j \in J}$ be a complete system of representatives. For every event B with $P(B) > 0$ we have*

$$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{\sum_{j \in J} P(A_j)P(B|A_j)} \ (i \in J),$$

which is called Bayes' [1] *formulas.*

Proof. Using formula of the total probability and the definition of the conditional probability, we have

$$P(A_i|B) = \frac{P(A_i \cap B)}{P(B)} = \frac{P(A_i)P(B|A_i)}{\sum_{j \in J} P(A_j)P(B|A_j)} \ (i \in J).$$

This ends the proof of theorem. □

Example 4.5 Suppose that we have chosen a white ball in the experiment considered in Example 4.3.

Problem. What is the probability that we have chosen white ball from the first urn?

Solution of the problem. By the Bayes' formula we have

$$P(A_1|B) = \frac{P(A_1) \cdot P(B|A_1)}{P(A_1) \cdot P(B|A_1) + P(A_2) \cdot P(B|A_2)} = \frac{1}{3} \cdot \frac{2}{5} : \frac{57}{105} = \frac{2}{9}.$$

Example 4.6 (A problem about ruin of the player). Let consider the game concerned with throwing of a coin, when playing heads or tails. If comes the side of the coin which has been chosen by the player, he wins 1 lari. In other case he loses the same amount of money. Assume that the initial capital of the player is x lari and he wishes to increase his capital to a lari $(x < a)$. The game is finished when the player is ruined or when he increases his capital to a lari.

Problem. What is the probability that the player will be ruined ?

[1] Bayes Thomas (1702, London-4.4.1761, Tanbridj)- English mathematician, the member of London Royal Society (1742). Main works in probability theory (1763).

Solution of the problem. Let $p(x)$ denote the probability of the ruin of the player when his initial capital is x lari. Then after one step in the case of winning the probability of the ruin will be $p(x+1)$, in other case same probability will be $p(x-1)$. Let B_1 denote the event, which corresponds to the case when the player wins in the first step. Analogously, denote by B_2 the event, which corresponds to the case when the player loses in the first step. We denote by A the event which corresponds to the ruin of the player. Then by the definition of the conditional probability, we have

$$P(A|B_1) = p(x+1),\ P(A|B_2) = p(x-1).$$

It is clear that (B_1, B_2) is a complete system of events. Since the coin is symmetrical, we have $P(B_1) = P(B_2) = \frac{1}{2}$. Using the formula of total probability, we have

$$p(x) = \frac{1}{2}[p(x+1) + p(x-1)].$$

Note that the following initial conditions $p(0) = 1$ and $p(a) = 0$ are fulfilled. One can easily check that the following linear function

$$p(x) = c_1 + c_2 x,$$

whose coefficients are defined by

$$p(0) = c_1 = 1,\ p(a) = c_1 + c_2 a = 0,$$

is a solution of the above mentioned equation. Finally, we get

$$p(x) = 1 - \frac{x}{a},\ 0 \le x \le a.$$

Conclusion. The probability that the player will be ruined in the above described game in the case when his initial capital is equal to x lari, is equal to

$$p(x) = 1 - \frac{x}{a},\ 0 \le x \le a.$$

Example 4.6 (The problem about division of game between hunters). The probability of shooting the game for the first hunter is equal to 0,8. The same probability for the second hunter is - 0,7. The beast was shot with simultaneous shots. The mass of the game was 190 kg. It was found that the game was killed with one bullet. How should the game be divided between hunters?

Solution of the problem. Let B denote the event that the game was killed by one hunter in the case of simultaneous shots. Let A_1 and A_2 denote events that the animal was killed by the first and the second hunters, respectively. Using Bayes' formulas we obtain

$$P(A_1|B) = \frac{0,3 \cdot 0,8}{0,3 \cdot 0,8 + 0,2 \cdot 0,7} = \frac{12}{19},$$

$$P(A_2|B) = \frac{0,2 \cdot 0,7}{0,3 \cdot 0,8 + 0,2 \cdot 0,7} = \frac{7}{19}.$$

It follows, that $P(A_1|B) \cdot 190 = 120$ (kg) of the game belongs to the first hunter and $P(A_2|B) \cdot 190 = 70$ (kg) of the game belongs to the second hunter, respectively.

Tests

4.1. Two shots shoot a target. The probability that the first shot will hit the shooting mark is equal to 0,9. Analogous probability for the second shot is 0,7. Then the probability that the target will be hit by both shots, is equal to

a) $0,42$, b) $0,63$, c) $0,54$, d) $0,36$.

4.2. The number of non-rainy days in June for Tbilisi is equal to 25. The probability that the first two days would be non-rainy is equal to

a) $\frac{5}{87}$, b) $\frac{20}{29}$, c) $\frac{19}{29}$, d) $\frac{18}{29}$.

4.3. We have accidently chosen two points A and B from set Δ which is defined by

$$\Delta = \{(x,y) : x \in [0,1],\ y \in [0,1]\}.$$

The functions g and f are defined by

$$g((x,y)) = \begin{cases} 1, & \text{if } x^2+y^2 \leq \frac{1}{4}, \\ 0, & \text{if } x^2+y^2 > \frac{1}{4}, \end{cases}$$

$$f((x,y)) = \begin{cases} 0, & \text{if } x+y \leq \frac{1}{2}, \\ 1, & \text{if } x+y > \frac{1}{2}, \end{cases}$$

Then the probability that $g(A)+f(B)=1$, is equal to

a) $\frac{7}{8} - \frac{3\pi}{64}$, b) $\frac{5}{8} - \frac{3\pi}{64}$, c) $\frac{7}{8} - \frac{\pi}{16}$, d) $1 - \frac{\pi}{8}$.

4.4. Here we have three boxes with the following compositions of balls

Box	Black balls	White balls
I	2	3
II	3	2
III	1	4

We accidently choose a box, from which accidentally choose also a ball.

1) The probability that the chosen ball is white, is equal to

a) $0,4$, b) $0,6$, c) $0,7$, d) $0,8$;

2) It is known that an accidently chosen ball is white. The probability that we have chosen a ball from box I, is equal to

a) $\frac{1}{3}$, b) $\frac{1}{4}$, c) $\frac{1}{5}$, d) $\frac{1}{6}$.

4.5. 100 and 200 details are produced in plants I and II, respectively. The probabilities of the producing of a standard detail in plants I and II are equal to 0,9 and 0,8, respectively.

1) A damage caused by the realization of non-standard details made up 3000 lari. A fine which must be payed by the administration of plant II caused realization of its non-standard details is equal to

a) 2400 lari, b) 2300 lari, c) 2000 lari, d) 1600 lari;

2) The prόfit received by the realization of standard details made up 5000 lari. A portion of the prόfit due to plant I is equal to

a) 1800, b) 1700, c) 1400, d) 3000 .

4.6. The player chooses a "heads" or "tails". If comes the side of the coin which was chosen by the player, then he wins 1 lari. In other case he loses the same money. Assume that the initial capital of the player is 1000 lari and he wishes to increase his capital to 2000 lari . The game is finished when the player is ruined or when the player will increase his capital to 2000 lari. What is the probability that the player will increase his capital to the until desired amount ?

a) $0,4$, b) $0,5$, c) $0,6$, d) $0,7$.

4.7. The probability of formation of k-bacteria $(k \in N)$ is equal to $\frac{0,3^k}{k!}e^{-0,3}$ $(k \in N)$. The probability of the adaptation with environment of the formed bacterium is equal to $0,1$.

1) The probability that exactly 5 bacteria will pass the adaptation process is equal to

a) $\frac{0,03^5}{5!}e^{-0,03}$, b) $\frac{0,04^5}{5!}e^{-0,04}$, c) $\frac{0,05^5}{5!}e^{-0,05}$, d) $\frac{0,06^5}{5!}e^{-0,06}$;

2) The observation of the accidentally chosen bacterium showed us that it has passed the adaptation process. The probability that this bacterium belongs to the adapted family consisting of 6 members, is equal to

a) $\frac{0,03^6}{6!} : (\sum_{k=1}^{\infty} \frac{0,03^k}{k \cdot k!} e^{-0,03})$, b) $\frac{0,03^6}{6 \cdot 6!} : (\sum_{k=1}^{\infty} \frac{0,03^k}{k \cdot k!} e^{-0,03})$.

Chapter 5

Applications of Caratheodory Method

5.1 Construction of Probability Spaces with Caratheodory method

Let Ω be a non-empty set and let F be any class of subsets of Ω.

Lemma 5.1.1 *There exists a σ-algebra $\sigma(F)$ of subsets of Ω, which contains the class F and is minimal in the sense of inclusion between such σ-algebras which contain F.*

Proof. Let $(\mathcal{F}_j)_{j\in J}$ denote a family of all σ-algebras of subsets of Ω which contain F and let define a class $\sigma(F)$ by the following formula

$$\sigma(F) = \cap_{j\in J}\mathcal{F}_j.$$

Let show that $\sigma(F)$ is a σ-algebra.

Indeed,

1) $\Omega \in \sigma(F)$, because $\Omega \in \mathcal{F}_j$ for $j \in J$.

2) Let $(A_k)_{k\in N}$ be any sequence of elements of $\sigma(F)$. Since this is a sequence of elements of $\mathcal{F}_j$ for arbitrary $j \in J$, we conclude that $\cap_{k\in N}A_k \in \mathcal{F}_j$ and $\cup_{k\in N}A_k \in \mathcal{F}_j$. The latter relation means that $\cap_{k\in N}A_k \in \cap_{j\in J}\mathcal{F}_j = \sigma(F)$ and $\cup_{k\in N}A_k \in \cap_{j\in J}\mathcal{F}_j = \sigma(F)$.

3) If $A \in \sigma(F)$, then $\overline{A} \in \mathcal{F}_j$ for arbitrary $j \in J$. The latter relation means that $\overline{A} \in \cap_{j\in J}\mathcal{F}_j = \sigma(F)$.

Now assume that $\sigma(F)$ is not minimal (in the sense of inclusion) between such σ-algebras which contain F. It means that there exists a σ-algebra $\mathcal{F}^*$, such that the following two conditions

1) $F \subset \mathcal{F}^*$,

2) $\mathcal{F}^* \subset \sigma(F)$ and $\sigma(F) \setminus \mathcal{F}^* \neq \emptyset$

are fulfilled.

By the definition of family $(\mathcal{F}_j)_{j\in J}$ there exists an index $j_0 \in J$ such that $\mathcal{F}_{j_0} = \mathcal{F}^*$. Hence, $\sigma(F) \subset \mathcal{F}^*$ which contradicts to 2). So we have obtained a contradiction and Lemma 5.1.1 is proved. □

Definition 5.1.1 Let S_1 and S_2 be two classes of subsets of Ω such that $S_1 \subset S_2$. Let P_1 and P_2 be two real-valued functions defined on S_1 and S_2, respectively. The function P_2 is called an extension of the function P_1 if

$$(\forall X)(X \in S_1 \to P_2(X) = P_1(X)).$$

Definition 5.1.2 Let $\mathcal{A}$ be an algebra of subsets Ω. A real-valued function P defined on $\mathcal{A}$ is called a probability if the following three conditions are fulfilled

1) $P(A) \geq 0$ for $A \in \mathcal{A}$;

2) $P(\Omega) = 1$;

3) If $(A_k)_{k\in N}$ $\mathcal{A}$ is a family of pairwise disjoint elements of $\mathcal{A}$ such that $\cup_{k\in N} A_k \in \mathcal{A}$, then

$$P(\cup_{k\in N} A_k) = \sum_{k\in N} P(A_k).$$

The general method of a construction of probability spaces is contained in the following Theorem.

Theorem 5.1.1 (Charatheodory [1]). *Let $\mathcal{A}$ be an algebra of subsets of Ω and P be a probability measure defined on $\mathcal{A}$. Then there exists a unique probability measure $\overline{P}$ on $\sigma(\mathcal{A})$ which is an extension of P. This extension is defined by the following formula:*

$$(\forall B)(B \in \sigma(\mathcal{A}) \to \overline{P}(B) = \inf\{ \sum_{k\in N} P(A_k) | (\forall k)(k \in N \to A_k \in \sigma(\mathcal{A}))$$

$$\& \ B \subseteq \cup_{k\in N} A_k \}.$$

Remark 5.1.1 The proof of Theorem 5.1.1 can be found in [6] .

Below we consider some applications of Theorem 5.1.1.

[1]Caratheodory Constantin (13.9.1873, Berlin2.2.1950, Munhen)-German mathematician. Professor of the Munhen University (1924-39), Lecturer of the Athena University(1933). Main works in theory of measures.

5.2 Construction of the Borel one-dimensional measure b_1 on $[0,1]$

Let $\mathcal{A}$ denote a class of subsets of $[0,1]$ elements of which can be presented as a union of a finite family of elements having one of the forms

$$[a_k,b_k[,\ [a_k,b_k],\ ]a_k,b_k[,\ ]a_k,b_k].$$

One can easily show that $\mathcal{A}$ is an algebra of subsets of $[0,1]$. We set

$$P([a_k,b_k[) = P([a_k,b_k]) = P(]a_k,b_k[) = P(]a_k,b_k]) = b_k - a_k.$$

In natural way, we can define P on elements of $\mathcal{A}$. It is not difficult to check that P is a probability measure on $\mathcal{A}$. Using Theorem 5.1.1 we deduce that there exists a unique extension $\overline{P}$ on $\sigma(\mathcal{A})$. A class $\sigma(\mathcal{A})$ is called a Borel σ-algebra of subsets of $[0,1]$ and is denoted by $\mathcal{B}([0,1])$. The probability $\overline{P}$ is called a one-dimensional classical Borel measure on $[0,1]$ and is denoted by b_1. The triplet $([0,1],\mathcal{B}([0,1]),b_1)$ is called a Borel classical probability space associated with $[0,1]$.

5.3 Construction of Borel probability measures on R

Let $F : R \to [0,1]$ be a continuous from the right function on R, which satisfies the following conditions:

$$\lim_{x\to-\infty} F(x) = 0 \ \&\ \lim_{x\to+\infty} F(x) = 1.$$

We suppose that $F(-\infty) = 0$ and $F(+\infty) = 1$.

We set $\Omega = R \cup \{+\infty\}$.

Let $\mathcal{A}$ denote a class of all subsets of Ω , which are represented by the union of finite number of semi-closed intervals of the form $(a,b]$, i.e.,

$$\mathcal{A} = \{A|A = \sum_{i=1}^{n}(a_i,b_i]\},$$

where $-\infty \le a_i < b_i \le \infty$ $(1 \le i \le n)$.

It is easy to show that $\mathcal{A}$ is an algebra of subsets of Ω.

We set

$$P(A) = P(\sum_{i=1}^{n}(a_i,b_i])) = \sum_{i=1}^{n} P((a_i,b_i]) = \sum_{i=1}^{n} F(b_i) - F(a_i).$$

One can easily demonstrate that the real-valued function P is a probability defined on $\mathcal{A}$. Using Theorem 5.1.1 we deduce that there exists a unique probability measure $\overline{P}$ on $\sigma(\mathcal{A})$ which is an extension of P. The class $\sigma(\mathcal{A})$ is called a Borel σ-algebra of subsets of the real axis $\mathbf{R}$ and is denoted by $\mathcal{B}(\mathbf{R})$. A real-valued function P_F , defined by

$$(\forall X)(X \in \mathcal{B}(\mathbf{R}) \to P_F(X) = \overline{P}(X)),$$

is called a probability Borel measure on $\mathbf{R}$ defined by the function F.

Example 5.3.1. Let F be defined by

$$(\forall x)(x \in R \to F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt\ ).$$

Let P_F be a Borel probability measure on R defined by F. Then the triplet $(\Omega, \mathcal{F}, P)$ is called an one-dimensional canonical(or standard) Gaussian probability space, associated with one-dimensional Euclidean vector space $\mathbf{R}(=\mathbf{R}^1)$.

The real-valued function P_F is called an one-dimensional canonical (or standard) Gaussian measure on $\mathbf{R}$ and is denoted by Γ_1.

5.4 The product of a finite family of probabilities

Let $(\Omega_i, \mathcal{F}_i, P_i)$ $(1 \le i \le n)$ be a finite family of probability spaces. We introduce some notions.

$$\prod_{i=1}^{n} \Omega_i = \{(\omega_1, \cdots, \omega_n) \mid \omega_1 \in \Omega_1, \cdots, \omega_n \in \Omega_n\}.$$

A set $A \subseteq \prod_{i=1}^{n} \Omega_i$ is called cylindrical if the following representation

$$B = \prod_{i=1}^{n} B_i,$$

is valid, where $B_i \in \mathcal{F}_i$ $(1 \le i \le n)$.

Let $\mathcal{A}$ be a class of all subsets of $\prod_{i=1}^{n} \Omega_i$ which are represented by the union of a finite number of pairwise disjoint cylindrical subsets. Note that $\mathcal{A}$ is an algebra of subsets of $\prod_{i=1}^{n} \Omega_i$. We set

$$P(\prod_{i=1}^{n} B_i) = \prod_{i=1}^{n} P_i(B_i)$$

and extend in the natural way a function P on class $\mathcal{A}$. Now one can easily demonstrate that function P is a probability measure defined on algebra $\mathcal{A}$. Using Charatheodory theorem there exists a unique probability measure $\overline{P}$ on $\sigma(\mathcal{A})$ which extends P. Class $\sigma(\mathcal{A})$ is called a product of the family of σ-algebras $(\mathcal{F}_i)_{1 \le i \le n}$ and is denoted by $\prod_{1 \le i \le n} \mathcal{F}_i$. The probability $\overline{P}$ is called a product of the family of probabilities $(P_i)_{1 \le i \le n}$ and is denoted by $\prod_{i=1}^{n} P_i$. The triplet $(\prod_{i=1}^{n} \Omega_i, \prod_{i=1}^{n} \mathcal{F}_i, \prod_{i=1}^{n} P_i)$ is called a product of the family of probability spaces $(\Omega_i, \mathcal{F}_i, P_i)_{1 \le i \le n}$.

Remark 5.4.1 Let consider a sequence of n independent random experiments. It is such a sequence of n random experiments when the result of any experiment does not influence on the result in the next experiment. Assume that i-th $(1 \le i \le n)$ experiment is described by a probability space $(\Omega_i, \mathcal{F}_i, P_i)$. Then a sequence of n independent random experiments is described by $(\prod_{i=1}^{n} \Omega_i, \prod_{i=1}^{n} \mathcal{F}_i, \prod_{i=1}^{n} P_i)$.

Here we consider some examples.

Example 5.4.1 (Bernoulli [2] probability measure).

We set

$$\Omega_i = \{0,1\},\ \mathcal{F}_i = \{A | A \subseteq \Omega_i\},\ P_i(\{1\}) = p$$

for i $(1 \leq i \leq n)$ and $0 < p < 1$. The product of probability spaces

$$(\prod_{i=1}^{n} \Omega_i, \prod_{i=1}^{n} \mathcal{F}_i, \prod_{i=1}^{n} P_i)$$

is called the Bernoulli n-dimensional classical probability space.

$\prod_{i=1}^{n} P_i$ is called the Bernoulli n-dimensional probability measure. If we consider a set A_k defined by

$$A_k = \{(\omega_1, \cdots, \omega_n) | (\omega_1, \cdots, \omega_n) \in \prod_{i=1}^{n} \Omega_i \ \&\ \sum_{i=1}^{n} \omega_i = k\},$$

then using the structure of the above-mentioned measure, we obtain

$$(\forall (\omega_1, \cdots, \omega_n))((\omega_1, \cdots, \omega_n) \in A_k \to \prod_{i=1}^{n} P_i((\omega_1, \cdots, \omega_n)) = p^k (1-p)^{n-k}).$$

Hence, $\prod_{i=1}^{n} P_i(A_k) = |A_k| p^k (1-p)^{n-k}$, where $|\cdot|$ denotes the cardinality of the corresponding set. It is easy to show that $|A_k| = C_n^k$, where C_n^k denotes the cardinality of all different subsets of cardinality k in the fixed set of cardinality n. The probability $(\prod_{i=1}^{n} P_i)(A_k)$ is denoted by $P_n(k)$, which means that during n-random two $\{0,1\}$-valued experiments the event $\{1\}$ had occurred k-times, if it is known that the probability of event $\{1\}$ in an arbitrary experiment is equal to p. If we denote by q the probability of event $\{0\}$, then we obtain the following formula

$$P_n(k) = C_n^k p^k q^{n-k}\ (1 \leq k \leq n),$$

which is called Bernoulli formula.

A natural number $k_0 \in [0, n]$ is called a number with hight probability if

$$P(k_0) = \max_{0 \leq k \leq n} P_n(k).$$

The number k_0 with hight probability is calculated by the following formula

$$k_0 = \begin{cases} [(1+n)p], & \text{if } (1+n)p \notin Z; \\ (1+n)p \text{ or } (1+n)p - 1, & \text{if } (1+n)p \in Z, \end{cases}$$

[2] Jacob Bernoulli (27.12.1654-16.8.1705)-Swedish mathematician, professor of Bazel University (1687). Him belongs the first proof of the so called Bernoulli theorem(which is a partial case of the Law of Large numbers) (cf. Arsconjectandi(Basileqe)(1713).

where $[\cdot]$ denotes an integer part of the corresponding number.

Example 5.4.2 (n-dimensional multinomial probability measure). Let triplet $(\Omega_i, \mathcal{F}_i, P_i)_{1\le i\le n}$ be defined by :

a) $\Omega_i = \{x_1, \cdots, x_k\}$ $(1 \le i \le n)$,

b) $\mathcal{F}_i$ is a powerset of Ω_i for $(1 \le i \le n)$,

c) $P_i(\{x_j\}) = p_j > 0,\ 1 \le i \le n,\ 1 \le j \le k,\ \sum_{j=1}^k p_j = 1$.

Then triplet $(\prod_{i=1}^n \Omega_i, \prod_{i=1}^n \mathcal{F}_i, \prod_{i=1}^n P_i)$ is called n-dimensional multinomial probability space.

A real-valued function $\prod_{i=1}^n P_i$ is called n-dimensional multinomial probability.

If we consider a set $A_n(n_1, \cdots, n_k)$, defined by

$$A_n(n_1, \cdots, n_k) = \{(\omega_1, \cdots, \omega_n) | (\omega_1, \cdots, \omega_n) \in \prod_{i=1}^n \Omega_i \ \&$$

$$|\{i : \omega_i = x_p\}| = n_p,\ 1 \le p \le k\},$$

then using the structure of the product measure, we obtain

$$(\forall(\omega_1, \cdots, \omega_n))((\omega_1, \cdots, \omega_n) \in A_n(n_1, \cdots, n_k) \to$$

$$\prod_{i=1}^n P_i((\omega_1, \cdots, \omega_n)) = p_1^{n_1} \times \cdots \times p_k^{n_k}).$$

Hence, $\prod_{i=1}^n P_i(A_n(n_1, \cdots, n_k)) = |A_n(n_1, \cdots, n_k)| \times p_1^{n_1} \times \cdots \times p_k^{n_k}$, where $|\cdot|$ denotes the cardinality of the corresponding set. It is not difficult to prove that $|A_n(n_1, \cdots, n_k)| = \frac{n!}{n_1! \times \cdots \times n_k!}$.

Then probability $\prod_{i=1}^n P_i(A_n(n_1, \cdots, n_k))$ (denoted by $P_n(n_1, \cdots, n_k)$) assumes that during n-random $\{x_1, \cdots, x_k\}$-valued experiments the event x_1 will occur n_1-times, $\cdots$, the event x_k will occur n_k-times if it is known that in i-th experiment the probability that the event x_i occurred is equal to $p_i (1 \le i \le k)$, is calculated by the following formula

$$P_n(n_1, \cdots, n_k) = \frac{n!}{n_1! \times \cdots \times n_k!} \times p_1^{n_1} \times \cdots \times p_k^{n_k},$$

This formula is called the formula for calculation of n-dimensional multinomial probability.

Remark 5.4.1 Note that the class of n-dimensional multinomial probability measures consists the class of n-dimensional Bernoulli probability measures. In particular, when $k = 2$, the n-dimensional multinomial probability measure stands n-dimensional Bernoulli probability measure.

Example 5.4.3 (n-dimensional Borel classical measures on $[0,1]^n$ and R^n). Assume that a family of probability spaces $(\Omega_i, \mathcal{F}_i, P_i)_{1\le i\le n}$ is defined by:

a) $\Omega_i = [0,1]$ $(1 \le i \le n)$,

b) $\mathcal{F}_i = \mathcal{B}([0,1])$ $(1 \le i \le n)$,

c) $P_i = b_1,\ (1\ \leq i \leq n)$.

Then triplet $(\prod_{i=1}^n \Omega_i, \prod_{i=1}^n \mathcal{F}_i, \prod_{i=1}^n P_i)$ is called n-dimensional Borel probability space associated with n-dimensional cube $[0,1]^n$. The real-valued function $\prod_{i=1}^n P_i$ is called n-dimensional classical Borel measure defined on $[0,1]^n$. The real-valued function b_n, defined by

$$(\forall X)(X \in B(R^n) \rightarrow b_n(X) = \sum_{g \in Z^n} \prod_{i=1}^{n} P_i([0,1[^n \cap g(X)),$$

is called n-dimensional classical Borel measure defined on R^n.

Example 5.4.4 Assume that a family of functions $(F_i)_{1 \leq i \leq n}$ is defined by

$$(\forall i)(\forall x)(1 \leq i \leq n \ \&\ x \in R \rightarrow F_i(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt\ ).$$

Assume also that P_i denotes the probability measure on R defined by F_i. Then the probability space $((\prod_{1 \leq i \leq n} \Omega_i, \prod_{i \in N} \mathcal{F}_i, \prod_{i \in N} P_i)$ is called an n-dimensional canonical (or standard) Gaussian probability space associated with R^n. The real-valued function $\prod_{1 \leq i \leq n} P_i$ is called an n-dimensional canonical (or standard) Gaussian probability measure on R^n and is denoted by Γ_n.

5.5 Definition of the Product of the Infinite Family of Probabilities

Let $(\Omega_i, \mathcal{F}_i, P_i)_{i \in I}$ be an infinite family of probability spaces. We set

$$\prod_{i \in I} \Omega_i = \{(\omega_i)_{i \in I}\ :\ \omega_i \in \Omega_i, i \in I\}.$$

A subset $A \subseteq \prod_{i \in I} \Omega_i$ is called a cylindrical set, if there exists a finite number of indices $(i_k)_{1 \leq k \leq n}$ and such elements $B_{i_k}(1 \leq k \leq n)$ of σ-algebras $\mathcal{F}_{i_k}(1 \leq k \leq n)$ that

$$B = \{(\omega_i)_{i \in I}\ :\ (\omega_i \in \Omega_i, i \in I \setminus \cup_{k=1}^n \{i_k\}\ ) \ \&\ (\omega_i \in B_i, i \in \cup_{k=1}^n \{i_k\})\ \}.$$

Let $\mathcal{A}$ denote a class of such subsets of $\prod_{i \in I} \Omega_i$ which are presented by the union of finite number of pairwise disjoint cylindrical subsets. Note that class $\mathcal{A}$ is an algebra of subsets of $\prod_{i \in I} \Omega_i$. Define a real-valued function P on the cylindrical subset B by the following formula

$$P(B) = \prod_{k=1}^{n} P_{i_k}(B_{i_k})$$

and extend in natural way a functional P on class $\mathcal{A}$. Clearly, a real-valued function P is the probability defined on an algebra $\mathcal{A}$. Using Charatheodory theorem we deduce an existence of the unique extended probability measure $\overline{P}$ on class $\sigma(\mathcal{A})$. The class of subsets $\sigma(\mathcal{A})$ is called the product of the infinite family of σ-algebras $(\mathcal{F}_i)_{i \in I}$ and is denoted by $\prod_{i \in I} \mathcal{F}_i$.

The real-valued function $\overline{P}$ is called the product of the infinite family of probabilities $(P_i)_{i\in I}$ and is denoted by $\prod_{i\in I} P_i$. A triplet $(\prod_{i\in I}\Omega_i, \prod_{i\in I}\mathcal{F}_i, \prod_{i\in I} P_i)$ is called the product of the infinite family of probability spaces $(\Omega_i, \mathcal{F}_i, P_i)_{i\in I}$.

Remark 5.5.1 An infinite sequence of independent experiments is such a sequence of experiments when the result of each experiment does not influence on the result of any next experiment. Assume that i-th $(i \in I)$ random experiment is described by the probability space $(\Omega_i, \mathcal{F}_i, P_i)$. Then an infinite sequence of independent experiments is described by the triplet

$$(\prod_{i\in I}\Omega_i, \prod_{i\in I}\mathcal{F}_i, \prod_{i\in I} P_i).$$

Let consider some examples.

Example 5.5.1 For $i \in N$ we set

$$\Omega_i = \{0,1\},\ \mathcal{F}_i = \{A|A \subseteq \Omega_i\},\ P_i(\{1\}) = p,$$

where $0 < p < 1$. The product of the infinite family of probability spaces

$$(\prod_{i\in N}\Omega_i, \prod_{i\in N}\mathcal{F}_i, \prod_{i\in N} P_i)$$

is called the infinite-dimensional Bernoulli classical probability space. A real-valued function $\prod_{i\in N} P_i$ is called the infinite-dimensional Bernoulli classical probability .

Example 5.5.2 (Infinite-dimensional multinomial probability space). Assume that an infinite family of probability spaces $(\Omega_i, \mathcal{F}_i, P_i)_{i\in N}$ is defined by :

a) $\Omega_i = \{x_1, \cdots, x_k\}$ $(i \in N)$,

b) $\mathcal{F}_i$ is the powerset of Ω_i for arbitrary $i \in N$,

c)$P_i(\{x_j\}) = p_j > 0,\ i \in N,\ 1 \le j \le k,\ \sum_{j=1}^{k} p_j = 1.$

Then $(\prod_{i\in N}\Omega_i, \prod_{i\in N}\mathcal{F}_i, \prod_{i\in N} P_i)$ is called the infinite-dimensional multinomial probability space. The real-valued function $\prod_{i\in N} P_i$ is called the infinite-dimensional multinomial probability.

Example 5.5.3 (Infinite-dimensional Borel classical probability measure on infinite-dimensional cube $[0,1]^N$) Let consider an infinite family of probability spaces $(\Omega_i, \mathcal{F}_i, P_i)_{i\in N}$ defined by:

a) $\Omega_i = [0,1]$ $(i \in N)$,

b) $\mathcal{F}_i = \mathcal{B}([0,1])$ $(i \in N)$,

c) $P_i = b_1$ $(i \in N)$.

Then the triplet

$$(\prod_{i\in N}\Omega_i, \prod_{i\in N}\mathcal{F}_i, \prod_{i\in N} P_i)$$

is called the infinite-dimensional Borel classical probability measure associated with infinite-dimensional cube $[0,1]^N$. Measure $\prod_{i\in N} P_i$ is called the infinite-dimensional Borel classical probability measure on infinite-dimensional cube $[0,1]^N$ and is denoted by b_N.

Example 5.5.4 Assume that an infinite family of functions $(F_i)_{i\in N}$ is defined by

$$(\forall i)(\forall x)(i \in N \ \&\ x \in R \to F_i(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt\ ).$$

Let P_i be a Borel probability measure on R defined by F_i. Then triplet $(\prod_{i\in N} \Omega_i, \prod_{i\in N} \mathcal{F}_i, \prod_{i\in N} P_i)$ is called the infinite-dimensional Gaussian canonical probability space associated with R^N. The real-valued function $\prod_{i\in N} P_i$ is called the infinite-dimensional Gaussian canonical probability measure defined on RN and is denoted by Γ_N.

Tests

5.1. There are 10000 trade booths on the territory of the market. The probability that each owner of a booth will get profit 500 lari during one quarter is equal to 0; 5. Further, the probability that the same owner of the booth loose 200 lari during the same quarter is equal to 0; 5. The number of such owners of trade booths , which at the end of the year

1) will loose 800 lari, is equal to

a) 625, b) 670, c) 450, d) 700;

2) will loose 100 lari, is equal to

a) 2500, b) 3000, c) 2000, d) 3500;

3)will get a profit of 600 lari, is equal to

a) 3750, b) 3650, c) 3600, d) 3400;

4)will get a profit of 1300 lari, is equal to

a) 2500, b) 2000, c) 3000, d) 1500;

5)will get a profit of 2000 lari, is equal to

a) 625, b) 650, c) 600, d) 550.

5.2. Wholesale storehouse supplies 20 magazines. It is possible to get order for the next day from each magazine with probability $0,5$.

1) The number of hight probability of orders at the end of the day is equal to

a) 10, b) 11, c) 12, d) 13;

2) The probability corresponding with the number of hight probability of orders at the end of the day is equal to

a) $C_{20}^{10}\frac{1}{2^{20}}$, b) $C_{20}^{10}\frac{1}{2^{10}}$, c) $C_{20}^{10}\frac{1}{2^{30}}$, d) $C_{20}^{5}\frac{1}{2^{20}}$.

5.3. There are three boxes numerated by numbers 1, 2, 3. The probabilities, that a particle will be placed in the box 1, 2, 3 are equal to $0.3, 0.4$ and 0.3, respectively. The probability that out of 6 particles 3 will be placed in box 1, 2 particles will be placed in box 2 and one particle will be find in box 3, is equal to

a) $\frac{3!}{3!2!1!}0,3^4 0,4^2$,

b) $\frac{4!}{3!2!1!}0, 3^4 0, 4^2$,
c) $\frac{5!}{3!2!1!}0, 3^4 0, 4^2$,
d) $\frac{6!}{3!2!1!}0, 3^4 0, 4^2$.

5.4. Let $\Omega \subset R^m$ be a Borel subset such that $0 < b_m(\Omega) < +\infty$. Suppose that the probability that a point will be choice from any Borel subset of $A \subset \Omega$ is proportional to its Borel b_m-measure. Let $(A_i)_{1 \le i \le n}$ be a complete system of representatives. If we accidentally choose n points from the region Ω, then the probability that n_i points will be chosen in region A_i for $1 \le i \le n$, can be calculated by
a) $\frac{n!}{n_1! \times \cdots \times n_k! b_m(\Omega)^n} \times b_m(A_1)^{n_1} \times \cdots \times b_m(A_k)^{n_k}$,
b) $\frac{m!}{n_1! \times \cdots \times n_k! b_m(\Omega)^m} \times b_m(A_1)^{n_1} \times \cdots \times b_m(A_k)^{n_k}$,

5.5. We accidentally choose n point from a square with inscribed circle. The probability that between randomly chosen n points exactly k points will be within the circle, is equal to
a) $\frac{n!}{k!(n-k)!}(1-\frac{\pi}{4})^k(\frac{\pi}{4})^{n-k}$
b) $\frac{n!}{k!(n-k)!}(1-\frac{\pi}{4})^{n-k}(\frac{\pi}{4})^k$

5.6. We accidentally choose n point from a cube, in which is inscribed a ball. The probability that between n chosen points k points belong to the ball, is equal to
a) $\frac{n!}{k!(n-k)!}(1-\frac{\pi}{6})^k(\frac{\pi}{6})^{n-k}$
b) $\frac{n!}{k!(n-k)!}(1-\frac{\pi}{6})^{n-k}(\frac{\pi}{6})^k$

5.7. We accidentally choose n points in a square Δ, which is defined by

$$\Delta = \{(x, y) : 0 \le x \le 1,\ 0 \le y \le 1\}.$$

The probability that coordinates (x, y) of k points satisfy the following condition

$$x + y \ge \frac{1}{2},$$

is equal to
a) $\frac{n!}{k!(n-k)!}(\frac{1}{8})^k(\frac{7}{8})^{n-k}$
b) $\frac{n!}{k!(n-k)!}(\frac{1}{8})^{n-k}(\frac{7}{8})^k$

5.8. There are passed parallel lines on the plane such that the distant between neighbouring lines is equal to $2a$. We accidentally throw $l(2l < 2a)$ long needle on the plane n-times. The probability that the needle k-times $(0 \le k \le n)$ intersects any of the above-mentioned parallel line, is equal to
a) $\frac{n!}{k!(n-k)!}(\frac{2l}{a\pi})^k(\frac{a\pi-2l}{a\pi})^{n-k}$
b) $\frac{n!}{k!(n-k)!}(\frac{2l}{a\pi})^{n-k}(\frac{a\pi-2l}{a\pi})^k$

Chapter 6

Random Variables

Let $(\Omega, \mathcal{F}, P)$ be a probability space.

Definition 6.1 Function $\xi : \Omega \to R$ called a random variable, if

$$(\forall x)(x \in R \to \{\omega : \omega \in \Omega,\ \xi(\omega) < x\} \in \mathcal{F}).$$

Example 6.1 Arbitrary random variable $\xi : \Omega \to R$ can be considered as a definite rule of dispersion of the unit mass of powder Ω on the real axis R, according to which each particle $\omega \in \Omega$ will be placed on particle $M \in R$ with coordinate $\xi(\omega)$.

Definition 6.2 Function $I_A : \Omega \to R$ $(A \subset \Omega)$, defined by

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

is called the indicator of A.

Theorem 6.1 *Let $A \subset \Omega$. Then I_A is a random variable if and only if $A \in \mathcal{F}$.*

Proof. The validity of Theorem 6.1 follows from the following formula

$$\{\omega : I_A(\omega) < x\} = \begin{cases} \emptyset, & \text{if } x \le 0, \\ \overline{A}, & \text{if } 0 < x \le 1, \\ \Omega, & \text{if } 1 < x. \end{cases}$$

□

Definition 6.3 Random variable $\xi : \Omega \to R$ is called a discrete random variable, if there exists a sequence of pairwise disjoint events $(A_k)_{k \in N}$ and a sequence of real numbers $(x_k)_{k \in N}$, such that:

1) $(\forall k)(k \in N \to x_k \in R, A_k \in \mathcal{F})$,
2) $\cup_{k \in N} A_k = \Omega$,
3) $\xi(\omega) = \sum_{k \in N} x_k I_{A_k}(\omega),\ \omega \in \Omega$.

Definition 6.4 Random variable $\xi : \Omega \to R$ is called a simple discrete random variable, if there exists a finite sequence of pairwise disjoint events $(A_k)_{1 \le k \le n}$ and a finite sequence of real numbers $(x_k)_{1 \le k \le n}$, such that:
1) $(\forall k)(1 \le k \le n \to x_k \in R, A_k \in \mathcal{F}))$;
2) $\cup_{k=1}^{n} A_k = \Omega$;
3) $\xi(\omega) = \sum_{k=1}^{n} x_k I_{A_k}(\omega),\ \omega \in \Omega$.

Definition 6.5 A sequence of random variables $(\xi_k)_{k \in N}$ is called increasing if

$$(\forall n)(\forall \omega)(n \in N, \omega \in \Omega \to \xi_n(\omega) \le \xi_{n+1}(\omega)).$$

The following theorem gives an interesting information about the structure of non-negative random variables

Theorem 6.2 *For arbitrary non-negative random variable* $\xi : \Omega \to R$ *there exists an increasing sequence of simple discrete variables* $(\xi_k)_{k \in N}$ *such that*

$$(\forall \omega)(\omega \in \Omega \to \xi(\omega) = \lim_{n \to \infty} \xi_n(\omega)).$$

Proof. For arbitrary $n \in N$ we define a simple discrete variable n by the following formula

$$\xi_n(\omega) = \sum_{k=1}^{n \cdot 2^n} \frac{k-1}{2^n} \cdot I_{\{y: y \in \Omega, \frac{k-1}{2^n} \le \xi(y) < \frac{k}{2^n}\}}(\omega) + n \cdot I_{\{y: y \in \Omega, \xi(y) \ge n\}}(\omega).$$

Clearly,

$$(\forall n)(n \in N \to \ \xi_n(\omega) \le \xi_{n+1}(\omega))$$

and

$$(\forall \omega)(\omega \in \Omega \to \xi(\omega) = \lim_{n \to \infty} \xi_n(\omega)).$$

This ends the proof of theorem. □

Theorem 6.3 *For arbitrary random variable* $\eta : \Omega \to R$ *there exists a sequence of simple discrete variables* $(\eta_k)_{k \in N}$, *such that*

$$(\forall \omega)(\omega \in \Omega \to \eta(\omega) = \lim_{n \to \infty} \eta_n(\omega)).$$

Proof. For arbitrary random variable $\eta : \Omega \to R$ we have the following representation $\eta = \eta^+ + \eta^-$, where $\eta^+(\omega) = \max\{\eta(\omega), 0\}$ and $\eta^-(\omega) = \min\{\eta(\omega), 0\}$. Using Theorem 6.2, for η^+ and $-\eta^-$ there exist increasing sequences of simple discrete variables $(\eta_k^+)_{k\in N}$ and $(\eta_k^-)_{k\in N}$, such that

$$(\forall\omega)(\omega \in \Omega \to \lim_{k\to\infty} \eta_k^+(\omega) = \eta^+(\omega),\ \lim_{k\to\infty} \eta_k^-(\omega) = -\eta^-(\omega)).$$

It is easy to show that $(\eta_n)_{n\in N} = (\eta_n^+ - \eta_n^-)_{n\in N}$ is a sequence of simple discrete random variables such that

$$(\forall\omega)(\omega \in \Omega \to \eta(\omega) = \lim_{n\to\infty} \eta_n(\omega)).$$

This ends the proof of theorem. □

Tests

6.1. Let ξ and η be discrete random variables for which the following representations

$$\xi(\omega) = \sum_{k\in N} x_k I_{A_k}(\omega),\ \ \eta(\omega) = \sum_{m\in N} y_m I_{B_m}(\omega)\ (\ \omega \in \Omega)$$

are valid. Then

1) for random variable $\xi + \eta$ we have

$$\text{a)}\ (\xi+\eta)(\omega) = \sum_{k\in N}\sum_{m\in N}(x_k + y_m) I_{A_k \cap B_m}(\omega),$$

$$\text{b)}\ (\xi+\eta)(\omega) = \sum_{k\in N}\sum_{m\in N} x_k y_m I_{A_k \cap B_m}(\omega);$$

2) for random variable $\xi \cdot \eta$ we have

$$\text{a)}\ (\xi\cdot\eta)(\omega) = \sum_{k\in N}\sum_{m\in N}(x_k + y_m) I_{A_k \cap B_m}(\omega),$$

$$\text{b)}\ (\xi\cdot\eta)(\omega) = \sum_{k\in N}\sum_{m\in N} x_k y_m I_{A_k \cap B_m}(\omega);$$

3) if $g : R \to R$ is a measurable function, then

$$\text{a)}\ g(\xi)(\omega) = \sum_{k\in N} g(x_k) I_{A_k}(\omega),$$

$$\text{b)}\ g(\xi)(\omega) = \sum_{k\in N} g^{-1}(x_k) I_{A_k}(\omega);$$

4) the following formula is valid

$$\text{a)}\ \sin(\xi)(\omega) = \sum_{k\in N} \sin(x_k) I_{A_k}(\omega),$$

$$\text{b)}\ \sin(\xi)(\omega) = \sum_{k\in N} \arcsin(x_k) I_{A_k}(\omega).$$

6.2. Let $(A_k)_{k\in N}$ be a sequence of events and let ξ be a random variable. Then

1)

$$\text{a)}\ \xi^{-1}(\cup_{k\in N} A_k) = \cup_{k\in N} \xi^{-1}(A_k),$$

$$\text{b)}\ \xi^{-1}(\cup_{k\in N} A_k) = \cap_{k\in N} \xi^{-1}(A_k);$$

2)

$$a)\ \xi^{-1}(\cap_{k\in N}A_k) = \cap_{k\in N}\xi^{-1}(A_k),$$

$$b)\ \xi^{-1}(\cap_{k\in N}A_k) = \cup_{k\in N}\xi^{-1}(A_k);$$

3)

$$a)\ \Omega\setminus\xi^{-1}(A_k) = \xi^{-1}(\Omega\setminus A_k),$$

$$b)\ \Omega\setminus\xi^{-1}(A_k) = \xi^{-1}(A_k).$$

6.3.

1) If $|\xi|$ is a random variable, then

a) ξ is a random variable,

b) it is possible that ξ is not a random variable;

2) if ξ is a random variable, then

a) ξ^+ is a random variable;,

b) it is possible that ξ^+ is not a random variable;

3) Let ξ and η be random variables and let A be any event. If $\Theta(\omega) = \xi(\omega)I_A(\omega) + \eta(\omega)I_{\overline{A}}(\omega)$ $(\omega\in\Omega)$, then

a) Θ is a random variable,

b) It is possible that Θ is not a random variable.

Chapter 7

Random variable distribution function

Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $\xi : \Omega \to R$ be a random variable

Definition 7.1 Function $\xi : \Omega \to R$, defined by

$$(\forall x)(x \in \overline{R} \to F_{\xi}(x) = P(\{\omega : \xi(\omega) \leq x\}),$$

where $\overline{R} = \{-\infty\} \cup R \cup \{+\infty\}$, is called a distribution function of random variable ξ.

Here we consider some properties of distribution functions.

Theorem 7.1 $F_{\xi}(+\infty) = \lim_{x \to +\infty} F_{\xi}(x) = 1$.

Proof. Let consider an increasing sequence of real numbers $(x_k)_{k \in N}$ such that $\lim_{k \to +\infty} x_k = +\infty$. On the one hand, we have

$$\{\omega : \xi(\omega) \leq x_k\} \subseteq \{\omega : \xi(\omega) \leq x_{k+1}\} \ (k \in N).$$

On the other hand, we have $\cup_{k \in N}\{\omega : \xi(\omega) \leq x_k\} = \Omega$. Using the property of continuity from below we get

$$\lim_{k \to \infty} P(\{\omega : \xi(\omega) \leq x_k\}) = P(\cup_{k \in N}\{\omega : \xi(\omega) \leq x_k\}) = P(\Omega) = 1,$$

i.e. $\lim_{x \to +\infty} F_{\xi}(x) = 1$. Note that $F_{\xi}(+\infty) = P(\{\omega : \xi(\omega) \leq +\infty\}) = P(\Omega) = 1$. Finally we get

$$F_{\xi}(+\infty) = \lim_{x \to +\infty} F_{\xi}(x) = 1.$$

This ends the proof of theorem. □

Theorem 7.2 $F_{\xi}(-\infty) = \lim_{x \to -\infty} F_{\xi}(x) = 0$.

Proof. Note that

$$F_\xi(-\infty) = P(\{\omega : \xi(\omega) \le -\infty\}) = P(\emptyset) = 0.$$

Let consider a decreasing sequence of real numbers $(x_k)_{k\in N}$ such that $\lim_{k\to+\infty} x_k = -\infty$. It is easy to check the validity of the following conditions:

1) $\{\omega : \xi(\omega) \le x_{k+1}\} \subseteq \{\omega : \xi(\omega) \le x_k\}$ $(k \in N)$,
2) $\cap_{k\in N} \{\omega : \xi(\omega) \le x_k\} = \emptyset$.

Using the property of the continuity from above of P, we get

$$\lim_{k\to\infty} P(\{\omega : \xi(\omega) \le x_k\}) = P(\cap_{k\in N}\{\omega : \xi(\omega) \le x_k\}) = P(\emptyset) = 0,$$

i.e.,

$$\lim_{x\to-\infty} F_\xi(x) = F_\xi(-\infty) = 0.$$

This ends the proof of theorem. □

Theorem 7.3 *Distribution function F(x) is an increasing function.*

Proof. Let, $x_1 < x_2$. Let show the validity of the following non-strict inequality $F_\xi(x_1) \le F_\xi(x_2)$. Indeed, using the validity of the following inclusion

$$\{\omega : \xi(\omega) \le x_1\} \subseteq \{\omega : \xi(\omega) \le x_2\}$$

and Property 2.5 (cf. Chapter 2), we have

$$P(\{\omega : \xi(\omega) \le x_1\}) \le P(\{\omega : \xi(\omega) \le x_2\}),$$

which is equivalent to condition $F_\xi(x_1) \le F_\xi(x_2)$. This ends the proof of theorem. □

Theorem 7.4 *Distribution function $F_\xi(x)$ is continuous from the right , i. e., for arbitrary sequence of real numbers $(x_k)_{k\in N}$ for which $x_k > x$ $(k \in N)$ and $\lim_{k\to\infty} x_k = x$, the following condition*

$$\lim_{k\to\infty} F_\xi(x_k) = F_\xi(x).$$

is fulfilled.

Proof. Without loss of generality, we can assume that $(x_k)_{k\in N}$ is a decreasing sequence. Then

$$\{\omega : \xi(\omega) \le x\} = \cap_{k\in N}\{\omega : \xi(\omega) \le x_k\},$$

$$\{\omega : \xi(\omega) \le x_{k+1}\} \subseteq \{\omega : \xi(\omega) \le x_k\} \ (k \in N).$$

Hence, using the property of the continuity from above of P, we obtain

$$\lim_{k\to\infty} P(\{\omega : \xi(\omega) \le x_k\}) = P(\cap_{k\in N}\{\omega : \xi(\omega) \le x_k\}) = P(\{\omega : \xi(\omega) \le x\}),$$

which is equivalent to condition $\lim_{k\to\infty} F_\xi(x_k) = F_\xi(x)$. This ends the proof of theorem. □

Let ξ be a discrete random variable, i.e., there exist an infinite family of pairwise disjoint events $(A_k)_{k\in N}$ and an infinite family of real numbers $(x_k)_{k\in N}$, such that:

1) $(\forall k)(k \in N \to x_k \in R,\ A_k \in \mathcal{F})$,

2) $\cup_{k\in N} A_k = \Omega$,

3) $\xi(\omega) = \sum_{k\in N} x_k I_{A_k}(\omega),\ \omega \in \Omega$.

Then the distribution function of ξ is calculated by

$$F_\xi(x) = \sum_{x_k \le x} P(A_k).$$

Remark 7.1 Assume that in column A of the Excel table we have entered values $x_1, \cdots, x_n$ of simple discrete random variable ξ. Assume also that in column B of Excel table we have entered the corresponding probabilities $p_1, \cdots, p_n$. Then the statistical function PROB$(x_1 : x_n; p_1 : p_n; y_1; y_2)$ calculates the following probability $P(\{\omega : \omega \in \Omega \ \&\ y_1 \le \xi \le y_2\})$.

Let consider some examples.

Example 7.1 (Poisson [1] distribution). We say that a discrete random variable $\xi : \Omega \to R$ defined by

$$\xi(\omega) = \sum_{n\in N} n \cdot I_{A_n}(\omega)\ (\omega \in \Omega)$$

generates a Poisson distribution with parameter $\lambda\ (\lambda > 0)$ if the following condition

$$P(A_n) = \frac{\lambda^n}{n!} e^{-\lambda}\ (n \in N),$$

i.e.,

$$P(\{\omega : \xi(\omega) = n\}) = \frac{\lambda^n}{n!} e^{-\lambda}\ (n \in N).$$

Poisson distribution function $F(x,\lambda)$ with parameter λ is defined by the following formula

$$F(x,\lambda) = \sum_{n \le x} \frac{\lambda^n}{n!} e^{-\lambda}\ (x \in R).$$

Remark 7.2 POISSON$(k;\lambda;0)$ calculates the probability that the Poisson random variable with parameter λ will get value k. For example, POISSON$(0;0;2;0) = 0,818730753$.

Remark 7.3 POISSON$(k;\lambda;1)$ calculates the probability that the Poisson random variable with parameter will get an integer value in interval $[0,k]$. For example, POISSON$(2;0;2;1) = 0,998851519$.

[1]Poisson; Semion Denis (21.6.1781 - 25.4.1840)-French mathematician, physician, the member of Paris Academy of Sciences (1812), the honourable member of Petersburg Academy of Sciences (1826).

Example 7.2 (The geometric distribution). We say that a discrete random variable $\xi : \Omega \to R$, defined by

$$\xi(\omega) = \sum_{n \in N} n \cdot I_{A_n}(\omega) \ (\omega \in \Omega)$$

generates the geometric distribution with parameter q $(0 \le q \le 1)$ if the condition

$$P(A_n) = (1-q)q^{n-1} \ (n \in N),$$

i. e.,

$$P(\{\omega : \xi(\omega) = n\}) = (1-q)q^{n-1} \ (n \in N).$$

The geometric distribution F_q with parameter q is defined by the following formula

$$F_q(x) = \sum_{n \le x} (1-q)q^{n-1} \ (x \in R).$$

Example 7.3 (Leibniz [2] distribution). We say that a discrete random variable $\xi : \Omega \to R$, defined by

$$\xi(\omega) = \sum_{n=1}^{\infty} n \cdot I_{A_n}(\omega) \ (\omega \in \Omega)$$

generates the Leibniz distribution if the condition

$$P(A_n) = \frac{1}{n \cdot (n+1)} \ (n \in N),$$

holds, i.e.

$$P(\{\omega : \xi(\omega) = n\}) = \frac{1}{n \cdot (n+1)} \ (n \in N).$$

The Leibniz distribution function is calculated by the following formula

$$F(x) = \sum_{n \le x} \frac{1}{n \cdot (n+1)} =$$

$$= \begin{cases} 0, & \text{if } x < 1 \\ 1 - \frac{1}{[x]+1}, & \text{if } x \ge 1 \end{cases},$$

where $[x]$ denotes an integer part of x.

Example 7.4 (Hypergeometric distribution). A simple discrete random variable

$$\xi(\omega) = \sum_{k=1}^{n} k I_{A_k}(\omega) \ \ (\omega \in \Omega).$$

[2]Leibniz; Gottfried Wilhelm (1.7.1646, -14.11.1716)-German mathematician, the member of London Royal Society (1673), the member of Paris Academy of Sciences (1700)

is called distributed by hypergeometric law with parameters (n,a,A) if

$$P(A_k) = \frac{C_a^k C_{A-a}^{m-k}}{C_A^k} \; (k = 0,1,\cdots,n)$$

where $0 \le n \le \min\{a, A-a\}$.

The hypergeometric distribution with parameters $(n;a;A)$ is denoted by $F_{((n;a;A)}$ and is defined by

$$F_{(n;a;A)} = \sum_{k \le x} \frac{C_a^k C_{A-a}^{m-k}}{C_A^k}.$$

Remark 7.4 HYPERGEOMDIST$(k;n;a;A)$ calculates the value $\frac{C_a^k C_{A-a}^{m-k}}{C_A^k}$ For example, HYPERGEOMDIST$(1;4;20;30) = 0,087575$

Example 7.5 (Binomial distribution). A simple discrete random variable

$$\xi(\omega) = \sum_{k=1}^{n} k I_{A_k}(\omega) \;\; (\omega \in \Omega).$$

is called distributed by binomial law with parameter (n,p) if

$$P(A_k) = C_n^k \cdot p^k (1-p)^{n-k}$$

where $0 < p < 1, \; 0 \le k \le n$, i.e.,

$$P(\{\omega : \xi(\omega) = k\}) = C_n^k \cdot p^k (1-p)^{n-k}.$$

The binomial distribution with parameter (n,p) is denoted by $F_n(x,p)$ and is defined by

$$F_n(x,p) = \sum_{k \le x} C_n^k \cdot p^k (1-p)^{n-k}.$$

Remark 7.5 BINOMDIST$(k;n;p;0)$ calculates the value $C_n^k \cdot p^k (1-p)^{n-k}$. For example, BINOMDIST$(3;10;0;5;0) = 0,1171875$. BINOMDIST$(k;n;p;1)$ calculates the sum $\sum_{k \le x} C_n^k \cdot p^k (1-p)^{n-k}$. For example, BINOMDIST$(3;10;0;5;1) = 0,171875$.

Remark 7.6 The random variable distributed by the binomial law with parameter $(1;p)$ is called also a random variable distributed by the Bernoulli law with parameter p. It can be proved that the random variable distributed by the Binomial law with parameter $(n;p)$ can be presented as a sum of n independent random variables each of them is distributed by the Bernoulli law with parameter p.

Definition 7.2 Random variable $\xi : \Omega \to R$ is called absolutely continuous [3] if there exists a non-negative function $f_\xi : R \to R^+$ such that

$$(\forall x)(x \in R \to F_\xi(x) = \int_{-\infty}^{x} f_\xi(x)dx),$$

where $R^+ = [0, +\infty[$.

Function $f_\xi(x)$ $(x \in R)$ is called a density function of random variable ξ.

Theorem 7.5 Let $f_\xi : R \to R$ be a density function of random variable $\xi : \Omega \to R$ Then

$$\int_{-\infty}^{+\infty} f_\xi(x)dx = 1.$$

Proof. Since $\lim_{L\to+\infty} F_\xi(L) = 1$, we have $\lim_{L\to+\infty} \int_{-\infty}^{L} f_\xi(x)dx = 1$. The latter relation means the validity of the following equality

$$\int_{-\infty}^{+\infty} f_\xi(x)dx = 1.$$

This ends the proof of theorem □

Theorem 7.6 *Let F_ξ be a distribution function of an absolutely continuous random variable ξ. Then for arbitrary real numbers x and y $(x < y)$ we have*

$$P(\{\omega : x < \xi(\omega) \le y\}) = F_\xi(y) - F_\xi(x),$$

If ξ is absolutely continuous random variable and f_ξ is its density function, then

$$P(\{\omega : x < \xi(\omega) \le y\}) = \int_{x}^{y} f_\xi(s)ds.$$

Proof.

$$P(\{\omega : x < \xi(\omega) \le y\}) = P(\{\omega : \xi(\omega) \le y\} \setminus \{\omega : \xi(\omega) \le x\}) =$$

$$= P(\{\omega : \xi(\omega) \le y\}) - P(\{\omega : \xi(\omega) \le x\}) = F_\xi(y) - F_\xi(x).$$

If $F_\xi(t) = \int_{-\infty}^{t} f_\xi(s)ds$, then

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

This ends the proof of theorem. □

[3]Note that the density function of the absolutely continuous random variable is defined exactly until null sets (in the Lebesgue sense) of R. We recall the reader that $X \subset R$ is null-set(in Lebesgue sense) if for arbitrary $\varepsilon > 0$ there exists sequence $(]a_k; b_k[)_{k\in N}$ of open intervals such that $X \subseteq \cup_{k\in N}]a_k; b_k[$ and $\sum_{k\in N} b_k - a_k < \varepsilon$.

Remark 7.7 If f_ξ and F_ξ are the density function and the distribution functions respectively, then almost everywhere on R we have

$$\frac{dF_\xi(x)}{dx} = f_\xi(x),$$

i.e., the linear measure l_1 of a set

$\{x : x \in R,\ \frac{dF_\xi(x)}{dx} \neq f_\xi(x)$ or $\frac{dF_\xi(x)}{dx}$ does not exist $\}$

is equal to zero, where l_1 denotes one-dimensional Lebesgue measure on R.

Example 7.6 (Normal distribution). Absolutely continuous random variable $\xi : \Omega \to R$ is called normally distributed with parameters (m, σ^2) $(m \in R,\ \sigma > 0)$ if

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

The density function and the distribution function of the normally distributed random variable with parameters (m, σ^2) are denoted by $\phi_{(m,\sigma^2)}$ and $\Phi_{(m,\sigma^2)}$, respectively, i.e.,

$$\phi_{(m,\sigma^2)}(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}} \ (x \in R),$$

$$\Phi_{(m,\sigma^2)}(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{x} e^{-\frac{(t-m)^2}{2\sigma^2}} dt \ (t \in R).$$

When $m = 0$ and $\sigma = 1$, they are denoted as ϕ and Φ, respectively. ϕ and Φ are called the density function and the distribution function of the standard normally distributed random variable, respectively.

Remark 7.8 NORMDIST$(x; m; \sigma; 0)$ calculates the function

$$\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}}.$$

For example, NORMDIST$(0; 0; 1; 0) = 0,3989428$.

NORMDIST$(x; m; \sigma; 1)$ calculates the integral

$$\frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{x} e^{-\frac{(t-m)^2}{2\sigma^2}} dt.$$

For example, NORMDIST$(0; 0; 1; 1) = 0,5$.

In addition, one can use *Cumulative Normal Distribution Calculator*, placed in the web site

http://stattrek.com/Tables/Normal.aspx

Example 7.6 (The uniform distribution). Absolutely continuous random variable $\xi : \Omega \to R$ is called uniformly distributed on the interval $[a, b]$ $(a < b)$ if

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

Distribution function F_ξ of the random variable uniformly distributed on $[a,b]$ is defined by

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

Example 7.7 (Cauchy [4] distribution). We say that an absolutely continuous random variable $\xi : \Omega \to R$ is distributed by the Cauch law, if

$$f_\xi(x) = \frac{1}{\pi(1+x^2)} \ (x \in R).$$

Its distribution function is defined by

$$F_\xi(x) = \int_{-\infty}^{x} \frac{1}{\pi(1+t^2)} dt = \frac{1}{2} + \frac{1}{\pi} arctg(x) \ (x \in R).$$

Example 7.8 (Exponential distribution). Absolutely continuous random variable $\xi : \Omega \to R$ is distributed by the exponential law with parameter λ, if

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

Its distribution function F_ξ is defined by

$$F_\xi(x) = \begin{cases} 1 - e^{-\lambda x}, & \text{if } x \geq 0; \\ 0, & \text{if } x < 0. \end{cases}$$

Remark 7.9 EXPONDIST$(x;\lambda;0)$ calculates value $\lambda e^{-\lambda x}$ for $x > 0$ and $\lambda > 0$. For example, EXPONDIST$(4;3;0) = 1,84326$. EXPONDIST$(x;\lambda;1)$ calculates value $1 - e^{-\lambda x}$ for for $x > 0$ and $\lambda > 0$. For example, EXPONDIST$(4;3;1) = 0,999993856$.

Example 7.9 (Singular distribution). Let consider closed interval $[0,1]$ and let define a sequence of functions constructed by G.Cantor [5]. Let divide interval $[0,1]$ into three equal parts and define function

$$F_1(x) = \begin{cases} \frac{1}{2}, & \text{if } x \in]\frac{1}{3}, \frac{2}{3}[; \\ 0, & \text{if } x = 0; \\ 1, & \text{if } x = 1. \end{cases}$$

[4]Cauchy; Augustin Louis (21.8.1789, - 23.5.1857) -French mathematician, the member of Paris Academy of Sciences (1816), the honourable member of Petersburg Academy of Sciences(1831).

[5]Cantor; George (19.2.(3.3).1845 -6.1.1918)-German mathematician, professor of Gales University (1879-1913). He had proved that a real numbers axis is not countable

We continue its values on other points of $[0,1]$ by linear interpolation. Further, let consider the division of intervals $[0,\frac{1}{3}]$ and $[\frac{2}{3},1]$ into three equal parts and define

$$F_2(x)=\begin{cases}\frac{1}{2}, & \text{if } x\in]\frac{1}{3},\frac{2}{3}[;\\ \frac{1}{4}, & \text{if } x\in]\frac{1}{9},\frac{2}{9}[;\\ \frac{3}{4}, & \text{if } x\in]\frac{7}{9},\frac{8}{9}[;\\ 0, & \text{if } x=0;\\ 1, & \text{if } x=1.\end{cases}$$

Analogously, we continue the values of F_2 on other points of $[0,1]$ by linear interpolation. If we shall continue this process, then we shall get a sequence of functions $(F_n)_{n\in N}$, which tends uniformly to concrete continuous function F on $[0,1]$, the increase points [6] of which is null-set in the Lebesgue sense. Indeed, we get that the Lebesgue measure of the union of intervals

$$]\frac{1}{2},\frac{2}{3}[,\ ]\frac{1}{9},\frac{2}{9}[,\ ]\frac{7}{9},\frac{8}{9}[,\cdots$$

on which function F is constant, is

$$\frac{1}{3}+\frac{2}{9}+\frac{4}{27}+\cdots=\frac{1}{3}\sum_{n=0}^{\infty}\Big(\frac{2}{3}\Big)^n=1.$$

F is called a Cantor function.

Let consider one construction of the random variable, whose distribution function coincides with Cantor function F.

We set

$$(\Omega,\mathcal{F},P)=([0,1],\mathcal{B}([0,1]),b_1).$$

Let define a sequence of functions

$$(\xi_{\frac{k}{2^n}})_{n\in N,\ 1\le k\le n,\ \&\ k\in 2N+1}=(\xi_i)_{i\in I}$$

defined by the following formulas

$$\xi_{\frac{1}{2}}(\omega)=\frac{1}{3}I_{\{\frac{1}{2}\}}(\omega),\ (\omega\in\Omega),$$

$$\xi_{\frac{1}{4}}(\omega)=\frac{1}{9}I_{\{\frac{1}{4}\}}(\omega),\ (\omega\in\Omega),$$

$$\xi_{\frac{3}{4}}(\omega)=\frac{1}{9}I_{\{\frac{3}{4}\}}(\omega),\ (\omega\in\Omega),$$

$$\xi_{\frac{1}{8}}(\omega)=\frac{1}{27}I_{\{\frac{1}{8}\}}(\omega),\ (\omega\in\Omega),$$

[6] x is called a point of increment for function F if $F(x+\varepsilon)-F(x-\varepsilon)>0$ for arbitrary $\varepsilon>0$.

$$\xi_{\frac{3}{8}}(\omega) = \frac{1}{27} I_{\{\frac{3}{8}\}}(\omega), \ (\omega \in \Omega),$$

$$\xi_{\frac{5}{8}}(\omega) = \frac{1}{27} I_{\{\frac{5}{8}\}}(\omega), \ (\omega \in \Omega),$$

$$\xi_{\frac{7}{8}}(\omega) = \frac{1}{27} I_{\{\frac{7}{8}\}}(\omega), \ (\omega \in \Omega),$$

and so on.

We define $\xi_{Cantor} : \Omega \to R$ by the following formula

$$\xi_{Cantor}(\omega) = \sum_{i \in I, \ i \leq \omega} \xi_i(\omega).$$

It is easy to show that the distribution function generated by ξ_{cantor} cantor coincides with Cantor function F.

Definition 7.3. A continuous distribution function, whose points of the increment have a Lebesgue measure zero, is called singular. The corresponding random variable is also called singular.

Theorem 7.7 *Arbitrary distribution function F admits the following representation*

$$F(x) = p_1 \cdot F_1(x) + p_2 \cdot F_2(x) + p_3 \cdot F_3(x) \ (x \in R),$$

where F_1, F_2, F_3 are distribution functions generated by a discrete, an absolutely continuous and a singular random variables, respectively and p_1, p_2, p_3 are such non-negative real numbers that $p_1 + p_2 + p_3 = 1$.

Theorem 7.8 *Let F_ξ be a distribution function of ξ and $a, b \in R, a \neq 0$. Then the distribution function of $\eta = a\xi + b$ is calculated by*

$$F_\eta(x) = F_\xi\left(\frac{x-b}{a}\right) \ (x \in \overline{R}).$$

Proof. Note, that

$$F_\eta(x) = P(\{\omega : a\xi(\omega) + b \leq x\}) = P(\{\omega : \xi(\omega) \leq \frac{x-b}{a}\}) = F_\xi\left(\frac{x-b}{a}\right).$$

□

Tests

7.1. The distribution law of random variable $\xi(\omega) = \sum_{k=1}^{4} x_k I_{A_k}(\omega) \ (\omega \in \Omega)$ is given by the following table

ξ	-1	0	4	5
P	$0,2$	$0,3$	$0,1$	$0,4$

.

Then

1) $F_\xi(-3)$ is equal to

a) $0,2$, b) $0,3$, c) $0,1$, d) 0;

2) $F_\xi(-1)$ is equal to

a) $0,2$, b) $0,3$, c) $0,1$, d) 0;

3) $F_\xi(-0,3)$ is equal to

a) $0,2$, b) $0,3$, c) $0,1$, d) 0;

4) $F_\xi(4)$ is equal to

a) $0,6$, b) $0,4$, c) 1, d) $0,8$;

5) $F_\xi(6)$ is equal to

a) $0,6$, b) $0,4$, c) 1, d) $0,8$.

7.2. The distribution function of ξ is defined by

$$F_\xi(x) = \begin{cases} a, & x < 0; \\ bx, & 0 \le x < 1; \\ c, & x \ge 1. \end{cases}$$

Then

a) $a = 1, b = 0, c = 0$; b) $a = 0, b = 1, c = 1$;
c) $a = 0, b = 0, c = 1$; d) $a = 1, b = 1, c = 0$;

7.3. The probability that event A will occur in partial experiment is equal to 0,3. Let ξ be the number of experiments in the three independent experiments, when the event A occurred. Then the distribution of ξ is given by the following table

a)

ξ	0	1	2	3
P	$0,343$	$0,441$	$0,189$	$0,027$

,

b)

ξ	0	1	2	3
P	$0,343$	$0,441$	$0,179$	$0,037$

.

7.4. A shot gets 5 points if he strucks a target and loses 2 points in other case. The probability that the shot strucks a target is equal to 0,5. The law of distribution of collected points ξ in 4 shots is given by the following table

a)

ξ	-8	-1	6	13	20
P	$0,24$	$0,41$	$0,26$	$0,08$	$0,01$

,

b)

ξ	-8	-1	6	13	20
P	$0,2$	$0,44$	$0,25$	$0,09$	$0,01$

.

7.5. The complect of 10 details contains 8 non-standard details. We accidentally choose 2 details. Then the law of distribution of number ξ of standard details in our probability sampling is given by the following table

a)

ξ	0	1	2
P	$\frac{1}{45}$	$\frac{16}{45}$	$\frac{28}{45}$

,

b)

ξ	0	1	2
P	$\frac{2}{45}$	$\frac{14}{45}$	$\frac{29}{45}$

.

7.6. The probability that the price of goods will increase or decrease by 1 lari during one unit of time is equal to 0,5 and 0,5, respectively. An initial price of goods is 10 lari. Then the distribution law of price ξ after 4 unites of time is given by the following table

a)

ξ	6	8	10	12	14
P	$0,24$	$0,41$	$0,26$	$0,08$	$0,01$

,

b)

ξ	6	8	10	12	14
P	$0,2$	$0,44$	$0,2$	$0,14$	$0,01$

.

7.7. A particle is placed at the origin of the real axis. The probabilities of shifting to the right or to the left along the real axis during one unit of time are equal (=0,5). The distribution law of states ξ of the particle after 4 unit of time is given by the following table

a)

ξ	-4	-2	0	2	4
P	$0,0625$	$0,25$	$0,375$	$0,25$	$0,0625$

,

b)

ξ	-4	-2	0	2	4
P	$0,0625$	$0,245$	$0,385$	$0,245$	$0,0625$

.

7.8. Let ξ be a Poisson random variable with parameter $\lambda = 1$. Then the probability that

1) ξ will obtain a value in the interval $[2,5;5,5]$ is equal to

a) $0,079707$, b) $0,13455$, c) $0,11213$, d) $0,28111$;

2) $3\xi+4$ will obtain a value in the interval $[6,5;7,5]$ is equal to

a) $0,367879$, b) $0,13894$, c) $0,13121$, d) $0,28991$.

7.9. Let ξ be a random variable uniformly distributed on $[3,10]$. Then

1) $F_\xi(4)$ is equal to

a) $\frac{1}{7}$, b) $\frac{1}{8}$, c) $\frac{1}{9}$, d) $\frac{1}{10}$;

2) the probability that ξ will obtain a value in the interval $[2,5;5,5]$ is equal to

a) $\frac{5}{14}$, b) $\frac{5}{8}$, c) $\frac{5}{9}$, d) $0,5$;

3) the probability that $5\xi+5$ will obtain a value in the interval [5;10] is equal to

a) 0, b) 1, c) $0,5$, d) $0,8$.

7.10. Let ξ be an exponential random variable with parameter $\lambda(\lambda > 0)$.

1) If the probability that ξ will obtain a value in the interval $[0,a]$ is equal $\frac{2}{3}$, then

a) $a = \frac{\ln(3)}{\lambda}$, b) $a = \frac{\ln(4)}{\lambda}$, c) $a = \frac{\ln(5)}{\lambda}$, d) $a = \frac{\ln(6)}{\lambda}$;

2) The probability that $3\xi - 4$ will obtain a value in the interval $[-5;5]$ is equal to

a) $1-e^{-3\lambda}$, b) $1-e^{-4\lambda}$, c) $1-e^{-5\lambda}$, d) $1-e^{-6\lambda}$.

7.11. Let ξ be a standard normal random variable.

1) If the probability that ξ will obtain a value in the interval $[-a,a]$ is equal to 0,99, then

a) $a = 2,37$, b) $a = 2,57$, c) $a = 2,77$, d) $a = 2,97$;

2) The probability that $3\xi+8$ will obtain a value in the interval $(-5,5)$ is equal to

a) $0,1586$, b) $0,7413$, c) $0,6413$, d) $0,5413$.

7.12. The amount of time you have to wait at a particular stoplight is uniformly distributed between zero and two minutes.

1) What is the probability that you have to wait than 30 seconds for light?

a) $0,5$, b) $0,25$, c) $0,75$, d) $1,01$;

2) What is the probability that you have to wait between 15 and 45 seconds for the light?

a) $0,25$, b) $0,28$, c) $0,64$, d) $0,54$.

3) Eighty percent of the time, the light will change before you have to wait how long?

a) 96 seconds , b) 28 seconds, c) 64 seconds, d) 54 seconds.

4) Sixty percent of the time, the light will change before you have to wait how long?

a) 72 seconds , b) 88 seconds, c) 64 seconds, d) 24 seconds.

Chapter 8

Mathematical expectation and variance

Let $(\Omega, \mathcal{F}, P)$ be a probability space and ξ be a simple discrete random variable, i.e.,

$$(\forall\omega)(\omega \in \Omega \to \xi(\omega) = \sum_{k=1}^{n} x_k \cdot I_{A_k}(\omega)),$$

where $x_k \in R$ $(1 \le k \le n)$ and $(A_k)_{1 \le k \le n}$ is the complete system of representatives, i.e.

1) $(\forall k)(\forall m)(1 \le k < m \le n \to A_k \cap A_m = \emptyset)$,

2) $\cup_{k=1}^{n} A_k = \Omega$.

Definition 8.1 A mathematical expectation of the simple random variable ξ is denoted by $M\xi$ and is defined by

$$M\xi = \sum_{k=1}^{n} x_k \cdot P(A_k).$$

Remark 8.1 Assume that in column A of Excel table we have entered values $x_1, \cdots, x_n$ of simple discrete random variable ξ. Assume also that in column B of Excel table we have entered the corresponding probabilities $p_1, \cdots, p_n$. Then the statistical function SUMPRODUCT$(x_1 : x_n; p_1 : p_n)$ calculates mathematical expectation of ξ.

Assume that η be an arbitrary random variable. Following Theorem 6.3 (cf. Chapter 6), there exists a sequence $(\eta_n)_{n \in N}$ of simple random variables such that

$$(\forall\omega)(\omega \in \Omega \to \eta(\omega) = \lim_{n\to\infty} \eta_n(\omega)).$$

Definition 8.2 If there exists a finite limit $\lim_{n\to\infty} M\eta_n$, then this limit is called a mathematical expectation of η and is denoted by $M\eta$ (or $\int_\Omega \eta(\omega)dP(\omega)$). It can be proved that if there exists a finite limit $\lim_{n\to\infty} M\eta_n$, then this limit is same for arbitrary sequence of simple random variables tending to η, which means a correctness of Definition 8.2.

Agreement. In the sequel we consider such a class of random variables each element ξ of which satisfies the conditions: $M(\xi) < \infty$ and $M(\xi^2) < \infty$.

Theorem 8.1 *If f_ξ is a density function of an absolutely continuous random variable ξ, then*

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

Definition 8.3 Value $M(\xi - M\xi)^2$ is called variance of ξ and is denoted by $D\xi$.

Definition 8.4 Value $\sqrt{D\xi}$ is called a mean absolute deviation of the random variable ξ and is denoted by $\sigma(\xi)$.

Let consider some properties of mathematical expectations and variances of random variables.

Theorem 8.2 *Let $\xi(\omega) = c$ $(\omega \in \Omega,\ c = const)$. Then $M\xi = c$.*

Proof. Following the definition of the expectation of the simple discrete random variable, we have

$$M\xi = M(c \cdot I_\Omega(\omega)) = c \cdot P(\Omega) = c.$$

□

Theorem 8.3 $M(\xi + \eta) = M\xi + M\eta$ *(i.e., mathematical expectation of the sum of two random variables is equal to the sum of expectations of corresponding random variables).*

Proof. Using the approximation property of a random variable by a sequence of simple discrete random variables and by the definition of the expectation of a random variable, it is sufficient to prove this theorem in the case of two simple discrete random variables. Now assume that ξ and η be simple random variables, i.e.

$$\xi(\omega) = \sum_{k=1}^{p} x_k \cdot I_{A_k}(\omega),\ A_k \cap A_m = \emptyset,\ 1 \le k < m \le p,$$

$$\cup_{k=1}^{p} A_k = \Omega,\ x_k \in R,\ k,m,p \in N,$$

$$\eta(\omega) = \sum_{n=1}^{q} y_n \cdot I_{B_n}(\omega),\ B_k \cap B_m = \emptyset,\ 1 \le k < m \le q,$$

$$\cup_{n=1}^{p} B_n = \Omega,\ y_n \in R,\ k,m,q \in N,$$

Note that

$$(\xi+\eta)(\omega)=\sum_{k=1}^{p}\sum_{n=1}^{q}(x_k+y_n)\cdot I_{A_k\cap B_n}(\omega)\ (\omega\in\Omega),$$

It follows

$$M(\xi+\eta)=\sum_{k=1}^{p}\sum_{n=1}^{q}(x_k+y_n)\cdot I_{A_k\cap B_n}(\omega)=\sum_{k=1}^{p}\sum_{n=1}^{q}(x_k+y_n)\cdot P(A_k\cap B_n)$$

$$=\sum_{k=1}^{p}x_k\sum_{n=1}^{q}P(A_k\cap B_n)+\sum_{n=1}^{q}y_n\sum_{k=1}^{p}P(A_k\cap B_n)=$$

$$=\sum_{k=1}^{p}x_kP(A_k)+\sum_{n=1}^{q}y_nP(B_n)=M\xi+M\eta.$$

This ends the proof of theorem. □

Definition 8.5 Two simple discrete random variables ξ and η are called independent, if

$$P(\{\omega:\xi(\omega)=x_k,\ \eta(\omega)=y_n\})=P(\{\omega:\xi(\omega)=x_k\})\cdot P(\{\omega:\ \eta(\omega)=y_n\}),$$

where $1\le k\le p,\ 1\le n\le q$.

Definition 8.6 Two random variables ξ and η are called independent if

$$P(\{\omega:\xi(\omega)\le x,\ \eta(\omega)\le y\})=P(\{\omega:\xi(\omega)\le x\})\cdot P(\{\omega:\ \eta(\omega)\le y\}),$$

where $x,\ y\ \in R$.

Remark 8.1 Definitions 8.5 and 8.6 are equivalent for simple discrete random variables.

Theorem 8.4 *Let ξ and η be independent random variables. Then there exist two sequences $(\xi_n)_{n\in N}$ and $(\eta_n)_{n\in N}$ of simple discrete random variables such that :*
1) ξ_n and η_n are independent for $n\in N$.
2) $(\xi\cdot\eta)(\omega)=\lim_{n\to\infty}\xi_n(\omega)\cdot\eta_n(\omega)\ \ (\omega\in\Omega)$.

Theorem 8.5 *If ξ and η are independent simple discrete random variables, then*

$$M(\xi\cdot\eta)=M\xi\cdot M\eta,$$

i.e., mathematical expectation of the product of two independent simple discrete random variables is equal to the product of expectations of corresponding simple random variables.

Proof. Note that

$$(\xi \cdot \eta)(\omega) = \sum_{k=1}^{p} \sum_{n=1}^{q} (x_k \cdot y_n) \cdot I_{A_k \cap B_n}(\omega) \ \ (\omega \in \Omega).$$

It follows

$$M(\xi \cdot \eta) = M(\sum_{k=1}^{p} \sum_{n=1}^{q} x_k \cdot y_n \cdot I_{A_k \cap B_n}) = \sum_{k=1}^{p} \sum_{n=1}^{q} x_k \cdot y_n P(A_k \cap B_n) =$$

$$= \sum_{k=1}^{p} \sum_{n=1}^{q} x_k \cdot y_n P(A_k) \cdot P(B_n) = \sum_{k=1}^{p} x_k P(A_k) \cdot \sum_{n=1}^{q} y_n P(B_n) = M\xi \cdot M\eta.$$

This ends the proof of theorem. □

Using Theorems 8.4 and 8.5, we get the validity of the following theorem.

Theorem 8.6 *If* ξ *and* η *are independent random variables, then*

$$M(\xi \cdot \eta) = M\xi \cdot M\eta,$$

i.e., mathematical expectation of the product of two random independent variables is equal to the product of expectations of the corresponding variables.

Proof. If ξ and η are two independent random variables then using Theorem 8. 4, there exist two sequences $(\xi_n)_{n \in N}$ and $(\eta_n)_{n \in N}$ of simple discrete random variables such that :

1) ξ_n and η_n are independent for $n \in N$.
2) $(\xi \cdot \eta)(\omega) = \lim_{n \to \infty} \xi_n(\omega) \cdot \eta_n(\omega) \ \ (\omega \in \Omega)$.

By using Definition 8.5 and the result of Theorem 8.5, we get

$$M(\xi \cdot \eta) = \lim_{n \to \infty} M\xi_n \cdot \eta_n = \lim_{n \to \infty} \left(M\xi_n \cdot M\eta_n \right) =$$

$$= \lim_{n \to \infty} M\xi_n \cdot \lim_{n \to \infty} M\eta_n = M\xi \cdot M\eta.$$

This ends the proof of theorem. □

Definition 8.7 A finite family of random variables $\xi_1, \cdots, \xi_n$ is called independent

$$P(\{\omega : \xi_1(\omega) \le x_1, \cdots, \xi_n(\omega) \le x_n\}) = \prod_{k=1}^{n} P(\{\omega : \xi_k(\omega) \le x_k\}).$$

for every $(x_k)_{1 \le k \le n} \in (R \cup \{+\infty\} \cup \{-\infty\})^n$.

Definition 8.8 A sequence of random variables $(\xi_n)_{n \in N}$ is called independent if family $(\xi_k)_{1 \le k \le n}$ is independent for arbitrary $n \in N$.

Remark 8.2 An analogy of Theorem 8.6 is valid for arbitrary finite family of random variables, i.e., if $(\xi_k)_{1\leq k\leq n}$ is a family of independent random variables, then

$$M(\prod_{k=1}^{n}\xi_k) = \prod_{k=1}^{n} M\xi_k.$$

Theorem 8.7 *If $c \in R$, then $M(c\xi) = cM\xi$, i.e., constant c goes out from the symbol of mathematical expectation.*

Proof. Note that a constant random variable c and an arbitrary random variable are independent. Following Theorem 8.6, we have

$$M(c\cdot\xi) = Mc\cdot M\xi = cM\xi.$$

This ends the proof of theorem. □

Theorem 8.8 *(Cauchy-Buniakovski [1]inequality). For arbitrary random variables ξ and η the following inequality*

$$|M(\xi\cdot\eta)| \leq \sqrt{M\xi^2}\cdot\sqrt{M\eta^2}$$

holds.

Proof. Let consider value $M(\xi+x\eta)^2$. Clearly, on the one hand, we have $M(\xi+x\eta)^2 \geq 0$ for arbitrary $x \in R$. Hence, the following expression

$$M(\xi+x\eta)^2 = M\xi^2 + 2M(\xi\cdot\eta)\cdot x + M\eta^2\cdot x^2$$

can be considered as a non-negative quadratic polynomial. Hence, its determinant must be non-positive, i.e.,

$$(2M(\xi\cdot\eta))^2 - 4M\eta^2\cdot M\xi^2 \leq 0,$$

which is equivalent to the condition

$$|M(\xi\cdot\eta)| \leq \sqrt{M\xi^2}\cdot\sqrt{M\eta^2}$$

This ends the proof of theorem. □

Theorem 8.9 *The following formula for calculation of variance*

$$D\xi = M\xi^2 - (M\xi)^2.$$

is valid for arbitrary random variable ξ.

[1]Buniakovski, Victor [4(16).12.1804 - 30.11 (12.12). 1889] -Russian mathematician, Academician of Petersburg Academy of Sciences (1830)).

Proof. By the definition of variance of ξ, we have

$$D\xi = M(\xi - M\xi)^2.$$

Using the properties of mathematical expectation $M\xi$ we have

$$D\xi = M(\xi - M\xi)^2 = M(\xi^2 - 2\xi M\xi + (M\xi)^2) =$$

$$= M\,\xi^2 - M(2\xi M\xi) + M((M\xi)^2) =$$

$$= M\xi^2 - 2M\xi M\xi + (M\xi)^2 = M\xi^2 - (M\xi)^2.$$

This ends the proof of theorem. □

Theorem 8.10 *For arbitrary random variable* ξ *the following equality*

$$D\xi = \min_{a \in R} M(\xi - a)^2.$$

Proof. Let calculate a minimum value of function $M(\xi - a)^2$. Clearly,

$$M(\xi - a)^2 = M(\xi^2 - 2a\xi + a^2) = M\xi^2 - 2M\xi a + a^2,$$

i.e., $M(\xi - a)^2$ is a quadratic polynomial with respect to a. Hence, point $a_{\min}$ is defined by

$$\frac{dM(\xi - a)^2}{and} = -2M\xi + 2a = 0.$$

It follows that $a_{\min} = M\xi$, i.e.,

$$\min_{a \in R} M(\xi - a)^2 = M(\xi - a_{\min})^2 = M(\xi - M\xi)^2 = D\xi.$$

This ends the proof of theorem □

Theorem 8.11 *For arbitrary random variable* ξ *the following conditions*

1) $D\xi \geq 0$,

2) $D\xi = 0 \Leftrightarrow (\exists c)(c \in R \to P(\{\omega : \xi(\omega) = c\}) = 1)$

are fulfilled.

Proof. Since $D\xi = M(\xi - M\xi)^2$ and $(\xi - M\xi)^2 \geq 0$, we easily deduce the validity of part 1). Let us prove part 2). Let $P(\{\omega : \xi(\omega) = c\}) = 1$, then $M\xi = c$ and $M\xi^2 = c^2$. Following Theorem 8.9 we get $D\xi = M\xi^2 - (M\xi)^2 = c^2 - c^2 = 0$. Now, if $D\xi = 0$, then $M(\xi - M\xi)^2 = 0$. i.e. $P(\{\omega : \xi(\omega) = M\xi\}) = 1$. Hence, it is sufficient to set $c = M\xi$.

This ends the proof of Theorem. □

Theorem 8.12 *Let $c \in R$ and ξ be an arbitrary random variable. Then :*
1) $D(c\xi) = c^2 D\xi$,
2) $D(c+\xi) = D\xi$.

Proof. Following Theorem 8.7 and the definition of the variance, we get

$$D(c\xi) = M(c\xi - M(c\xi))^2 = M(c\xi - cM\xi)^2 = M(c^2(\xi - M\xi)^2) =$$

$$= c^2 M(\xi - M\xi)^2 = c^2 D\xi.$$

This ends the proof of the part 1).

By definition of the variance of $\xi + c$, we get

$$D(c+\xi) = M((c+\xi) - M(c+\xi))^2 = M(c+\xi - Mc - M\xi)^2 =$$

$$= M(c+\xi - c - M\xi)^2 = M(\xi - M\xi)^2 = D\xi.$$

This ends the proof of the part 2) and theorem is proved. □

Theorem 8.13 *Let ξ and η be independent random variables. Then*

$$D(\xi + \eta) = D\xi + D\eta.$$

Proof. Since random variables ξ and η are independent, we get

$$D(\xi+\eta) = M((\xi+\eta) - M(\xi+\eta))^2 = M((\xi - M\xi) + (\eta - M\eta))^2 =$$

$$= M((\xi - M\xi)^2 + 2(\xi - M\xi)(\eta - M\eta) + (\eta - M\eta)^2) =$$

$$= M(\xi - M\xi)^2 + 2M((\xi - M\xi)(\eta - M\eta)) + M(\eta - M\eta)^2 =$$

$$= D\xi + 2M(\xi - M\xi)M(\eta - M\eta) + D\eta =$$

$$= D\xi + 2(M\xi - M(M\xi))(M\eta - M(M\eta)) + D\eta =$$

$$= D\xi + 2(M\xi - M\xi)(M\eta - M\eta) + D\eta = D\xi + D\eta.$$

This ends the proof of theorem

□

Remark 8.3 Note that an analogy of Theorem 8.13 is valid for arbitrary finite family $(\xi_k)_{1 \le k \le n}$ of independent random variables, i.e., the following equality

$$D\sum_{k=1}^{n} \xi_k = \sum_{k=1}^{n} D\xi_k.$$

holds.

Theorem 8.14 *Let F_ξ be a distribution function of the absolutely continuous random variable ξ. Then the following formula for calculation of variance*

$$D\xi = \int_{-\infty}^{+\infty} (x - M\xi)^2 f_\xi(x)dx.$$

is valid.

Let consider some examples for calculation of mathematical expectations and mathematical variances.

Example 8.1 (Poisson distribution). Let

$$\xi(\omega) = \sum_{n \in N} n \cdot I_{A_n}(\omega) \ \ (\omega \in \Omega)$$

be a discrete random variable distributed by Poisson law with parameter λ ($\lambda > 0$), i.e.,

$$P(A_n) = P(\{\omega : \xi(\omega) = n\}) = \frac{\lambda^n}{n!} e^{-\lambda} \ \ (n \in N).$$

Then

$$M\xi = \sum_{n=0}^{\infty} n \cdot \frac{\lambda^n}{n!} e^{-\lambda} = \sum_{n=1}^{\infty} n \cdot \frac{\lambda^n}{n!} e^{-\lambda} =$$

$$= \lambda \sum_{n=1}^{\infty} n \cdot \frac{\lambda^{n-1}}{n!} e^{-\lambda} = \lambda e^{-\lambda} \sum_{m=0}^{\infty} \frac{\lambda^m}{m!} = \lambda.$$

On the other hand, we have

$$M\xi^2 = \sum_{n=0}^{\infty} n^2 \cdot \frac{\lambda^n}{n!} e^{-\lambda} = \sum_{n=1}^{\infty} n^2 \cdot \frac{\lambda^n}{n!} e^{-\lambda} =$$

$$= \sum_{n=1}^{\infty} n \cdot \frac{\lambda^n}{(n-1)!} e^{-\lambda} = \sum_{n=1}^{\infty} \frac{(n-1)\lambda^n}{(n-1)!} e^{-\lambda} +$$

$$+ \sum_{n=1}^{\infty} \frac{\lambda^n}{(n-1)!} e^{-\lambda} = \lambda \sum_{m=0}^{\infty} \frac{m\lambda^m}{m!} e^{-\lambda} +$$

$$+ \lambda \sum_{m=0}^{\infty} \frac{\lambda^m}{m!} e^{-\lambda} = \lambda^2 + \lambda = \lambda(1 + \lambda).$$

Following Theorem 8.9, we get

$$D\xi = M\xi^2 - (M\xi)^2 = \lambda(1 + \lambda) - \lambda^2 = \lambda.$$

Example 8.2 (Geometric distribution). Let

$$\xi(\omega) = \sum_{n \in N} n \cdot I_{A_n}(\omega) \ \ (\omega \in \Omega)$$

be a discrete random variable distributed by the geometric law with parameter q, where $(0 \le q \le 1)$, i.e.,

$$P(A_n) = P(\{\omega : \xi(\omega) = n\}) = (1-q)q^{n-1} \quad (n \in N).$$

Then

$$M\xi = \sum_{n=1}^{\infty} n(1-q)q^{n-1} = (1-q)\cdot\sum_{n=1}^{\infty} nq^{n-1} = (1-q)(\sum_{n=1}^{\infty} q^n)' = (1-q)(\frac{1}{1-q})' =$$

$$= (1-q)\cdot\frac{1}{(1-q)^2} = \frac{1}{1-q}.$$

On the other hand, we have

$$M\xi^2 = \sum_{k=1}^{\infty} k^2(1-q)q^{k-1} = (1-q)\cdot\sum_{k=1}^{\infty}(k\cdot kq^{k-1}) = (1-q)\cdot\sum_{k=1}^{\infty}(kq^k)' =$$

$$= (1-q)\cdot\sum_{k=1}^{\infty}(kq^{k-1}\cdot q)' = (1-q)\cdot\big[(\sum_{k=1}^{\infty} kq^{k-1})'q + \sum_{k=1}^{\infty} kq^{k-1}\big] =$$

$$= (1-q)\big[\frac{2q}{(1-q)^3} + \frac{1}{(1-q)^2}\big] = \frac{2q}{(1-q)^2} + \frac{1}{(1-q)}.$$

Hence

$$D\xi = M\xi^2 - (M\xi)^2 = \frac{2q}{(1-q)^2} + \frac{1}{(1-q)} - \frac{1}{(1-q)^2} = \frac{q}{(1-q)^2}.$$

Example 8.3 (Leibniz [2] distribution). Let

$$\xi(\omega) = \sum_{n=1}^{\infty} n\cdot I_{A_n}(\omega) \ (\omega \in \Omega)$$

be a discrete random variable distributed by the Leibniz law, i.e.,

$$P(A_n) = P(\{\omega : \xi(\omega) = n\}) = \frac{1}{n\cdot(n+1)} \quad (n \in N).$$

Then

$$\sum_{n=1}^{\infty} n\cdot\frac{1}{n\cdot(n+1)} = \sum_{n=1}^{\infty}\frac{1}{(n+1)} = +\infty.$$

Hence, mathematical expectation $M\xi$ and variance $D\xi$ are not finite.

Example 8.4 (Binomial distribution) Let

$$\xi(\omega) = \sum_{k=0}^{n} kI_{A_k}(\omega) \quad (\omega \in \Omega)$$

[2]Leibniz, Gottfried Wilhelm (1.7.1646 - 14.11.1716)-German mathematician, the member of London Royal Society (1673), the member of Paris Academy of Sciences (1700).

be a simple random variable distributed by the Binomial law with parameters (n, p), i.e.,

$$P(A_k) = P(\{\omega : \xi(\omega) = k\}) = C_n^k \cdot p^k (1-p)^{n-k},$$

where $0 \le p \le 1$ and $0 \le k \le n$. Then

$$M\xi = \sum_{k=0}^{n} k \cdot C_n^k \cdot p^k (1-p)^{n-k} = \sum_{k=0}^{n} k \cdot \frac{n!}{k!(n-k)!} \cdot p^k (1-p)^{n-k} =$$

$$= n \cdot p \sum_{k=1}^{n} \frac{(n-1)!}{(k-1)!(n-k)!} \cdot p^{k-1} (1-p)^{(n-1)-(k-1)} =$$

$$= n \cdot p \sum_{k-1=0}^{n-1} \frac{(n-1)!}{(k-1)!(n-k)!} \cdot p^{k-1} (1-p)^{(n-1)-(k-1)} =$$

$$= n \cdot p \sum_{s=0}^{n-1} \frac{(n-1)!}{s!((n-1)-s)!} \cdot p^s (1-p)^{(n-1)-s} =$$

$$= n \cdot p \sum_{s=0}^{n-1} C_{n-1}^s \cdot p^s (1-p)^{(n-1)-s} = n \cdot p.$$

Remark 8.4 Let η be a random variable distributed by the Bernoulli law with parameter p, i. e. ,

$$\eta(\omega) = 0 \cdot I_{A_0}(\omega) + 1 \cdot I_{A_1}(\omega) \ \ (\omega \in \Omega),$$

and

$$P(A_0) = P(\{\omega : \eta(\omega) = 0\}) = 1 - p, \ P(A_1) = P(\{\omega : \eta(\omega) = 1\}) = p.$$

Then

$$M\eta = 0 \cdot P(\{\omega : \eta(\omega) = 0\}) + 1 \cdot P(\{\omega : \eta(\omega) = 1\}) = 1 \cdot (1-p) + 1 \cdot p = p.$$

On the other hand, we have

$$P(\{\omega : \eta^2(\omega) = 0\}) = 1 - p, \ P(\{\omega : \eta^2(\omega) = 1\}) = p,$$

Hence

$$M(\eta^2) = 0 \cdot P(\{\omega : \eta^2(\omega) = 0\}) + 1 \cdot P(\{\omega : \eta^2(\omega) = 1\}) = 1 \cdot (1-p) + 1 \cdot p = p.$$

Finally we get

$$D(\eta) = M\eta^2 - (M\eta)^2 = p - p^2 = p(1-p).$$

As simple discrete random variable distributed by Binomial law with parameter (n, p) can be presented as a sum of n exemplars of independent simple discrete random variables distributed by Bernoulli law with parameter p. Hence, following Theorem 8.3, we get

$$M\xi = M(\sum_{k=1}^{n} \xi_k) = \sum_{k=1}^{n} M\xi_k = np.$$

Following Remark 8.3, we get

$$D\xi = D(\sum_{k=1}^{n} \xi_k) = \sum_{k=1}^{n} D\xi_k = np(1-p).$$

Example 8.5 (Normal distribution). Let $\xi : \Omega \to R$ be a normally distributed random variable with parameter $(m, \sigma^2)(m \in R, \sigma > 0)$, i.e.,

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

Then, following Theorem 8.1, we get

$$M\xi = \int_{-\infty}^{+\infty} x f_\xi(x)dx = \int_{-\infty}^{+\infty} x \cdot \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}} dx =$$

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

$$= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} z \cdot e^{-\frac{z^2}{2\sigma^2}} dz + m = m.$$

Using the formula for calculation of variance, we get

$$D\xi = \int_{-\infty}^{+\infty} (x-m)^2 f_\xi(x)dx = \int_{-\infty}^{+\infty} (x-m)^2 f_\xi(x)dx =$$

$$= \int_{-\infty}^{+\infty} (x-m)^2 \cdot \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}} dx = \int_{-\infty}^{+\infty} z^2 \cdot \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{z^2}{2\sigma^2}} dz,$$

where $z = x - m$. Setting $t = \frac{z}{\sigma}$, we get

$$D\xi = \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} t^2 e^{-\frac{t^2}{2\sigma^2}} dz = \frac{\sigma^2}{\sqrt{2\pi}} \sqrt{2\pi} = \sigma^2,$$

because

$$\int_{-\infty}^{+\infty} t^2 e^{-\frac{t^2}{2\sigma^2}} dz = \sqrt{2\pi}.$$

Example 8.6 (Uniform distribution on [a; b]). Let $\xi : \Omega \to R$ be a random variable uniformly distributed on $[a, b]$ $(a < b)$, i.e.,

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

Then

$$M\xi = \int_{-\infty}^{+\infty} x f_\xi(x)dx = \int_a^b x \cdot \frac{1}{b-a} dx = \frac{x^2}{2(b-a)}\Big|_a^b = \frac{b^2-a^2}{2(b-a)} = \frac{a+b}{2}.$$

On the other hand, we have

$$M\xi^2 = \int_{-\infty}^{+\infty} x^2 f_\xi(x)dx = \int_a^b x^2 \cdot \frac{1}{b-a}dx = \frac{x^3}{3(b-a)}\Big|_a^b = \frac{b^2+ab+a^2}{3}.$$

Following Theorem 8.9, we get

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

Example 8.7 (Cauchy distribution). Let $\xi : \Omega \to R$ be an absolutely continuous random variable distributed by the Cauchy law, i.e.,

$$f_\xi(x) = \frac{1}{\pi(1+x^2)} \quad (x \in R).$$

Note that the following indefinite integral

$$\int_{-\infty}^{+\infty} x f_\xi(x)dx = \int_{-\infty}^{+\infty} x \frac{1}{\pi(1+x^2)}dx$$

does not exists. Hence, we deduce that there exists no a mathematical expectation of the random variable distributed by the Cauchy law.

Example 8.8 (Exponential distribution). Let $\xi : \Omega \to R$ be an absolutely continuous random variable distributed by the exponential law with parameter λ, i.e.,

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

Then

$$M\xi = \int_{-\infty}^{+\infty} x f_\xi(x)dx = \int_0^{+\infty} x\lambda e^{-\lambda x}dx =$$

$$= \lambda\big(-\frac{1}{\lambda}xe^{-\lambda x}\big|_0^\infty + \frac{1}{\lambda}\int_0^\infty e^{-\lambda x}dx\big) = \lambda\big(-\lim_{l\to\infty}\frac{1}{\lambda}\frac{l}{e^{\lambda l}} + \frac{1}{\lambda^2}\big) =$$

$$= \lambda\big(-\lim_{l\to\infty}\frac{1}{\lambda}\frac{1}{\lambda e^{\lambda l}} + \frac{1}{\lambda^2}\big) = \lambda \cdot \frac{1}{\lambda^2} = \frac{1}{\lambda}.$$

Using analogous calculations, we get

$$D\xi = \int_{-\infty}^{+\infty} x^2 f_\xi(x)dx - (M\xi)^2 = \lambda\int_0^{+\infty} x^2 e^{-\lambda} x dx - \frac{1}{\lambda^2} =$$

$$= \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$$

Example 8.9 (Singular distribution). Let consider Cantor's random variable ξ_{Cantor}, defined on $[0,1]$. It is easy to show that

$$\int_0^1 \xi_{Cantor}(y)dy + \int_0^1 F(x)dx = 1,$$

where F denotes the Cantor function defined on $[0,1]$. Hence

$$M\xi_{Cantor} = \int_0^1 \xi_{Cantor}(y)dy = 1 - \int_0^1 F(x)dx.$$

Note that for set obtained by counterclockwise rotation about point $(\frac{1}{2},\frac{1}{2})$ on angle π of set $\Delta_1 = \{(x,y) : x \in [0,1], 0 \le y \le F(x)\}$ we have:

a) $b_2(\Delta_1 \cap \Delta_2) = 0$,

b) $b_2(\Delta_1) = b_2(\Delta_2)$,

c) $\Delta_1 \cup \Delta_2 = [0,1] \times [0,1]$.

Hence $b_2(\Delta_1) = b_2(\Delta_2) = \frac{1}{2}$. It follows that

$$M\xi_{Cantor} = 1 - \int_0^1 F(x)dx = 1 - b_2(\Delta_1) = 1 - \frac{1}{2} = \frac{1}{2}.$$

Now let calculate $D(\xi_{Cantor})$. Note that $\pi M\xi^2_{Cantor}$ coincides with the volume of the object, obtained by rotation of the set Δ_2 about the real axis $0Y$, which is equal to the difference of volumes of sets $[0,1] \times [0,1]$ and

$$[\frac{1}{3},\frac{2}{3}] \times [0,\frac{1}{2}] \cup [\frac{1}{9},\frac{2}{9}] \times [0,\frac{1}{4}] \cup [\frac{7}{9},\frac{8}{9}] \times [0,\frac{3}{4}] \cup \cdots,$$

respectively. Hence,

$$M\xi^2_{Cantor} = 1 - \Big[\frac{1}{2}\big((\frac{2}{3})^2 - (\frac{1}{3})^2\big) + \frac{1}{4}\big((\frac{2}{9})^2 - (\frac{1}{9})^2\big) + \frac{3}{4}\big((\frac{8}{9})^2 - (\frac{7}{9})^2\big) + \cdots\Big].$$

Remark 8.5. (Physical sense of mathematical expectation and variance). We remind the reader that arbitrary random variable $\xi : \Omega \to R$ can be considered as a special rule of dispersion of the unit mass of powder Ω on the real axis R, by means of which every particle $\omega \in \Omega$ is send on the particle $A \in R$ with coordinate $\xi(\omega)$. Here naturally arises the following

Problem. What physical sense is put in $M\xi$ and $D\xi$, respectively ?

It is well known from the course of theoretical mechanics that if mass p_k is placed at point $x_k \in R$ for $1 \le k \le n$ and $\sum_{k=1}^n p_k = 1$, then center x_c of the whole mass is calculated by :

$$x_c = \sum_{k=1}^n x_k \cdot p_k.$$

If the rule of dispersion of the unit mass of powder Ω on real axis R is a simple discrete random variable given by the following table

ξ	x_1	x_2	$\cdots$	x_n
P	p_1	p_2	$\cdots$	p_k

,

then $M\xi = x_c$, which means that $M\xi$ is a center of the unite mass distributed by the law ξ on real axis R. Note that the physical sense of $M\xi$ is same in the case of arbitrary random variable ξ.

On the other hand, if ξ is a simple discrete random variable, then

$$D\xi = \sum_{k=1}^{n} (x_k - M\xi)^2 p_k.$$

Note that value $D\xi$ depends on values $((x_k - M\xi)^2)_{1 \le k \le n}$. The latter relation means that the particles $(x_k)_{1 \le k \le n}$ of mass are concentrated nearer to its center $M\xi$ as well a variance $D\xi$ is near at zero. In particular, if $x_1 = \cdots = x_n = M\xi$, then $D\xi = 0$. Hence, variance $D\xi$ can be considered as a characterization why the particles of the unite mass of the powder are removed about its center $M\xi$. As an example let consider random variables ξ_1 and ξ_2, defined by

ξ_1	-1	1
P	$\frac{1}{2}$	$\frac{1}{2}$

ξ_2	-2	2
P	$\frac{1}{2}$	$\frac{1}{2}$

It is obvious that

$$M\xi_1 = M\xi_2 = 0,$$

$$D\xi_1 = 1 \cdot \frac{1}{2} + 1 \cdot \frac{1}{2} = 1,$$

$$D\xi_2 = 4 \cdot \frac{1}{2} + 4 \cdot \frac{1}{2} = 4.$$

Note that, on the one hand, the centers of the particles of the unit mass of powder Ω dispersed by laws ξ_1 and ξ_2, respectively, coincide and are equal to zero, i.e., $M\xi_1 = M\xi_2 = 0$. On the other hand, the particles of the unit mass of powder Ω dispersed by rule ξ_1 are more nearer to the center than the particles of the unit mass of powder Ω dispersed by rule ξ_1.

Remark 8.1 Let $x_1, \cdot, x_2$ be the results of observation on the random variable with finite mathematical expectation and with finite variance. Then:

1) AVERAGE$(x_1 : x_n)$ calculates $\frac{1}{n}\sum_{i=1}^{n} x_i$.
2) VARP$(x_1 : x_n)$ calculates $\frac{1}{n}\sum_{i=1}^{n}(x_i - \frac{1}{n}\sum_{j=1}^{n} x_j)^2$.
3) VAR$(x_1 : x_n)$ calculates $\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \frac{1}{n}\sum_{j=1}^{n} x_j)^2$.

Tests

8.1. Distribution laws of two independent random variables ξ and η are given in the following tables

ξ	-1	0	1	2
P	$0,3$	$0,2$	$0,1$	$0,4$

,

η	-1	0	-2
P	$0,5$	$0,3$	$0,2$

.

Then

1) $M(3\xi-4\eta)$ is equal to

a) $5,3$, b) $5,4$, c) $5,5$, d) $5,6$;

2) $D(3\xi-4\eta)$ is equal to

a) $20,4,3$, b) $21,5$, c) $22,6$, d) $23,7$;

8.2. Distribution function F_ξ of the absolutely continuous random variable ξ has the following form

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

Then

1) $M(3\xi-4)$ is equal to

a) 3, b) -3, c) 4, d) -4;

2) $D(\sqrt{18}\xi-4)$ is equal to

a) $0,3$, b) $0,7$, c) 1, d) $1,3$.

8.3. Let ξ_1 be a random variable normally distributed with parameters (3; 25), ξ_2 be a random variable uniformly distributed in interval (18; 20) and ξ_3 be a random variable distributed by the Poisson law with parameter $\lambda=5$. Then

1) $M(1\xi_1+2\xi_2+3\xi_3)$ is equal to

a) 34, b) 35, c) 36, d) 37;

2) If ξ_1, ξ_2, ξ_3 are independent random variables, then $D(1\xi_1+2\xi_2+3\xi_3+4)$ is equal to

a) $25\frac{1}{3}$, b) $26\frac{1}{3}$, c) $27\frac{1}{3}$, d) $28\frac{1}{3}$.

8.4. Distribution laws of two independent random variables ξ and η are given in the following tables

ξ	-1	1	2
P	$0,2$	$0,1$	$0,7$

,

η	2	3	-1
P	$0,3$	$0,3$	$0,4$

.

Then

1) distribution law of $\xi\eta$ is given in the following table

a)

$\xi\eta$	-3	-2	-1	1	2	3	4	6
P	$0,06$	$0,34$	$0,04$	$0,08$	0.03	$0,03$	$0,21$	$0,21$

,

b)

$\xi\eta$	-3	-2	-1	1	2	3	4	6
P	$0,05$	$0,35$	$0,03$	$0,09$	0.03	$0,02$	$0,22$	$0,21$

.

2) distribution law $\xi+\eta$ is given in the following table

a)

$\xi+\eta$	-2	0	1	2	3	4	5
P	0,08	0,04	0,34	0,06	0.03	0,24	0,21

,

b)

$\xi+\eta$	-2	0	1	2	3	4	5
P	0,06	0,06	0,34	0,06	0.02	0,25	0,21

.

8.5. You have recently joined a country club. The number of times you expect to play golf in a month is represented by a random variable with a mean of 10 and a standard deviation of 2.2. Assume you pay monthly membership fees of 500 dollars per month and pay an additional 50 dollars per round of golf.

1) What is the average monthly bill from the country club?

a) 1000 dollars, b) 900 dollars, c) 800 dollars, d) 700 dollars;

2) What is the standard deviation for your average monthly bill from the country club?

a) 110 dollars, b) 200 dollars, c) 300 dollars, d) 400 dollars;

8.6. Let the random variable ξ follow a standard normal distribution.

1) What is $P(\{\omega : \xi(\omega) > 1,2\})$?

a) $0,1151$, b) $0,2110$, c) $0,800$, d) $0,567$;

2) What is $P(\{\omega : \xi(\omega) > -0,21\})$?

a) $0,5832$, b) $0,5678$, c) $0,5438$, d) $0,5675$;

3) What is $P(\{\omega : 0,33 < \xi(\omega) < 0,45\})$?

a) $0,0443$, b) $0,0678$, c) $0,0438$, d) $0,0675$;

8.7. Let the random variable ξ follow a standard normal distribution with a mean of 17.1 and a standard deviation of $3,2$.

1) What is $P(\{\omega : \xi(\omega) > 16\})$?

a) $0,6331$, b) $0,4562$, c) $0,5678$, d) $0,5678$;

2) What is $P(\{\omega : 15 < \xi(\omega) < 20\})$?

a) $0,5640$, b) $0,4321$, c) $0,2225$, d) $0,1234$;

8.8. Let the random variable ξ follow a standard normal distribution with a mean of 61.7 and a standard deviation of $5,2$.

1) What is the value of k such that $P(\{\omega : \xi(\omega) > k\}) = 0,63$?

a) $59,984$, b) $23,4562$, c) $40,5678$, d) $90,5678$;

2) What is the value of k such that $P(\{\omega : 59 < \xi(\omega) < k\}) = 0,54$

a) $66,9$, b) $25,4$, c) $20,2$, d) $50,1$;

8.9. The number of orders that come into a mail-order sales office each month is normally distributed with with a mean of 298 and a standard deviation of $15,4$.

1) What is the probability that in a particular month the office receives more than 310 orders?

a) $0,2177$, b) $0,4562$, c) $0,5678$, d) $0,5678$;

2) The probability is $0,3$ that the sales office receives less than how many orders?

a) $290,0$, b) $125,4$, c) $220,2$, d) $250,1$;

8.10. Investment A has an expected return of 8% with a standard deviation of $2,5\%$. Investment B has an expected return of 6% with a standard deviation of $1,2\%$. Assume that you invest equally in both investments and that that the rates of return are independent.

1) What is the expected return of your portfolio?

a) 6% , b) 7% , c) 8%, d) 9% ;

2) What is the standard deviation of the return on your portfolio? Assume that the returns on the two investments are independent.

a) $2,77$, b) $5,45$, c) $2,21$, d) $2,15$;

8.11. The length of time it takes to be seated at a local restaurant on Friday night is normally distributed with a mean of 15 minutes and a standard deviation of 4.75 minutes.

1) What is the probability that you have to wait more 20 minutes to be seated?

a) $0,1469$, b) $0,3669$, c) $0,4691$, d) $0,9891$;

2) What is the probability that you have to wait between 13 and 16 miutes to be seated?

a) $0,1469$, b) $0,3669$, c) $0,4691$, d) $0,9891$;

8.12. Let the random variable ξ follow a standard normal distribution. Find $P(\{\omega : 0 < \xi(\omega) < 0,57\})$?

a) $0,2157$, b) $0,4562$, c) $0,5678$, d) $0,5678$;

8.13. Let the random variable ξ follow a standard normal distribution. Find $P(\{\omega : -2,21 < \xi(\omega) < 0\})$?

a) $0,4864$, b) $0,4562$, c) $0,5678$, d) $0,5178$;

8.13. Let the random variable ξ follow a standard normal distribution. Find $P(\{\omega : -1,33 < \xi(\omega) < 0,78\})$?

a) $0,6905$, b) $0,6562$, c) $0,6678$, d) $0,6178$;

8.14. Let the random variable ξ follow a standard normal distribution. Find the value k such that $P(\{\omega : \xi(\omega) > k\}) = 0,73$?

a) $-0,61$, b) $-0,65$, c) $0,66$, d) $0,61$;

Chapter 9

Correlation Coefficient

Let $(\Omega, \mathcal{F}, P)$ be a probability space, and let ξ and η be such random variables that
$0 < D\xi < \infty$ and $0 < D\eta < \infty$.

Definition 9.1 Numerical value $\rho(\xi,\eta)$, defined by

$$\rho(\xi,\eta) = \frac{M(\xi - M\xi)(\eta - M\eta)}{\sqrt{D\xi}\sqrt{D\eta}},$$

is called a correlation coefficient between random values ξ and η.

Definition 9.2 Numerical value $\mathrm{cov}(\xi,\eta)$, defined by

$$\mathrm{cov}(\xi,\eta) = M(\xi - M\xi)(\eta - M\eta),$$

is called a covariation coefficient between random variables ξ and η.

Remark 9.1 Let $(x_1, y_1), \cdots, (x_n, y_n)$ be the results of observations on the random vector $(X, ; Y)$, every component of which has a finite mathematical expectation and a finite variance. Then:

1) CORREL$(x_1 : x_n; y_1 : y_n)$ calculates the value $\rho_n(X,Y)$, which is a good estimation of correlation coefficient $\rho(X,Y)$.

2) COVAR$(x_1 : x_n; y_1 : y_n)$ calculates a value $\mathrm{cov}_n(X,Y)$, which is a good estimation of covariation coefficient $\mathrm{cov}(X,Y)$.

Below in columns A and B we have entered the results of observations of random vector (X,Y), every component of which has a finite mathematical expectation and a finite variance.

A	*B*
7	2
11	5
6	6
7	7

.

Then:
1) $\rho_4(X,Y) = \text{CORREL}(A_1 : A_4; B_1 : B_4) = -0,0695889$;
2) $\text{cov}_4(X,Y) = \text{COV}(A_1 : A_4; B_1 : B_4) = -0,25$.

We have the following propositions

Theorem 9.1 *Let,* ξ *and* η *be such random variables that* $0 < D\xi < \infty$ *and* $0 < D\eta < \infty$. *Then* $|\rho(\xi,\eta)| \leq 1$.

Proof.

$$0 \leq D(\frac{\xi - M\xi}{\sqrt{D\xi}} \pm \frac{\eta - M\eta}{\sqrt{D\eta}}) = M(\frac{\xi - M\xi}{\sqrt{D\xi}} \pm \frac{\eta - M\eta}{\sqrt{D\eta}})^2 = 2 \pm 2\rho(\xi,\eta),$$

Hence $|\rho(\xi,\eta)| \leq 1$. □

Theorem 9.2 If ξ and η are such independent random variables, that $0 < D\xi < \infty$ and $0 < D\eta < \infty$, then $|\rho(\xi,\eta)| = 0$.

Proof. From the independence of ξ and η we get that random variables $\frac{\xi - M\xi}{\sqrt{D\xi}}$ and $\frac{\eta - M\eta}{\sqrt{D\eta}}$ are also independent. By using Theorem 86(cf. Chapter 8) we deduce that

$$\rho = M\Big[\Big(\frac{\xi - M\xi}{\sqrt{D\xi}}\Big) \cdot \Big(\frac{\eta - M\eta}{\sqrt{D\eta}}\Big)\Big] =$$

$$= M\Big(\frac{\xi - M\xi}{\sqrt{D\xi}}\Big) \cdot M\Big(\frac{\eta - M\eta}{\sqrt{D\eta}}\Big) = 0.$$

□

Example 9.1 Note here that the inverse result given in Theorem 9.2 is not always valid, i. e., the existence of such non-independent random variables ξ and η is possible that $0 < D\xi < \infty$, $0 < D\eta < \infty$ and $\rho(\xi,\eta) = 0$. Indeed, assume

$$(\Omega, \mathcal{F}, P) = ([0;1], \mathcal{B}([0;1]), b_1).$$

Let define random variables ξ and η with the following formulas:

$$\xi(\omega) = 4 \cdot I_{[0,\frac{1}{4}[}(\omega) + 0 \cdot I_{[\frac{1}{4},\frac{1}{2}[}(\omega) - 4 \cdot I_{[\frac{1}{2},\frac{3}{4}[}(\omega) + 0 \cdot I_{[\frac{3}{4},1]}(\omega),$$

$$\eta(\omega) = 0 \cdot I_{[0,\frac{1}{4}[}(\omega) + 4 \cdot I_{[\frac{1}{4},\frac{1}{2}[}(\omega) + 0 \cdot I_{[\frac{1}{2},\frac{3}{4}[}(\omega) - 4 \cdot I_{[\frac{3}{4},1]}(\omega).$$

Note that

$$M\xi = M\eta = 0,\ D\xi = D\eta = 8$$

and

$$\rho(\xi,\eta)=\frac{M(\xi-M\xi)(\eta-M\eta)}{\sqrt{D\xi}\sqrt{D\eta}}=$$

$$=\frac{M\xi\eta}{8}=\frac{M0}{8}=0.$$

Now let show that ξ and η are not independent. Indeed, on the one hand we have

$$P(\{\omega:\xi<3,\eta<3\})=P(\{\omega:\xi(\omega)<3\}\cap\{\omega:\eta(\omega)<3\})=$$

$$P([\frac{1}{4};1]\cap([0;\frac{1}{4}[\cup[\frac{1}{2};1]))=P([\frac{1}{2};1])=\frac{1}{2},$$

On the other hand we have

$$P(\{\omega:\xi<3\})=\frac{3}{4},\ P(\{\omega:\eta<3\})=\frac{3}{4},$$

It follows that

$$P(\{\omega:\xi<3,\eta<3\})=\frac{1}{2}\neq\frac{3}{4}\times\frac{3}{4}=P(\{\omega:\xi<3\})\cdot P(\{\omega:\eta<3\}).$$

Theorem 9.3 *If the conditions of Theorem 9.1 are fulfilled then* $|\rho(\xi,\eta)|=1$ *if and only if there exist real numbers* $a\neq 0$ *and b, such that*

$$P(\{\omega:\eta(\omega)=a\xi(\omega)+b\})=1.$$

Proof. Sufficient. Assume that

$$P(\{\omega:\eta(\omega)=a\xi(\omega)+b\})=1.$$

We set $M\xi=\alpha$ and $\sqrt{D\xi}=\beta$.

Then

$$\rho(\xi,\eta)=M\frac{\xi-\alpha}{\beta}\cdot\frac{\alpha\xi+b-a\alpha-b}{|\alpha|\beta}=sign(a).$$

Necessity. Assume that $\Big|\rho(\xi,\eta)\Big|=1$. Let consider the case when $\rho(\xi,\eta)=1$. Then

$$D\Big(\frac{\xi-M\xi}{\sqrt{D\xi}}-\frac{\eta-M\eta}{\sqrt{D\eta}}\Big)=2(1-\rho(\xi,\eta))=0.$$

Using the property of variance for concrete $c\in R$, we get

$$P\Big(\{\omega:\frac{\xi-M\xi}{\sqrt{D\xi}}-\frac{\eta-M\eta}{\sqrt{D\eta}}=c\}\Big)=1,$$

Hence,

$$P\Big(\{\omega : \xi(\omega) = \frac{\sqrt{D\xi}}{\sqrt{D\eta}} \cdot \eta(\omega) - \sqrt{D\xi}\Big(\frac{M\eta}{\sqrt{D\eta}} - c\Big) + M\xi\}\Big) = 1.$$

If $\rho(\xi,\eta) = -1$, we get

$$D\Big(\frac{\xi - M\xi}{\sqrt{D\xi}} + \frac{\eta - M\eta}{\sqrt{D\eta}}\Big) = 2(1 + \rho(\xi,\eta)) = 0.$$

Analogously, using the property of variance, we deduce an existence of such $d \in R$ that

$$P\Big(\{\omega : \frac{\xi - M\xi}{\sqrt{D\xi}} + \frac{\eta - M\eta}{\sqrt{D\eta}} = d\}\Big) = 1,$$

e.i.

$$P\Big(\{\omega : \xi(\omega) = -\frac{\sqrt{D\xi}}{\sqrt{D\eta}} \cdot \eta(\omega) + \sqrt{D\xi}\frac{M\eta}{\sqrt{D\eta}} + d\sqrt{D\xi} + M\xi\}\Big) = 1.$$

□

Remark 9.2 The correlation coefficient is a quantity characterization of the degree of the dependence between two random variables. It can be considered as cosine of the angle between them. Indeed, since $|\rho(\xi,\eta)| \leq 1$, there exists a unique real number ϕ in interval $[0,\pi]$, that $\cos\phi = \rho(\xi,\eta)$. This number ϕ is called an angle between random variables ξ and η and is denoted with symbol $\widehat{(\xi,\eta)}$, i.e., $\widehat{(\xi,\eta)} = arccos(\rho(\xi,\eta))$.

The following geometrical interpretations of Theorems 9.2-9.3 are interesting:

1) If ξ and η are such independent variables that $0 < D\xi < \infty$ da $0 < D\eta < \infty$, then they are orthogonal , i.e., $\widehat{(\xi,\eta)} = \frac{\pi}{2}$.

2) If $\widehat{(\xi,\eta)}$ is equal to 0 or π, then a random variable $\eta$is presented (P-almost everywhere) as a linear combination of random variable ξ and constant random variable.

Example 9.2 Let consider a transmission system of the signal. Let denote a useful signal with ξ. As here we have hindrances, we receive signal $\eta(\omega) = \alpha\xi(\omega) + \Delta(\omega)$, where α is a coefficient of the intensification , $\Delta(\omega)$ is a hindrance (white noise). Assume that variables Δ and ξ are independent, $M\xi = a, D\xi = 1$ and $M\Delta = 0,\ D\Delta = \sigma^2$. A correlation coefficient between random variables ξ and η is calculated by

$$\rho(\xi,\eta) = M\Big((\xi - a) \cdot \frac{\alpha\xi + \Delta - a\alpha}{\sqrt{\alpha^2 + \sigma^2}}\Big) = \frac{\alpha}{\sqrt{\alpha^2 + \sigma^2}}.$$

If σ is smaller than α and is near at 0, then $\rho(\xi,\eta)$ will be near at 1 and following Theorem 9.3, it is possible to restore ξ by η.

Let consider other numerical characterizations of random variables.

Definition 9.2 A moment of order k $(k \in N)$ of the random variable ξ is defined with $M\xi^k$ and is denoted with symbol α_k , i.e.,

$$\alpha_k = M\xi^k \ (k \in N).$$

Definition 9.3 Value $M(\xi - M\xi)^k$ $(k \in N)$ is called a central moment of order k and is denoted with symbol μ_k, i.e.,

$$\mu_k = M(\xi - M\xi)^k \ (k \in N).$$

Remark 9.2 Note that variance $D\xi$ is the central moment of the order two.

Let $\{\xi_1, \cdots, \xi_n\}$ be a finite sequence of random variables.

Definition 9.4 Value

$$M\xi_1^{k_1} \cdots \xi_n^{k_n}$$

is called a mixed moment of order $k_1 + \cdots + k_n$ and is denoted with symbol $\alpha_{(k_1,\cdots,k_n)}$, i.e.,

$$\alpha_{(k_1,\cdots,k_n)} = M\xi_1^{k_1} \cdots \xi_n^{k_n} \ \ (k_1, \cdots, k_n \in N).$$

Definition 9.5 Value

$$M(\xi_1 - M\xi_1)^{k_1} \cdots (\xi_n - M\xi_n)^{k_n}$$

is called a central moment of order $k_1 + \cdots + k_n$ and is denoted with symbol $\mu_{(k_1,\cdots,k_n)}$, i.e.,

$$\mu_{(k_1,\cdots,k_n)} = M(\xi_1 - M\xi_1)^{k_1} \cdots (\xi_n - M\xi_n)^{k_n} \ (k_1, \cdots, k_n \in N).$$

Definition 9.6 A skewness coefficient of the random variable ξ is called a number $\frac{\mu_3}{\sigma^3}$ and is denoted with symbol A_s, i.e.,

$$A_s = \frac{\mu_3}{\alpha^3}.$$

Remark 9.3 Let $x_1, \cdots, x_n$ be the results of observations on the random variable X. Then the statistical function KURT$(x_1 : x_n)$ gives estimation of the excess of X. For example, KURT$(-1;-3;-80;-80) = -5,990143738$.

Definition 9.7 An excess of the random variable ξ is called a number $\frac{\mu_4}{\alpha^4} - 3$ and is denoted with symbol E_x, i.e.,

$$E_x = \frac{\mu_4}{\alpha^4} - 3.$$

Remark 9.4 Let $x_1, \cdots, x_n$ be the results of observations on the random variable X. Then the statistical function SKEW$(x_1 : x_n)$ gives estimation of the excess of X. For example, SKEW$(1;-1;3;-3;80;80;-80) = -0,17456105$.

Definition 9.8 If F_ξ is a distribution function of ξ, then a median of random variable ξ is called a number γ, for which the following condition is fulfilled

$$F_\xi(\gamma-0) \leq \frac{1}{2} \ , \ F_\xi(\gamma+0) \geq \frac{1}{2},$$

where $F_\xi(\gamma-0)$ and $F_\xi(\gamma+0)$ denote the right and the left limits of function F_ξ at point γ, respectively.

Remark 9.5 Let $x_1,\cdots,x_n$ be the values of the discrete random variable X such that $x_1 < \cdots < x_n$. Then median is x_k+1, when $n=2k+1$, and $\frac{x_k+x_{k+1}}{2}$, when $n=2k$. The statistical function MEDIAN$(x_1 : x_n)$ calculates the median of $x_1,\cdots,x_n$. For example, MEDIAN$(6;7;8;11) = 7,5$ and MEDIAN$(6;7;100) = 7$.

Definition 9.9 A mode of simple discrete random variable X is called its such possible meaning whose corresponding probability is maximal.

Definition 9.10 A mode of absolutely continuous random variable X is called a point of the local maximum of the corresponding density function.

Remark 9.6 Let $x_1,\cdots,x_n$ be the results of observations on the random variable X. Then the statistical function MODE$(x_1 : x_n)$ gives the estimation of the smallest mode of X. For example, MODE$(7;11;6;7;11;18;18) = 7$.

Definition 9.11 A random variable is called unimodular, if it has only one mode. In other cases, the random variable is called polymodular

Tests

9.1. Suppose that $(\Omega,\mathcal{F},P) = ([0,1],\mathcal{B}([0,1]),b_1)$. Assume also that ξ and η are defined with

$$\xi(\omega) = \begin{cases} 0, & \omega \in [\frac{1}{2},\frac{3}{4}[\\ 1, & \omega \in [0,\frac{1}{2}[\\ 2, & \omega \in [\frac{3}{4},1] \end{cases} ,$$

$$\eta(\omega) = \begin{cases} 2, & \omega \in [0,\frac{1}{2}[, \\ -1, & \omega \in [\frac{1}{2},0] \end{cases} .$$

Then the correlation coefficient $\rho(\xi,\eta)$ is equal to

a) $-0,2$, b) $-0,1$, c) 0, d) $0,1$.

9.2. The distribution law of the random variable ξ is given in the table

ξ	-1	0	-1
P	$0,6$	$0,1$	$0,3$

.

Then

1) $M(\xi^3)$ is equal to

a) $-0,1$, b) $-0,2$, c) $-0,3$, d) $-0,4$;

2) $M(\xi - M\xi)^4$ is equal to

a) $1,948$, b) $0,9481$, c) $0,8481$, d) $0,7481$.

9.3. Let ξ be a random variable normally distributed with parameters $(0,1)$. Then

1) α_{2k+1} is equal to

a) 1, b) 0, c) $2k+1$, d) $2k$;

2) μ_2 is equal to

a) 0, b) 1, c) 2, d) 3;

3) γ is equal to

a) 0, b) 1, c) 2, d) 3;

4) mode is equal to

a) 0, b) 1, c) 2, d) 3.

9.4. ξ is a random variable uniformly distributed on $(0,4)$. Then

1) μ_2 is equal to

a) 6, b) 7, c) 8, d) 9;

3) median γ is equal to

a) 1, b) 2, c) 3, d) 4;

4) mode is equal to

a) $[0,4]$, b) $[0,3]$, c) $[0,2]$, d) $[0,1]$.

9.5. The distribution law of simple discrete random variable ξ is given in the following table

ξ	-1	2	3
P	$0,3$	$0,4$	$0,3$

Then

1)the median of ξ is equal to

a) 1, b) 2, c) 3, d) 4;

2) mode is equal to

a) -1, b) 2, c) 3, d) 4.

9.6. Distribution function F_ξ of absolutely continuous random variable ξ is defined with

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

Then

1) median γ is equal to

a) $\frac{\sqrt{2}}{2}$, b) $\frac{\sqrt{3}}{3}$, c) $\frac{\sqrt{5}}{5}$, d) $\frac{\sqrt{7}}{7}$;

2) the mode of ξ is equal to

a) 1, b) 2, c) 3, d) 4.

Chapter 10

Random Vector Distribution Function

Let (Ω, F, P) be the probability space and let $(\xi_k)_{1 \leq k \leq n}$ be a finite family of random variables.

Definition 10.1 A mapping $(\xi_1, \cdots, \xi_n) : \Omega \to R^n$, defined with

$$(\forall \omega)(\omega \in \Omega \to (\xi_1, \cdots, \xi_n)(\omega) = (\xi_1(\omega), \cdots, \xi_n(\omega))),$$

is called n-dimensional random vector.

Definition 10.2 A mapping $F_{\xi_1, \cdots, \xi_n} : R^n \to R$, defined with

$$(\forall (x_1, \cdots, x_n))((x_1, \cdots, x_n) \in R^n \to F_{\xi_1, \cdots, \xi_n}((x_1, \cdots, x_n)) =$$

$$= P(\{\omega : \xi_1 < x_1, \cdots, \xi_n < x_n\})),$$

is called a joint distribution function of the n-dimensional random vector $(\xi_1, \cdots, \xi_n)$.

Definition 10.3 Random vector $(\xi_1, \cdots, \xi_n)$ is called discrete if i-th component ξ_i is a discrete random variable for every i with $1 \leq i \leq n$.

Analogously we can define an absolutely continuous random vector.

The joint distribution function $F_{\xi_1, \cdots, \xi_n}$ has the following properties:

$$1. \lim_{x_i \to \infty,\ 1 \leq i \leq n} F_{\xi_1, \cdots, \xi_n}((x_1, \cdots, x_n)) = 1,$$

$$2. \lim_{x_i \to -\infty,\ 1 \leq i \leq n} F_{\xi_1, \cdots, \xi_n}((x_1, \cdots, x_n)) = 0.$$

Here naturally arises a question what is the probability that the 2-dimensional random vector will obtain the value in the rectangular?

The following result is valid.

Theorem 10.1 *The following formula*

$$(\forall k)(\forall x_k)(\forall y_k)(1 \le k \le 2 \ \& \ x_k \in R \ \& \ y_k \in R \ \& \ x_1 < x_2 \ \& \ y_1 < y_2 \rightarrow$$
$$P(\{\omega : (\xi_1,\xi_2)(\omega) \in [x_1;x_2[\times[y_1;y_2[\}) = F_{\xi_1,\xi_2}((x_2,y_2)) - F_{\xi_1,\xi_2}((x_1,y_2)) +$$
$$+F_{\xi_1,\xi_2}((x_1,y_1)) - F_{\xi_1,\xi_2}((x_2,y_1)),$$

holds, where

$$[x_1;x_2[\times[y_1;y_2[= \{(x,y) | x_1 \le x < x_2, \ y_1 \le y < y_2\}.$$

Proof. Setting

$$A_{(a,b)} = \{\omega : (\xi_1,\xi_2)(\omega) \in]-\infty;a[\times]-\infty;b[\} \ \ (a \in R, b \in R),$$

we get

$$\{\omega : (\xi_1,\xi_2)(\omega) \in]x_1;x_2[\times]x_2;y_2[\} = (A_{(x_2,y_2)} \setminus A_{(x_2,y_1)}) \setminus (A_{(x_1,y_2)} \setminus A_{(x_1,y_1)}).$$

Hence

$$P(\{\omega : (\xi_1,\xi_2)(\omega) \in]x_1;x_2[\times]x_2;y_2[\}) = P((A_{(x_2,y_2)} \setminus A_{(x_2,y_1)})) -$$
$$P((A_{(x_1,y_2)} \setminus A_{(x_1,y_1)})) = (P(A_{(x_2,y_2)}) - P(A_{(x_2,y_1)})) - (P(A_{(x_1,y_2)}) -$$
$$P(A_{(x_1,y_1)})) = P(A_{(x_2,y_2)}) - P(A_{(x_2,y_1)}) - P(A_{(x_1,y_2)}) + P(A_{(x_1,y_1)}) =$$
$$F_{\xi_1,\xi_2}((x_2,y_2)) - F_{\xi_1,\xi_2}((x_2,y_1)) - F_{\xi_1,\xi_2}((x_1,y_2)) + F_{\xi_1,\xi_2}((x_1,y_1)).$$

This ends the proof of theorem. □

Assume that $(x,y) \in R^2$. If there exists double limit

$$\lim_{\Delta x, \Delta y \to 0} \frac{P(\{\omega : (\xi_1,\xi_2)(\omega) \in [x-\Delta x; x+\Delta x[\times]y-\Delta y; y+\Delta y[\})}{4\Delta x \Delta y},$$

then we say that joint distribution function F_{ξ_1,ξ_2} of 2-dimensional random vector (ξ_1,ξ_2) has the density function $f_{\xi_1,\xi_2}(x,y)$ at point (x,y) which is equal to the above-mentioned double limit.

We have the following proposition

Theorem 10.2 *If a function of two variables F_{ξ_1,ξ_2} has the continuous partial derivatives of the first and second orders in any neighborhood of the point (x_0,y_0), then 2-dimensional random vector (ξ_1,ξ_2) has density function $f_{\xi_1,\xi_2}(x_0,y_0)$ at point (x_0,y_0), which can be calculated with the following formula*

$$f_{\xi_1,\xi_2}(x_0,y_0) = \frac{\partial^2 F_{\xi_1,\xi_2}(x_0,y_0)}{\partial x \partial y} = \frac{\partial^2 F_{\xi_1,\xi_2}(x_0,y_0)}{\partial y \partial x}.$$

Proof. Using Theorem 10.1, we get

$$P(\{\omega : (\xi_1,\xi_2)(\omega) \in [x-\Delta x; x+\Delta x[\times]y-\Delta y; y+\Delta y[\}) =$$

$$= F_{\xi_1,\xi_2}((x_0+\Delta x, y_0+\Delta y)) - F_{\xi_1,\xi_2}((x_0+\Delta x, y_0-\Delta y)) -$$
$$F_{\xi_1,\xi_2}((x_0-\Delta x, y_0+\Delta y)) + F_{\xi_1,\xi_2}((x_0-\Delta x, y_0-\Delta y)).$$

Without loss of generality, we can assume that points $(x_0-\Delta x, y_0)$, $(x_0-\Delta x, y_0+\Delta y)$, $(x_0, y_0-\Delta y), (x_0, y_0+\Delta y)$ belong to such neighborhood of point (x_0, y_0) in which F_{ξ_1,ξ_2} has continuous partial derivatives of the first and second orders, respectively. Following Lagrange[1] well known theorem, there exists $\theta_1 \in]0;1[$ such that

$$[F_{\xi_1,\xi_2}((x_0+\Delta x, y_0+\Delta y)) - F_{\xi_1,\xi_2}((x_0+\Delta x, y_0-\Delta y))] -$$

$$[F_{\xi_1,\xi_2}((x_0-\Delta x, y_0+\Delta y)) - F_{\xi_1,\xi_2}((x_0-\Delta x, y_0-\Delta y))] =$$

$$= 2\Delta x \cdot [\frac{\partial F_{\xi_1\xi_2}}{\partial x}(x_0-\Delta x+2\theta_1\Delta x, y_0+\Delta y) - \frac{\partial F_{\xi_1\xi_2}}{\partial x}(x_0-\Delta x+2\theta_1\Delta x, y_0-\Delta y)].$$

Again using the Lagrange theorem, we deduce an existence of $\theta_2 \in]0;1[$ such that

$$P(\{\omega : (\xi_1,\xi_2)(\omega) \in [x-\Delta x; x+\Delta x[\times]y-\Delta y; y+\Delta y[\}) =$$

$$= 4\cdot\Delta x\cdot\Delta y \frac{\partial^2 F_{\xi_1,\xi_2}}{\partial y \partial x}(x_0-\Delta x+2\theta_1\Delta x, y_0-\Delta y+2\theta_2\Delta y).$$

Clearly,

$$\lim_{\Delta x,\Delta y\to 0} \frac{P(\{\omega : (\xi_1,\xi_2)(\omega) \in [x-\Delta x; x+\Delta x[\times]y-\Delta y; y+\Delta y[\})}{4\Delta x\Delta y} =$$

$$= \lim_{\Delta x,\Delta y\to 0} \frac{4\cdot\Delta x\cdot\Delta y \frac{\partial^2 F_{\xi_1,\xi_2}}{\partial y\partial x}(x_0-\Delta x+2\theta_1\Delta x, y_0-\Delta y+2\theta_2\Delta y)}{4\Delta x\Delta y} =$$

$$= \frac{\partial^2 F_{\xi_1,\xi_2}(x_0,y_0)}{\partial y\partial x}.$$

The application of the well-known Schwarz [2] theorem ends the proof of theorem.

□

[1]Lagrange; Joseph Louis (25.1.1736 - 10.4.1813) - French mathematician, the member of Paris Academy of Sciences (1772).

[2]Schwarz; Karl Hermann Amandus (25.1.1843 30.11.1921) German mathematician, the member of Berlin Academy of Sciences (1893).

Example 10.1 2-dimensional random vector (ξ_1, ξ_2) is called distributed by Gaussian law, if its density function f_{ξ_1,ξ_2} has the following form

$$f_{\xi_1,\xi_2}(x_1,x_2) = \frac{1}{2\pi\sigma_1\sigma_2} e^{-\frac{(x_1-a_1)^2}{2\sigma_1^2} - \frac{(x_2-a_2)^2}{2\sigma_2^2}} \quad (x_1, x_2 \in R),$$

where $a_1, a_2 \in R, \sigma_1 > 0, \sigma_2 > 0$.

Here we present some theorems (without proofs).

Theorem 10.3 *Let $D \subseteq R^2$ be some region of R^2 and f_{ξ_1,ξ_2} be a density function of 2-dimensional random vector (ξ_1, ξ_2). Then the following formula*

$$P(\{\omega : (\xi_1, \xi_2)(\omega) \in D\}) = \int\int_D f_{\xi_1,\xi_2}(x,y)dxdy.$$

Definition 10.4 A mapping $g : R^n \to R$ is called measurable, if the following condition

$$(\forall x)(x \in R \to \{(x_1, \cdots, x_n) : g(x_1, \cdots, x_n) < x\} \in B(R^n)).$$

holds. It is easy to show that $g : R^n \to R$ is measurable if and only if when

$$(\forall B)(B \in \mathcal{B}(R) \to g^{-1}(B) \in \mathcal{B}(R^n)),$$

where $g^{-1}(B) = \{(x_1, \cdots, x_n) : g(x_1, \cdots, x_n) \in B\}$.

Theorem 10.4 *Let $f_{\xi_1,\cdots,\xi_n}$ be a density function of random vector $(\xi_1, \cdots, \xi_n)$. Then for arbitrary measurable mapping $g : R^n \to R$ and for arbitrary $B \in B(R)$ we have:*

$$P(\{\omega : g((\xi_1, \cdots, \xi_n)(\omega)) \in B\}) = \int \cdots \int_{g^{-1}(B)} f_{\xi_1,\cdots,\xi_n}(x_1, \cdots, x_n)dx_1 \cdots dx_n.$$

Theorem 10.5 Let $(\xi_k)_{1 \le k \le n}$ be a family of independent random variables and $f_{\xi_1,\cdots,\xi_n}$ be the density function of random vector $(\xi_1, \cdots, \xi_n)$. If f_{ξ_i} $(1 \le i \le n)$ is the density function of ξ_i for $1 \le i \le n$, then

$$f_{\xi_1,\cdots,\xi_n}(x_1, \cdots, x_n) = \prod_{1 \le i \le n} f_{\xi_i}(x_i) \quad ((x_1, \cdots, x_n) \in R^n).$$

Definition 10.5 Let $(\xi_k)_{1 \le k \le n}$ be a family of independent random variables and let ξ_k be normally distributed random variable with parameter (a_k, σ_k^2) for $1 \le k \le n$. Then

$(\xi_1,\cdots,\xi_n)$ is called n-dimensional Gaussian vector and its density function, following Theorem 10.5, has the following form

$$f_{\xi_1,\cdots,\xi_n}(x_1,\cdots,x_n)=\frac{1}{(\sqrt{2\pi})^n\prod_{k=1}^n\sigma_k}e^{-\sum_{k=1}^n\frac{(x_k-a_k)^2}{2\sigma_k^2}},$$

where $(x_1,\cdots,x_n)\in R^n,\ a_1,\cdots,a_n\in R,\sigma_1>0,\cdots,\sigma_n>0.$

n-dimensional Gaussian vector $(\eta_1,\cdots,\eta_n)$ is called standard if

$$f_{\eta_1,\cdots,\eta_n}(x_1,\cdots,x_n)=\frac{1}{(\sqrt{2\pi})^n}e^{-\sum_{k=1}^n\frac{x_k^2}{2}}\ ((x_1,\cdots,x_n)\in R^n).$$

Definition 10.6 Assume that $(\xi_1,\cdots,\xi_n)$ is a Gaussian random vector. Function $P_{\xi_1,\cdots,\xi_n}$ defined with

$$(\forall B)(B\in\mathcal{B}(R^n)\to P_{\xi_1,\cdots,\xi_n}(B)=P(\{\omega:(\xi_1(\omega),\cdots,\xi_n(\omega))\in B\}),$$

is called n-dimensional Gaussian probability measure.

By using Theorem 10.3, we have

$$P_{\xi_1,\cdots,\xi_n}(B)=\overbrace{\int\cdots\int_B}^{n}\frac{1}{(\sqrt{2\pi})^n\prod_{k=1}^n\sigma_k}e^{-\sum_{k=1}^n\frac{(x_k-a_k)^2}{2\sigma_k^2}}dx_1\cdots dx_n.$$

Let consider some examples.

Example 10.3 Let $(\xi_1,\cdots,\xi_n)$ be the n-dimensional Gaussian standard probability measure and $\prod_{k=1}^n[a_k,b_k]\subset R^n$. Then

$$P_{\xi_1,\cdots,\xi_n}(\prod_{k=1}^n[a_k,b_k])=\prod_{k=1}^n[\Phi(b_k)-\Phi(a_k)].$$

Example 10.4 (distribution χ_n^2). Let $(\xi_1,\cdots,\xi_n)$ be the n-dimensional Gaussian standard random vector and V_ρ^n be n-dimensional sphere with radius ρ and with center $O(0,\cdots,0)\in R^n$. Then

$$P_{\xi_1,\cdots,\xi_n}(V_\rho^n)=\overbrace{\int\cdots\int_{V_\rho}}^{n}\frac{1}{(\sqrt{2\pi})^n}e^{-\sum_{k=1}^n\frac{x_k^2}{2}}dx_1\cdots dx_n=$$

$$=\frac{1}{2^{\frac{n}{2}-1}\Gamma(\frac{n}{2})}\times\int_0^\rho r^{n-1}\cdot e^{-\frac{r^2}{2}}dr,$$

where $\Gamma(\cdot)$ is Eulerian integral of the second type. Distribution function $F_{\chi^2_n}$ of random variable $\chi^2_n = \xi_1^2 + \cdots + \xi_n^2$ is called χ^2_n(chi square) -distribution, which has the following form:

$$F_{\chi^2_n}(x) = \begin{cases} 0, & \text{if } x \le 0, \\ \frac{1}{2^{\frac{n}{2}-1}\Gamma(\frac{n}{2})} \times \int_0^{\sqrt{x}} r^{n-1} \cdot e^{-\frac{r^2}{2}} dr, & \text{if } x > 0. \end{cases}$$

Hence

$$f_{\chi^2_n}(x) = \begin{cases} 0, & \text{if } x \le 0, \\ \frac{1}{2^{\frac{n}{2}-1}\Gamma(\frac{n}{2})} \times x^{\frac{n-1}{2}} \cdot e^{-\frac{x}{2}} \frac{1}{2\sqrt{x}}, & \text{if } x > 0 \end{cases} =$$

$$= \begin{cases} 0, & \text{if } x \le 0, \\ \frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} \times x^{\frac{n}{2}-1} \cdot e^{-\frac{x}{2}}, & \text{if } x > 0. \end{cases}$$

Remark 10.1 We have applied the validity of the following fact

$$\overbrace{\int \cdots \int}^{n}_{V_\rho} f(\sqrt{\sum_{k=1}^{n} x_k^2}) dx_1 \cdots dx_n = 2 \cdot \frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2})} \cdot \int_0^{\rho} r^{n-1} f(r) dr,$$

where f is an arbitrary continuous function defined on V_ρ.

Remark 10.2 Let $\xi_1, \cdots, \xi_m$ be an independent family of standard normally distributed real-valued variables. Then CHIDIST$(x;n)$ calculates value

$$P(\{\omega : \omega \in \Omega \ \& \ \chi^2_n(\omega) > x\})$$

for $x \ge 0$. For example,CHIDIST$(2;10) = 0,996340153$.

If we denote by Γ_n a standard n-dimensional Gaussian measure on R^n, then the command $1 -$ CHIDIST$(r^2;n)$ calculates its value on n-dimensional ball $V(r;n)$ with radius r and the center at the zero of R^n. For example, $\Gamma_5(V(2;5)) = 1 -$ CHIDIST$(2^2;5) = 0,450584038$.

Example 10.5 Let $(e_k)_{1 \le k \le m}$ $(m \le n)$ be a family of linearly independent normalized vectors in R^n and let $\xi_1, \cdots, \xi_m$ be the family of one-dimensional independent Gaussian random variables defined on $(\Omega, \mathcal{F}, P)$. Then measure μ , defined with

$$(\forall X)(X \in \mathcal{B}(R^n) \to \mu(X) = P(\{\omega : \sum_{k=1}^{m} \xi_k(\omega) e_k \in X\})),$$

is a Gaussian measure defined on R^n . Note that the converse relation is valid, i.e., for an arbitrary Gaussian measure on R^n we have an analogous representation.

Example 10.6 (Student's distribution t_n). Let $\xi_1, \cdots, \xi_m$ be the independent family of one-dimensional standard Gaussian random variables defined on $(\Omega, \mathcal{F}, P)$ and $G : R^{n+1} \to be$ a measurable function defined with

$$g(x_1, \cdots, x_{n+1}) = \frac{x_{n+1}}{\sqrt{\frac{\sum_{k=1}^{n} x_k^2}{n}}}.$$

The random variable $t_n = g(\xi_1, \cdots, \xi_{n+1})$ is called Students random variable with degree of freedom n. Following Theorem 10.4, we have

$$F_{t_n}(x) = \int_{g^{-1}([-\infty;x))} \frac{1}{(2\pi)^{\frac{n}{2}}} e^{-\sum_{k=1}^{n} x_k^2} dx_1 \cdots dx_n.$$

It can be proved that

$$f_{t_n}(x) = \begin{cases} 0, & \text{if } x \le 0, \\ \frac{1}{\sqrt{nx}} \frac{\Gamma(\frac{n+1}{2}}{\Gamma(\frac{n}{2})} \times (1 + \frac{x^2}{n})^{-\frac{n+1}{2}}, & \text{if } x > 0. \end{cases}$$

It is reasonable to note that $M(t_n) = 0$, when $n > 1$. For variance $D(t_n)$ we have

$$D(t_n) = \begin{cases} \frac{n}{n-1}, & \text{if } n > 2, \\ +\infty, & \text{if } 0 < n \le 2. \end{cases}$$

Remark 10.3 Statistical functions TDIST$(x;n;1)$ and TDIST$(x;n;2)$ calculate the values $P(\{\omega : t_n(\omega) > x\})$ and $P(\{\omega : |t_n(\omega)| > x\})$, respectively. For example, TDIST$(3;4;2) = 0,19970984$ and TDIST$(3;4;1) = 0,039941968$.

Example 10.7 (Fisher's distribution $F_{\xi(k1;k2)}.$). Let $\xi_1, \cdots, \xi_{k_1+k_2}$ be the independent family of one-dimensional standard Gaussian random variables defined on $(\Omega, \mathcal{F}, P)$ and $G : R^{k_1+k_2} \to R$ be a measurable function defined with

$$g(x_1, \cdots, x_{k_1+k_2}) = \frac{\frac{\sum_{i=1}^{k_1} x_i^2}{k_1}}{\frac{\sum_{i=k_1+1}^{k_1+k_2} x_i^2}{k_2}}.$$

The random variable $\xi_{(k_1;k_2)} = g(\xi_1, \cdots, \xi_{k_1+k_2})$ is called Fishers random variable with degrees of freedom k_1 and k_2.

Following Theorem 10.4, we have

$$F_{\xi_{(k_1;k_2)}}(x) = \int_{g^{-1}([-\infty;x))} \frac{1}{(\sqrt{2\pi})^n} e^{-\sum_{k=1}^{n} \frac{x_k^2}{2}} dx_1 \cdots dx_n.$$

It can be proved that

$$f_{\xi_{(k_1;k_2)}}(x) = \begin{cases} 0, & \text{if } x \le 0, \\ 2(\frac{k_1}{k_2})^{\frac{k_1}{k_2}} \frac{\Gamma(\frac{k_1+k_2}{2})}{\Gamma(\frac{k_1}{2})\Gamma(\frac{k_2}{2})} \times x^{k_1-1}(1+\frac{k_1}{k_2}x^2)^{-\frac{k_1+k_2}{2}}, & \text{if } x > 0. \end{cases}$$

Remark 10.4 The statistical function FDIST$(x;k_1;k_2)$ calculates value $P(\{\omega : \xi_{(k_1;k_2)} < x\})$. For example, FDIST$(2;5;6) = 0,211674328$.

We have the following proposition.

Theorem 10.6 *Let* ξ_1 *and* ξ_2 *be the independent random variables with density functions* f_{ξ_1} *and* f_{ξ_2}*, respectively. Then distribution function* $F_{\xi_1+\xi_2}$ *and density function* $f_{\xi_1+\xi_2}$ *of sum* $\xi_1+\xi_2$ *are defined with:*

$$F_{\xi_1+\xi_2}(x) = \int_{-\infty}^{x} dx_2 \int_{-\infty}^{+\infty} f_1(x_1)f_2(x_2-x_1)dx_1,$$

$$f_{\xi_1+\xi_2}(x) = \int_{-\infty}^{+\infty} f_1(x_1)f_2(x_2-x_1)dx_1.$$

Proof. Sum $\xi_1+\xi_2$ can be represented as continuous image g of the random vector (ξ_1,ξ_2), where $g(x_1,x_2) = x_1+x_2$. We set $B = (-\infty, x)$. Using theorems 10.4 and 10.5, we get

$$F_{\xi_1+\xi_2}(x) = P(\{\omega : \xi_1(\omega)+\xi_2(\omega) < x\}) = P(\{\omega : g(\xi_1,\xi_2)(\omega) < x\}) =$$

$$= \int\int_{g^{-1}((-\infty;x))} f_{\xi_1}(x_1)f_{\xi_2}(x_2)dx_1dx_2.$$

Note that

$$g_{-1}((-\infty;x)) = \{(x_1,x_2)|x_1+x_2 < x\},$$

Hence

$$F_{\xi_1+\xi_2}(x) = \int\int_{\{(x_1,x_2)|x_1+x_2<x\}} f_{\xi_1}(x_1)f_{\xi_2}(x_2)dx_1dx_2 =$$

$$= \int_{-\infty}^{+\infty} dx_1 \int_{\infty}^{x-x_1} dx_2 f_{\xi_1}(x_1)f_{\xi_2}(x_2) = \int_{-\infty}^{+\infty} dx_1 \int_{\infty}^{x-x_1} d(x_1+x_2)f_{\xi_1}(x_1)f_{\xi_2}(x_2)$$

$$= \int_{-\infty}^{+\infty} dx_1 \int_{\infty}^{x} d\tau f_{\xi_1}(x_1)f_{\xi_2}(\tau-x_1) = \int_{-\infty}^{x} d\tau \int_{\infty}^{+\infty} f_{\xi_1}(x_1)f_{\xi_2}(\tau-x_1)dx_1 =$$

$$\int_{-\infty}^{x} dx_2 \int_{\infty}^{+\infty} f_{\xi_1}(x_1)f_{\xi_2}(x_2-x_1)dx_1.$$

Clearly, for ℓ_1-almost every x point of R the following equality

$$f_{\xi_1+\xi_2}(x) = \frac{dF_{\xi_1+\xi_2}(x)}{dx} = \int_{-\infty}^{+\infty} f_{\xi_1}(x_1)f_{\xi_2}(\tau-x_1)dx_1.$$

holds. The integral standing in the right in the above equality is called winding of functions f_1 and f_2 and is denoted by $f_{\xi_1} * f_{\xi_2}$. It is not difficult to show that $f_{\xi_1} * f_{\xi_2} = f_{\xi_2} * f_{\xi_1}$, i.e.

$$\int_{-\infty}^{+\infty} f_{\xi_1}(x_1) f_{\xi_2}(x - x_1) dx_1 = \int_{-\infty}^{+\infty} f_{\xi_1}(x - x_1) f_{\xi_2}(x_1) dx_1.$$

□

Tests

10.1. Assume that the distribution of 2-dimensional discrete random vector (ξ_1, ξ_2) is given in the following table

(ξ_1,ξ_2)	$(4,3)$	$(4,10)$	$(4,12)$	$(5,3)$	$(5,12)$
P	$0,17$	$0,13$	$0,25$	$0,2$	$0,25$

.

Then

1) the distribution law of ξ_1 is given in the table

a)

ξ_1	4	5
P	$0,55$	$0,45$

,

b)

ξ_1	4	5
P	$0,45$	$0,55$

;

2) the distribution law of ξ_2 is given in the table

a)

ξ_2	3	10	12
P	$0,37$	$0,13$	$0,5$

,

b)

ξ_2	3	10	12
P	$0,35$	$0,15$	$0,5$

;

3) $F_{\xi_1,\xi_2}(4,5;10,5)$ is equal to

a) $0,36$, b) $0,34$, c) $0,32$, d) $0,3$;

4) $P(\{\omega : (\xi_1(\omega), \xi_2(\omega)) \in [1,5] \times [5,8]\})$ tolia

a) 0, b) $0,38$, c) $0,37$, d) $0,36$.

10.2.Distribution laws of two independent random variables ξ_1 and ξ_2 are given in the following tables

ξ_1	2	3
P	$0,7$	$0,3$

,

ξ_2	-2	2
P	$0,3$	$0,7$

,

respectively. Then the distribution law of $\xi_1 \cdot \xi_2$ is given in the table a)

$\xi_1 \cdot \xi_2$	-6	-4	4	6
P	$0,08$	$0,22$	$0,48$	$0,22$

,

b)

$\xi_1 \cdot \xi_2$	-6	-4	4	6
P	$0,09$	$0,21$	$0,49$	$0,21$

.

10.3. A distribution function of 2-dimensional random vector (ξ_1,ξ_2) is defined with the following formula

$$F_{\xi_1,\xi_2}(x_1,x_2)=\begin{cases}(1-e^{-4x_1})(1-e^{-2x_2}), & x_1\geq 0,\ x_2\geq 0,\\ 0, & x_1<0 \text{ an } x_2<0.\end{cases}$$

Then

a)

$$f_{\xi_1,\xi_2}(x_1,x_2)=\begin{cases}(8e^{-4x_1-2x_2}), & x_1\geq 0,\ x_2\geq 0,\\ 0, & x_1<0 \text{ an } x_2<0,\end{cases}$$

b)

$$f_{\xi_1,\xi_2}(x_1,x_2)=\begin{cases}(6e^{-4x_1-2x_2}), & x_1\geq 0,\ x_2\geq 0,\\ 0, & x_1<0 \text{ an } x_2<0.\end{cases}$$

10.4. The density function of 2-dimensional random vector (ξ_1,ξ_2) is defined with

$$f_{\xi_1,\xi_2}(x_1,x_2)=\frac{20}{\pi^2(16+x_1^2)(25+x_2^2)}\quad ((x_1,x_2)\in R^2).$$

Then

a)

$$F_{\xi_1,\xi_2}(x_1,x_2)=(\frac{1}{2}+\frac{1}{\pi}arctg(\frac{x_1}{8})(\frac{1}{2}+\frac{1}{\pi}arctg(\frac{x_2}{10})),$$

b)

$$F_{\xi_1,\xi_2}(x_1,x_2)=(\frac{1}{2}+\frac{1}{\pi}arctg(\frac{x_1}{4})(\frac{1}{2}+\frac{1}{\pi}arctg(\frac{x_2}{5})).$$

10.5. It is known that coefficients of the general solution of differential equation $y''+5y'+6y=0$ are independent random variables uniformly distributed in interval $(0,1)$. The probability that a general solution of the differential equation will get value $\geq 0,5$ at point $x=0$, is equal to

a) $0,5$, b) $0,875$, c) $0,6$, d) $0,75$.

10.6. It is known that coefficients of the general solution of the differential equation $y''+y=0$ are independent random variables normally distributed with parameters $(0,1)$. Then the probability that the general solution y satisfies the following conditions

$$y(0)\in(0,2)\ \&\ y(-\frac{\pi}{2})\in(-2,1)$$

is equal to

a) $0,2245785$, b) $0,7767678$, c) $0,3905882$, d) $0,8598760$.

10.5. It is known that coefficients of the general solution of the differential equation $y'' - \ln 6 y' + \ln 2 \ln 3 y = 0$ are independent random variables uniformly distributed on the interval $(0,1)$. The probability that general solution y satisfies the following conditions

$$y(0) \in (-\infty, 1) \ \& \ y(-1) \in (-\infty, 2)$$

is equal to

a) $\frac{1}{2}$, b) $\frac{1}{3}$, c) $\frac{1}{4}$, d) $\frac{1}{5}$.

Chapter 11

Chebishev's inequalities

Let (Ω, F, P) be a probability space. The following proposition is valid.

Theorem 11.1 (Chebishev's [1] inequality). *For arbitrary non-negative random variable and for arbitrary positive real number the following inequality*

$$P(\{\omega : \xi(\omega) \geq \varepsilon\}) \leq \frac{M\varepsilon}{\varepsilon}.$$

holds.

Proof. Clearly

$$M\xi = M(\xi \cdot I_{\Omega}) = M(\xi \cdot I_{\{\omega:\xi(\omega)\geq\varepsilon\}} + \xi \cdot I_{\{\omega:\xi(\omega)<\varepsilon\}}) \geq$$

$$\geq M(\xi \cdot I_{\{\omega:\xi(\omega)\geq\varepsilon\}}) \geq \varepsilon \cdot P(\{\omega : \xi(\omega) \geq \varepsilon\}).$$

Finally, we get

$$P(\{\omega : \xi(\omega) \geq \varepsilon\}) \leq \frac{M\varepsilon}{\varepsilon}.$$

This ends the proof of theorem. □

Theorem 11.2 (Chebishev's II inequality). *For arbitrary random variable* η *and for arbitrary positive number* $\delta > 0$ *the following inequality*

$$P(\{\omega : |\eta(\omega) - M\eta| \geq \sigma\}) \leq \frac{D\eta}{\sigma^2}.$$

holds.

[1] P:Chebishev [4(16).5.1821. - 26.11.(8.12)1894] - Russian mathematician, Academician of Petersburg Academy of Sciences (1856), of Berlin Academy of Sciences (1871) and of Paris Academy of Sciences (1874).

Proof. We set : $\xi(\omega) = (\eta(\omega) - M\eta)^2$, $\varepsilon = \sigma^2$. Following Chebishevs I inequality, we get

$$P(\{\omega : (\eta(\omega) - M\eta)^2 \geq \sigma^2\}) \leq \frac{M(\eta - M\eta)^2}{\sigma^2}.$$

Note that

$$\{\omega : (\eta(\omega) - M\eta)^2 \geq \sigma^2\} = \{\omega : |\eta(\omega) - M\eta| \geq \sigma\}.$$

Finally, we get

$$P(\{\omega : |\eta(\omega) - M\eta| \geq \sigma\}) \leq \frac{D\eta}{\sigma^2}.$$

Proof. Example 11.1 Assume that we survey the moon and measure its diameter. Assume also that the results of survey are independent random variables $\xi_1, \cdots, \xi_n$. Assume that a is the value of moons diameter. Then $|\xi_k(\omega) - a|$ will be mistake in the k-th experiment $(1 \leq k \leq n)$. The value $\sqrt{M(\xi_k - a)^2} = \sqrt{D\xi_k}$ will be error mean square deviation. Assume also that the following conditions

$a)$ $M\xi_k = a$,

$b)$ $\sqrt{D\xi_k} = 1$,

$c)$ $(\xi_k)_{1 \leq k \leq n}$ are independent.

It is natural that value $J_n = \frac{1}{n}(\xi_1 + \cdots + \xi_n)$ may be considered as an estimation of parameter a. There naturally arises the following problem: Haw many measures are sufficient to establish the validity of the following stochastic inequality

$$P(\{\omega : |J_n(\omega) - a| \leq 0,01\}) \geq 0,95,$$

?

Clearly, on the one hand, we have

$$P(\{\omega : |J_n(\omega) - a| > 0,01\}) \leq 0,05.$$

On the other hand, we have

$$P(\{\omega : |J_n(\omega) - a| > 0,01\}) \leq \frac{D(J_n)}{(0,1)^2} =$$

$$= \frac{\frac{1}{n^2}\sum_{k=1}^{n} D\xi_k}{0,01} = \frac{\frac{1}{n^2}n}{0,01} = \frac{100}{n}.$$

From the latter inequality we deduce that the smallest natural number $n = n_{\mathrm{C}}$ for which inequality $\frac{100}{n_{\mathrm{C}}} \leq 0,05$, holds, is equal to 2000. Hence, we get

$$P(\{\omega : |J_{2000}(\omega) - a| \leq 0,1\}) \geq 0,95.$$

Theorem 11.3 (The law of three σ). *For arbitrary random variable* ξ *the following inequality*

$$P(\{\omega : |\xi(\omega) - M\xi| \geq 3\sigma\}) \leq \frac{1}{9}.$$

Proof. Indeed, using Chebishevs II inequality, we obtain

$$P(\{\omega : |\xi(\omega) - M\xi| \geq 3\sigma\}) \leq \frac{D\xi}{9\sigma^2} = \frac{1}{9}.$$

□

Tests

11.1. It is known that $D\xi = 0,001$. Using Chebishevs inequality the probability of event $\{\omega : |\xi(\omega) - M\xi| < 0,1\}$ is estimated from bellow by the number, which is equal to

a) $0,8$, b) $0,9$, c) $0,98$, d) $0,89$.

11.2. We have $D\xi = 0,004$. It is established that $P(\{\omega : |\xi(\omega) - M\xi| < \varepsilon\}) \geq 0,9$; Then ε is equal to

a) $0,1$, b) $0,2$, c) $0,3$, d) $0,4$.

11.3. The distribution law of random variable ξ has the following form

ξ	$0,3$	$0,6$
P	$0,2$	$0,8$

.

Using Chebishevs inequality the probability of event $\{\omega : |\xi(\omega) - M\xi| < \varepsilon\}$ is estimated from the below with the following number

a) $0,86$, b) $0,87$, c) $0,88$, d) $0,89$.

11.4. Mean consumption of water in populated area per one day is 50000 liters. Using Chebishevs inequality estimate from below the probability that in this area water consumption per one concrete day will be 150000 liters.

a) $\frac{1}{3}$, b) $\frac{2}{3}$, c) $\frac{1}{4}$, d) $\frac{1}{2}$.

11.5. The probability that an event A occurred in separate experiment is equal to $0,7$. Let denote with ν_n a fraction the numerator of which is equal to the occurred number of event A in n independent experiments, and the denominator of which is equal to n. Minimal natural number n, such that $P(\{\omega : |\nu_n(\omega) - p| < 0,06\} \geq 0,78$ is equal to

a) 327, b) 427, c) 527, d) 627.

11.6. Assume we throw a dice 1200 times. Let ξ denote the number of experiments when number 1 has been thrown. Use Chebishevs inequality for estimation from below of the probability of event $|\omega : \xi(\omega) \leq 800\}$

a) $0,74$, b) $0,75$, c) $0,76$, d) $0,77$.

11.7.Assume that we throw a dice 10000 times. Use Chebishevs inequality to estimate from below the probability that the relative frequency of event- Number 6 is thrown by us would be deviated from number $\frac{1}{6}$ with probability $0,01$

a) $0,84$, b) $0,85$, c) $0,86$, d) $0,87$.

11.8. Assume that we shot the gun 600 times and the probability of hitting the target in a separate experiment is equal to $0,6$. Use the Chebishevs inequality for estimation from the below of the probability that the number of successful shots will be deviated from number 360 by no more than 20.

a) $0,63$, b) $0,64$, c) $0,65$, d) $0,66$.

11.9.It is known that the mean weight of a bun is 50 grams. Use the Chebishevs inequality for estimation from below of the probability that the weight of randomly chosen bun will be ≤ 90 gram

a) $\frac{1}{3}$, b) $\frac{4}{9}$, c) $\frac{5}{9}$, d) $\frac{2}{3}$.

11.10.Use the Chebishevs inequality for estimation from below of the probability that the mean speed of a projectile, accidently shot from a gun is 800 $\frac{km}{sec}$ relative to the hypothesis that the mean speed of the projectile is equal to 500 $\frac{km}{sec}$.

a) $\frac{3}{7}$, b) $\frac{3}{8}$, c) $\frac{1}{3}$, d) $\frac{3}{10}$.

Chapter 12

Limit theorems

Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $(X_k)_{k \in N}$ be an infinite sequence of random variables.

Definition 12.1. We say that a sequence of random variables $(X_k)_{k \in N}$ converges to number $a \in R$ in the sense of probability if for arbitrary positive number $\varepsilon > 0$ the following condition

$$\lim_{k \to \infty} P(\{\omega : |X_k(\omega) - a| < \varepsilon\}) = 1.$$

holds.

This fact is denoted with $\lim_{k \to \infty} X_k \stackrel{p}{=} a$.

We have the following proposition

Theorem 12.1 (Chebishev). *Assume that mathematical variances of the random variables* X_k $(k \in N)$ *are jointly bounded, i.e.,*

$$(\exists c)(c \in R \to (\forall n)(n \in N \to DX_n < c)).$$

Then

$$\lim_{n \to \infty} \frac{1}{n}(\sum_{k=1}^{n} X_k - \sum_{k=1}^{n} MX_k) \stackrel{p}{=} 0.$$

Proof. Following Definition 12.1, it is necessary and sufficient to show the validity of the following condition

$$(\forall \varepsilon)(\varepsilon > 0 \to \lim_{n \to \infty} P(\{\omega : |\frac{1}{n}(\sum_{k=1}^{n} X_k(\omega) - \sum_{k=1}^{n} MX_k) - 0| < \varepsilon\}) = 1).$$

Setting

$$Y_n(\omega) = \frac{1}{n}(\sum_{k=1}^{n} X_k(\omega) - \sum_{k=1}^{n} MX_k),$$

we get

$$MY_n = M\frac{1}{n}(\sum_{k=1}^{n} X_k - \sum_{k=1}^{n} MX_k) = \frac{1}{n}\sum_{k=1}^{n} MX_k - \frac{1}{n}\sum_{k=1}^{n} MX_k = 0,$$

$$DY_n = D(\frac{1}{n}(\sum_{k=1}^{n} X_k - \sum_{k=1}^{n} MX_k)) = D(\frac{1}{n}\sum_{k=1}^{n} X_k) = \frac{1}{n^2}D(\sum_{k=1}^{n} X_k) \leq \frac{nc}{n^2} = \frac{c}{n}.$$

Using Chebishev's II inequality, we get

$$P(\{\omega : |Y_n(\omega) - MY_n| < \varepsilon\}) \geq 1 - \frac{DY_n}{\varepsilon^2} \geq 1 - \frac{c}{n\varepsilon^2}.$$

Hence,

$$P(\{\omega : |Y_n(\omega) - MY_n| < \varepsilon\}) \geq 1,$$

e.i.

$$\lim_{n\to\infty} \frac{1}{n}(\sum_{k=1}^{n} X_k - \sum_{k=1}^{n} MX_k) \stackrel{p}{=} 0.$$

This ends the proof of theorem. □

As corollary of Theorem 12.1, we get

Theorem 12.2 (Bernoulli). *Let* $(Z_k)_{k\in N}$ *be a sequence of independent simple discrete random variables distributed by Bernoulli law with parameter* p*. Then*

$$\lim_{n\to\infty}(\frac{1}{n}\sum_{k=1}^{n} Z_k) \stackrel{p}{=} p.$$

Proof. The sequence of random variables $(Z_k)_{k\in N}$ satisfies the conditions of Theorem 12.1. Hence,

$$\lim_{n\to\infty} \frac{1}{n}(\sum_{k=1}^{n} Z_k - \sum_{k=1}^{n} MZ_k) \stackrel{p}{=} 0.$$

But

$$\lim_{n\to\infty} \frac{1}{n}(\sum_{k=1}^{n} MZ_k) \stackrel{p}{=} p,$$

$$\lim_{n\to\infty}(\frac{1}{n}\sum_{k=1}^{n} Z_k - p) \stackrel{p}{=} 0,$$

which is equivalent to the following condition

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} Z_k \stackrel{p}{=} p.$$

This ends the proof of Theorem □

Theorem 12.3 *If f is a continuous real-valued function defined on $[0,1]$, then the sequence of random variables $(Mf(\frac{1}{n}(\sum_{k=1}^{n} Z_k))_{n\in N}$ is uniformly converged to function $f(p)$ in interval $[0,1]$, where $(Z_k)_{k\in N}$ is the sequence of independent random variables distributed by the Bernoulli law with parameter p.*

Proof. For arbitrary $\varepsilon > 0$, we have

$$M|f(\frac{1}{n}(\sum_{k=1}^{n} Z_k)) - f(p)| \leq M(|f(\frac{1}{n}\sum_{k=1}^{n} Z_k) - f(p)| \cdot I_{\{\omega:|f(\frac{1}{n}\sum_{k=1}^{n} Z_k)-f(p)|\leq\varepsilon\}} +$$

$$+M(|f(\frac{1}{n}\sum_{k=1}^{n} Z_k) - f(p)| \cdot I_{\{\omega:|f(\frac{1}{n}\sum_{k=1}^{n} Z_k)-f(p)|>\varepsilon\}} \leq \sup_{|x|\leq\varepsilon} |f(p+x) - f(p)| + o(n).$$

This ends the proof of Theorem 12.3.

□

Remark 12.1 If f is a continuous real-valued function defined on $[0,1]$, then

$$\lim_{n\to\infty}(\sum_{k=0}^{n} f(\frac{k}{n})C_n^k x^k(1-x)^{n-k}) = f(x)$$

for $x \in [0,1]$; Note here that the above-mentioned convergence is uniform on $[0,1]$. The later relation is a different entry of the uniform convergence of sequence

$$(Mf(\frac{1}{n}(\sum_{k=1}^{n} Z_k)))_{n\in N} = (\sum_{k=0}^{n} f(\frac{k}{n})C_n^k p^k(1-p)^{n-k})_{n\in N}$$

to function f with respect to p on interval $[0,1]$. From this fact we get the well known Weierstrass [1]theorem about approximation of the continuous real-valued function by polinomials. Note here also that these polinomials have the following form

$$\sum_{k=0}^{n} f(\frac{k}{n})C_n^k x^k(1-x)^{n-k} \quad (n \in N).$$

These polinomials are called the Bershtein [2] polinomials.

As corollary of Theorem 12.1 we get the following proposition

Theorem 12.4 (The law of large numbers). *Let $(X_k)_{k\in N}$ be a sequence of identically distributed random variables. Assume also that $MX_k = a$ and $DX_k = \sigma^2 < \infty$; Then an arithmetic mean of random variables converges in probability sense to number a, i. e. ,*

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} X_k \stackrel{p}{=} a.$$

[1]Weierstrass; Karl Theodor Wilhelm (31.10.1815 - 19.2.1897) - German mathematician; Academician of Petersburg Academy of Sciences(1864); Professor of Berlin University (1856).

[2]Bershtein; S (22.2(5,3).1880 - 26.10.1968) - Russian mathematician; Academician of the Ukrainian Academy of Sciences(1925) and academician of the USSR Academy of Sciences (1929).

Proof. Since the sequence of random variables $(X_k)_{k\in N}$ satisfies the conditions of Theorem 12.1, we get

$$\lim_{n\to\infty}\frac{1}{n}(\sum_{k=1}^{n}X_k-\sum_{k=1}^{n}MX_k)\overset{p}{=}0.$$

Clearly, $\frac{1}{n}\sum_{k=1}^{n}MX_k=\frac{na}{n}=a$. Note also that

$$\lim_{n\to\infty}\frac{1}{n}(\sum_{k=1}^{n}X_k-\sum_{k=1}^{n}MX_k)\overset{p}{=}\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}X_k\overset{p}{=}0.$$

The last equality is equivalent to the following equality

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}X_k\overset{p}{=}0.$$

This ends the proof of theorem.

□

Remark 12.2. If $(X_k)_{k\in N}$ is a sequence of independent random variables normally distributed with parameters $(0,\sigma^2)$, then

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}X_k^2\overset{p}{=}\sigma^2.$$

Remark 12.3 Assume that the probability of occurring of event A in each experiment is

equal to p. Let ν_n denote a relative frequency of the event A in n independent experiments. Using the law of Large numbers it is not difficult to show that for arbitrary positive number ε the following condition

$$\lim_{n\to\infty}P(\{\omega:|\nu_n(\omega)-p|<\varepsilon\})=1,$$

holds, i.e.,

$$\lim_{n\to n}\nu_n\overset{p}{=}p.$$

Tests

12.1. Let $(\xi_k)_{k\in N}$ be a sequence of independent random variables uniformly distributed on (a,b) .

1) Then

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}\xi_k\overset{p}{=}A,$$

where A is equal to

a) $\frac{a+b}{2}$, b) $\frac{b-a}{2}$, c) $\frac{a+b}{3}$, d) $\frac{b-a}{3}$.

2) Then

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}\xi_k^2 \overset{p}{=} B,$$

where B is equal to

a) $\frac{(a+b)^2}{2}$, b) $\frac{a^2+ab+b^2}{3}$, c) $\frac{(a+b)^3}{3}$, d) $\frac{(b-a)}{12}$.

12.2. Let $(\xi_k)_{k\in N}$ be a sequence of independent Poisson random variables with parameter $\lambda = 5$ .Then

1) Then

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}\xi_k \overset{p}{=} A,$$

where A is equal to

a) 3, b) 4, c) 5, d) 6;

2) maSin

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}\xi_k^2 \overset{p}{=} B,$$

where B is equal to

a) 28, b) 29, c) 30, d) 31.

12.3. Let $(\xi_k)_{k\in N}$ be a sequence of independent Bernoulli random variables with parameter p . Then for arbitrary non-zero real number s we have

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}\xi_k^s \overset{p}{=} A,$$

where A is equal to

a) p, b) pq, c) p^s, d) q^s.

12.4. Let $(\xi_k)_{k\in N}$ be a sequence of independent Cantors random variables. Then

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}\xi_k \overset{p}{=} A,$$

where A is equal to

a) $0,3$, b) $0,5$, c) $0,6$, d) $0,7$.

12.5. Let $(\xi_k)_{k\in N}$ be a sequence of independent geometric random variables with parameter $q = 0,3$. Then

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}\xi_k \overset{p}{=} A,$$

where A is equal to

a) $\frac{29}{49}$, b) $\frac{30}{49}$, c) $\frac{31}{49}$, d) $\frac{32}{49}$.

12.6. The sequence of functions $(\sum_{k=0}^{n}(\frac{k}{n})^3 C_n^k x^k (1-x)^{n-k})_{n\in N}$ is uniformly converged to function f in the interval $[0,1]$, where $f(x)$ is equal to

a) x^2; b) x^3; c) x^4; d) x^5;

12.7. The sequence of functions $(\sum_{k=0}^{n}\sin((\frac{k}{n})^2) C_n^k x^k (1-x)^{n-k})_{n\in N}$ is uniformly converged to function f in the interval $[0,1]$, where $f(x)$ is equal to

a) $\sin(x^2)$, b) $\sin(x^3)$, c) $\sin(x^4)$, d) $\sin(x^4)$.

12.8. Let $(\xi_n)_{n\in N}$ be a sequence of independent random variables with identical distribution functions. The distribution law of ξ_n is given in the following table

ξ_n	$-\sqrt{n+1}$	0	$\sqrt{n+1}$
P	$\frac{1}{n+1}$	$1-\frac{2}{n+1}$	$\frac{1}{n+1}$

.

Then the application of the Chebishev theorem with respect to the above-mentioned sequence

12.9. Let $(\xi_k)_{k\in N}$ be a sequence of independent Poisson random variables with parameter k. Then the application of the Chebishev theorem with respect to the above-mentioned sequence

12.10. Let $(\xi_k)_{k\in N}$ be a sequence of independent random variables and k be uniformly distributed on $[0;\sqrt{k}]$ for $k \in N$. Then an application of Chebishev theorem with respect to the above-mentioned sequence

Chapter 13

The Method of Characteristic Functions and its applications

Definition 13.1. Let (Ω, F, P) be a probability space. A characteristic function of random variable $\xi : \Omega \to R$ is called the mathematical expectation of complex function $e^{it\xi} = \cos(t\xi) + i\sin(t\xi)$ and is denoted with Φ_ξ, i. e.,

$$\Phi_\xi(t) = Me^{it\xi} \quad (t \in R).$$

Let ξ be a discrete random variable, i. e.,

$$\xi(\omega) = \sum_{k \in N} x_k I_{A_k}(\omega) \quad (\omega \in \Omega),$$

where $(A_k)_{k \in N}$ is a family of pairwise disjoint events covered Ω and $(x_k)_{k \in N}$ be a sequence of real numbers. In this situation, we have

$$\Phi_\xi(t) = Me^{it\xi} = \sum_{k \in n} e^{itx_k} P(A_k) \quad (t \in R).$$

When f_ξ is the density function of absolutely continuous random variable ξ, then we get

$$\Phi_\xi(t) = \int_{-\infty}^{+\infty} e^{itx} f_\xi(x)dx \quad (t \in R).$$

From the last relation we see that $\Phi_\xi(t)$ is Fourier transformation of f_ξ. From the course of mathematical analysis it is well known that if we have the Fourier [1] transformation $\Phi_\xi(t)$

[1] Fourier; Jean Baptiste Joseph (1.3. 1768-16. 5. 1830)-French mathematician, the member of Paris Academy of Sciences (1817), the member of Petersburg Academy of Sciences (1829).

of function f_ξ then in some situations we can restore function f_ξ with function $\Phi_\xi(t)$. In particular,

$$f_\xi(x) = \int_{-\infty}^{+\infty} e^{-itx}\Phi_\xi(t)dt \ (x \in R).$$

The above mentioned relation is called Fourier inverse transformation .

Let consider some properties of characteristic function.

Theorem 13.1 *For arbitrary random variable* $\xi : \Omega \to R$ *we have*

$$\Phi_\xi(0) = 1.$$

Proof. Since $\Phi_\xi(t) = Me^{it\xi}$ $(t \in R)$, we have

$$\Phi_\xi(0) = M\ 1 = 1.$$

□

Theorem 13.2. *For every random variable* ξ *the following condition*

$$(\forall t)(t \in R \to |\Phi_\xi(t)| \leq 1).$$

holds. Note that for every random variable η the following condition

$$|M\eta| \leq M|\eta|.$$

holds. Hence,

$$|\Phi_\xi(t)| = |Me^{it\xi}| \leq M|e^{it\xi}| = M1 = 1.$$

□

Theorem 13.3. *For arbitrary random variable* ξ *we have*

$$\Phi_\xi(-t) = \overline{\Phi_\xi(t)}.$$

Proof.

$$\Phi_\xi(-t) = M(e^{-it\xi}) = M(\cos(-t\xi) + i\sin(-t\xi)) = M(\cos(-t\xi)) + iM(\sin(-t\xi)) =$$

$$= M(\cos(t\xi)) - iM(\sin(t\xi)) = \overline{M(\cos(t\xi)) + iM(\sin(t\xi))} = \overline{Me^{it\xi}} = \overline{\Phi_\xi(t)}.$$

□

The following two facts are presented without proofs.

Theorem 13.4 *Characteristic function* $\Phi_\xi(t)$ *of random variable* ξ *is uniformly continuous on the real axis.*

Theorem 13.5 (Uniqueness Theorem). *The distribution function of the random variable is uniquely defined with its characteristic function.*

Theorem 13.6 *If random variables* ξ *and* η *are linearly related with* $\xi(\omega) = a\eta(\omega) + b$ $(a \in R,\ b \in R, \omega \in \Omega)$, *then*

$$\Phi_\xi(t) = e^{itb}\Phi_\eta(at).$$

Proof. Indeed,

$$\Phi_\xi(t) = \Phi_{a\eta+b}(t) = Me^{i(a\eta+b)t} = Me^{ibt}Me^{ia\eta t} = e^{itb}\Phi_\eta(at).$$

□

Theorem 13.7 *The characteristic function of the sum of two independent random variables is equal to the product of characteristic functions of the corresponding random variables.*

Proof. Let ξ and η be independent random variables. Then complex random variables $e^{it\xi}$ and $e^{it\eta}$ are independent, too. Now using the property of mathematical expectation we get

$$\Phi_{\xi+\eta}(t) = Me^{it(\xi+\eta)} = Me^{it\xi}Me^{it\eta} = \Phi_\xi(t)\cdot\Phi_\eta(t).$$

□

Theorem 13.7 admits the following generalization

Theorem 13.8 If $(\xi_k)_{1\le k\le n}$ *is the finite family of independent random variables, then*

$$\Phi_{\sum_{k=1}^n \xi_k}(t) = \prod_{k=1}^n \Phi_{\xi_k}(t) \ \ (t \in R).$$

Let ξ be a random variable and let $(\xi_k)_{k\in N}$ be a sequence of random variables.

Definition 13.2 The sequence of random variables $(\xi_k)_{k\in N}$ is called weakly converged to random variable ξ if sequence $(F_{\xi_n})_{n\in N}$ is convergent to function F_ξ at its continuity points.

We present one fundamental fact from the probability theory without proof.

Theorem 13.9 *The sequence of random variables* $(\xi_k)_{k\in N}$ *weakly converges to the random variable* ξ *if and only if the sequence of characteristic functions* $(\Phi_{\xi_n})_{n\in N}$ *converges to the characteristic function* Φ_ξ.

Let consider some examples.

Example 13.1 Let ξ be a Binomial random variable with parameters (n,p), i.e.,

$$P(\{\omega:\xi(\omega)=k\})=C_n^k p^k(1-p)^{n-k}\ \ (0\le k\le n).$$

Then

$$\Phi_\xi(t)=Me^{it\xi}=\sum_{k=0}^{n}e^{itk}\cdot C_n^k p^k(1-p)^{n-k}=$$

$$=\sum_{k=0}^{n}C_n^k(e^{it}p)^k(1-p)^{n-k}=[pe^{it}+(1-p)]^n=(pe^{it}+q)^n,\ q=1-p.$$

Example 13.2 Let ξ be a Poisson random variable with parameter λ, i. e.,

$$P(\{\omega:\xi(\omega)=k\})=\frac{\lambda^k}{k!}e^{-\lambda}\ \ (k=0,1,\cdots),$$

Then

$$\Phi_\xi(t)=Me^{it\xi}=\sum_{k=0}^{\infty}\frac{e^{itk}}{k!}e^{-\lambda}=\sum_{k=0}^{\infty}\frac{(e^{it}\lambda)^k}{k!}e^{-\lambda}=$$

$$=e^{-\lambda}\sum_{k=0}^{\infty}\frac{(e^{it}\lambda)^k}{k!}=e^{\lambda e^{it}-\lambda}=e^{\lambda(e^{it}-1)}.$$

Example 13.3 Let ξ be a random variable uniformly distributed in the interval $(a;b)$, i. e.,

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

Then

$$\Phi_\xi(t)=Me^{it\xi}=\int_{-\infty}^{+\infty}e^{itx}f_\xi(x)dx=\int_a^b\frac{e^{itx}}{b-a}dx=\frac{1}{(b-a)it}e^{itx}\Big|_a^b=$$

$$=\frac{1}{(b-a)it}(e^{itb}-e^{ita}).$$

Example 13.4 Let ξ be a normally distributed random variable with parameters (a,σ^2), i. e.,

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

Then

$$\Phi_\xi(t) = Me^{it\xi} = \int_{-\infty}^{+\infty} e^{itx} f_\xi(x)dx =$$

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

Setting $z = \frac{x-a}{\sigma} - it\sigma$, we get

$$\frac{x-a}{\sigma} = z + it\sigma,\ x = a + \sigma z + it\sigma^2,\ dx = \sigma dz.$$

With simple transformation we get

$$\Phi_\xi(t) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty - it\sigma}^{+\infty - it\sigma} e^{it(a + z\sigma + it\sigma^2) - \frac{(z+it\sigma)^2}{2}} \sigma dz =$$

$$= e^{iat - \frac{\sigma^2 t^2}{2}} \frac{1}{\sqrt{2\pi}} \int_{-\infty - it\sigma}^{+\infty - it\sigma} e^{-\frac{z^2}{2}} dz \ (t \in R).$$

Using the well known fact from mathematical analysis asserted that

$$(\forall b)(b \in R \to \int_{-\infty - it\sigma}^{+\infty - it\sigma} e^{-\frac{z^2}{2}} dz = \sqrt{2\pi}),$$

we get

$$\Phi_\xi(t) = e^{iat - \frac{\sigma^2 t^2}{2}} \quad (t \in R).$$

Remark 13.1 Characteristic function Φ_ξ of normally distributed random variable ξ with parameter $(0,1)$ has the following form

$$\Phi_\xi(t) = e^{-\frac{t^2}{2}} \quad (t \in R).$$

Example 13.5 Let ξ be exponential random variable with parameter λ, i.e.,

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

Then

$$\Phi_\xi(t) = Me^{it\xi} = \int_{-\infty}^{+\infty} e^{itx} f_\xi(x)dx = \int_0^{+\infty} e^{itx - \lambda x} dx =$$

$$= \lambda \int_0^{+\infty} (e^{it} e^{-\lambda})^x dx = \lambda \int_0^{+\infty} (e^{it - \lambda})^x dx = \lambda \frac{(e^{it-\lambda})^x}{(it - \lambda)} \Big|_0^{+\infty} = \frac{\lambda}{\lambda - it}.$$

Example 13.6 Let $\xi = c$ be a constant random variable, i.e.,

$$P(\{\omega : \xi(\omega) = c\}) = 1,$$

Then

$$\Phi_{\xi}(t) = Me^{it\xi} = Me^{itc} = e^{itc}.$$

Let consider one application of the method of characteristic functions

Theorem 13.10 (Lindeberg [2] -Levy [3]). *If* $(\xi_k)_{k\in N}$ *be a sequence of independent identically distributed random variables, then the sequence of random variables*

$$\Big(\frac{\sum_{k=1}^{n}\xi_k - M(\sum_{k=1}^{n}\xi_k)}{\sqrt{D\sum_{k=1}^{n}\xi_k}}\Big)_{n\in N}$$

is weakly converged to the standard normally distributed random variable, i.e.,

$$(\forall x)(x\in R \to \lim_{n\to\infty} P\big(\{\omega : \frac{\sum_{k=1}^{n}\xi_k - M(\sum_{k=1}^{n}\xi_k)}{\sqrt{D\sum_{k=1}^{n}\xi_k}} < x\}\big) =$$

$$= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{t^2}{2}}dt\big).$$

Proof. Here we present the proof of this theorem in the case of absolutely continuous random variables. Assume that $m = M\xi_1,\ \sigma = \sqrt{D\xi_1}$. We have

$$\lim_{n\to\infty}\Phi_{\frac{\sum_{k=1}^{n}\xi_k - mn}{\sqrt{n}\sigma}}(t) = \lim_{n\to\infty}\Phi_{\sum_{k=1}^{n}(\frac{\xi_k - m}{\sqrt{\sigma}})\frac{1}{\sqrt{n}}}(t) =$$

$$\lim_{n\to\infty}\Phi_{\sum_{k=1}^{n}(\frac{\xi_k - m}{\sqrt{\sigma}})}(\frac{t}{\sqrt{n}}) = \lim_{n\to\infty}\prod_{i=1}^{n}\Phi_{\frac{\xi_i - m}{\sigma}}(\frac{t}{\sqrt{n}}) = \lim_{n\to\infty} e^{\ln\prod_{i=1}^{n}\Phi_{\frac{\xi_i - m}{\sigma}}(\frac{t}{\sqrt{n}})} =$$

$$= \lim_{n\to\infty} e^{\sum_{i=1}^{n}\ln\Phi_{\frac{\xi_i - m}{\sigma}}(\frac{t}{\sqrt{n}})}.$$

If we denote with $f(t)$ the distribution function of random variable $\frac{\xi_i - m}{\sigma}$, then we get

$$\Phi(t) := \Phi_{\frac{\xi_i - m}{\sigma}}(t) = \int_{-\infty}^{+\infty} e^{itx} f(x)dx,$$

$$\Phi^{'}(t) = \int_{-\infty}^{+\infty} ixe^{itx} f(x)dx,$$

$$\Phi^{''}(t) = \int_{-\infty}^{+\infty} i^2x^2e^{itx} f(x)dx = -\int_{-\infty}^{+\infty} x^2e^{itx} f(x)dx.$$

[2]Lindeberg; J:W:- Finnish mathematician. He was the first who proved Theorem 10 which in literature is known as Central Limiting Theorem.

[3]Levy; Paul Pierre (15.9.1889 - 15.12.1971)-French mathematician, the member of Paris Academy of Sciences (1964). He was the first who applied the method of characteristic functions to prove Central Limiting Theorem.

Note that

$$\Phi(0) = 1,$$
$$\Phi^{'}(0) = iM(\frac{\xi_i - m}{\sigma}) = 0,$$
$$\Phi^{''}(0) = -\int_{-\infty}^{+\infty} x^2 f(x)dx = -1.$$

The Maclaurin [4] formula with the first three members has the following form

$$\Phi(t) = \Phi(0) + \frac{\Phi^{'}(0)}{1!}t + \frac{\Phi^{''}(0)}{2!}t^2 + a(t)t^3,$$

where $\lim_{t\to 0} a(t) = 0$. Hence, we get

$$\Phi(\frac{t}{\sqrt{n}}) = 1 + 0 \cdot \frac{t}{\sqrt{n}} - \frac{t^2}{2n} + a(\frac{t}{\sqrt{n}}) \cdot \frac{t^3}{n\sqrt{n}}.$$

From the course of mathematical analysis it is well known that $\ln(1+o(n)) \approx o(n)$ when $o(n)$ is an infinitely small sequence (i.e., $\lim_{n\to\infty} o(n) = 0$). Finally we get

$$\lim_{n\to\infty} \Phi_{\frac{\sum_{k=1}^n \xi_k - mn}{\sqrt{n}\sigma}}(t) = \lim_{n\to\infty} e^{n\ln(1+0\cdot\frac{t}{\sqrt{n}} - \frac{t^2}{2n} + a(\frac{t}{\sqrt{n}})\cdot\frac{t^3}{n\sqrt{n}})} =$$

$$= \lim_{n\to\infty} e^{n(-\frac{t^2}{2n} + a(\frac{t}{\sqrt{n}})\cdot\frac{t^3}{n\sqrt{n}})} = e^{\lim_{n\to\infty}(-\frac{t^2}{2} + a(\frac{t}{\sqrt{n}})\cdot\frac{t^3}{\sqrt{n}})} = e^{-\frac{t^2}{2}},$$

which ends the proof of theorem. □

Example 13.7 Assume that the following conditions are fulfilled:

1) Let ξ_k be the moons diameter estimation obtained with k-th measure $(k \in N)$;

2) a is the moons diameter;

3) the results of measures ξ_k $(k \in N)$ is a sequence of normally distributed independent random variables with parameters $(a, 1)$.

Using the Chebishev inequality (cf. Chapter 11, Example 11.1) we have proved that $n_C = 2000$ is such smallest natural number for which the following stochastic inequality

$$P(\{\omega : |\frac{1}{n}\sum_{k=1}^{n} \xi_k(\omega) - a| \leq 0,1\}) \geq 0,95.$$

holds. Note here that with the help of Chebishev inequality it is not possible to choose natural number smaller than $n_C = 2000$ which will satisfy the above mentioned inequality. Since $\frac{\sum_{k=1}^n \xi_k - na}{\sqrt{n}}$ is the normally distributed random variable with parameter $(0,1)$, we can calculate the smallest natural number n_C, for which the same inequality holds.

[4] Maclaurin; Colin (1698 - 14.6.1746) - Scottish mathematician.

Indeed, we get

$$P(\{\omega : |\frac{1}{n}\sum_{k=1}^{n}\xi_k(\omega) - a| \leq 0,1\} = P(\{\omega : |\frac{\sum_{k=1}^{n}\xi_k(\omega) - na}{n}| \leq 0,1\} =$$

$$= P(\{\omega : |\frac{\sum_{k=1}^{n}\xi_k(\omega) - na}{\sqrt{n}}| \leq 0,1\sqrt{n}\} = 1 - 2\Phi(-0,1\sqrt{n}).$$

Clearly, we must to choose a such smallest natural number n_L which will be a solution of the following inequality

$$1 - 2\Phi(-0,1\sqrt{n}) \geq 0,95.$$

We have

$$\Phi(-0,1\sqrt{n}) \leq \frac{1-0,95}{2} \Leftrightarrow \Phi(-0,1\sqrt{n}) \leq 0,025 \Leftrightarrow$$

$$-0,1\sqrt{n} \leq \Phi^{-1}(0,025) \Leftrightarrow \sqrt{n} \geq 100(\Phi^{-1}(0,025))^2 \Leftrightarrow$$

$$n \geq 100(1,96)^2 \Leftrightarrow n \geq 383,16 \Leftrightarrow n \geq 384.$$

Finally we deduce that $n_L = 385$, which is a such smallest natural number which is solution of the following inequality

$$P(\{\omega : |\frac{1}{n}\sum_{k=1}^{n}\xi_k(\omega) - a| \leq 0,1\} \geq 0,95.$$

It is clear that natural number $n_L = 385$ is smaller than natural number $n_C = 2000$ obtained with the help of the Chebishev inequality.

Remark 13.2. If the sequence of random variables $(\xi_k)_{k\in N}$ is weakly convergent to random variable ξ, then for sufficiently large natural number n the distribution function F_{ξ_n} of ξ_n can be assumed to be equal to distribution function F_ξ of random variable ξ.

Tests

13.1. Let define sequence of random variables $(\xi_n)_{n\in N}$ with

$$(\forall\omega)(\omega \in \Omega \to \xi_n(\omega) = C - \frac{1}{n}).$$

Then the sequence of random variables $(\xi_n)_{n\in N}$ is weakly convergent to random variable ξ, which is equal (with probability 1) to

a) $c-1$, b) c, c) c^2, d) $c+1$.

13.2. Let ξ_n be the Poisson random variable with parameter $\lambda + o(n)$ for $n \in N$, where $\lambda > 0$, and let $(o(n))_{n\in N}$ be an infinitely small sequence. Then the sequence of random variables $(\xi_n)_{n\in N}$ is weakly convergent to the Poisson random variable with parameter μ, where μ is equal to

a) λ, b) λ^2, c) $\lambda(1+\lambda)$, d) $\lambda^2(1+\lambda)^2$.

13.3. Let ξ_n be a random variable uniformly distributed in interval (a_n,b_n) for $n\in N$. Assume also that $\lim_{n\to\infty} a_n = a$ and $\lim_{n\to\infty} b_n = b$. Then the sequence of random variables $(\xi_n)_{n\in N}$ is weakly convergent to the random variable uniformly distributed on interval (c,d), where (c,d) coincides with

a) (a,b), b) $(\frac{b-a}{2},\frac{a+b}{2})$, c) $(a,\frac{a+b}{2})$, d) $(\frac{a+b}{2},b)$.

13.4. Let $(\xi_n)_{n\in N}$ be an independent sequence of random variables and let ξ_k be a normally distributed random variable with parameters $(\frac{1}{2^k};\frac{1}{2^{2k}})$ for $k\in N$. Then the sequence of random variables $(\sum_{k=1}^n \xi_k)_{n\in N}$ is weakly convergent to the normally distributed random variable with parameters (m,σ^2), where (m,σ^2) is equal to

a) $(1,3)$, b) $(1,4)$, c) $(1,5)$, d) $(1,6)$.

13.5. (The Poisson Theorem). Let $(\xi_n)_{n\in N}$ be the sequence of independent Binomial random variables with parameters (n,p_n). Assume that $\lim_{n\to\infty} n\cdot p_n = \lambda > 0$. Then the sequence of random variables $(\xi_n)_{n\in N}$ is weakly convergent to the Poisson random variable with parameter μ, where μ is equal to

a) $\lambda+1$, b) λ, c) $\lambda-1)$, d) λ^2.

13.6. Let $(\xi_n)_{n\in N}$ be the independent sequence of normally distributed random variables with parameter (a,σ^2). Then the sequence of random variables $(\frac{\xi_1+\cdots+\xi_n}{n})_{n\in N}$ is weakly convergent to constant random variable m, where m is equal to

a) a, b) a^2, c) a^3, d) a^4.

13.7. Let ξ_k be a normally distributed random variable with parameter (m_k,σ_k^2) for $1\le k\le n$. Then sum $\sum_{k=1}^n \xi_k$ is a normally distributed random variable with parameter (m,σ^2), where (m,σ^2) is equal to

a) $(\sum_{k=1}^n m_k, \sum_{k=1}^n \sigma_k)$, b) $(\sum_{k=1}^n m_k, \sum_{k=1}^n \sigma_k^2)$,
c) $(\sum_{k=1}^n m_k, \sum_{k=1}^n \sigma_k^3)$, d) $(\sum_{k=1}^n m_k, \sum_{k=1}^n \sigma_k^4)$.

13.8. If ξ is a normally distributed random variable with parameter (m,σ^2), then random variable $a\xi+b$is distributed normally with parameter (c,d^2), where (c,d^2) is equal to

a) $(b+am,a^2\sigma^2)$, b) $(b+am,a\sigma^2)$,
c) $(b+am,a^2\sigma)$, d) $(b+m,a\sigma^2)$.

13.9. Let $(\xi_k)_{1\le k\le n}$ be an independent sequence of random variables and let ξ_k be a Poisson random variable with parameter λ_k. Then sequence $\sum_{k=1}^n \xi_k$ is a Poisson random variable with parameter μ, where μ is equal to

a) $\sum_{k=1}^n \lambda_k^2$, b) $\sum_{k=1}^n \lambda_k$,
c) $\sum_{k=1}^n (1-\lambda_k)$, d) $\sum_{k=1}^n (1+\lambda_k)$.

13.10. Let $(\xi_k)_{1\le k\le n}$ be an independent sequence of random variables and let ξ_k be a Binomial random variable with parameters (n_k,p). Then $\sum_{k=1}^n \xi_k$ is a binomial random variable with parameters (m,x), where (m,x) is equal to

a) $(\sum_{k=1}^n{}_k, p)$, b) $(\sum_{k=1}^n{}_k, p^2)$,

c) $(\sum_{k=1}^{n} {}^{2}_{k}, p^2)$, d) $(\sum_{k=1}^{n} {}_{k}, p^3)$.

13.11. Let ξ_k be a number of demands of the k-th goods during one day which is a Poisson random variable with parameter $\lambda_k (1 \leq k \leq n)$. Then the probability that the common number of demands of all goods during one day will be equal to 8 relative to hypothesis $m = 10$, $\lambda_1 = \cdots = \lambda_5 = 0,3$, $\lambda_6 = \cdots = \lambda_9 = 0,8$, $\lambda_{10} = 1,3$, is equal to

a) $0,345103$, b) $0,457778$, c) $0,567788$, d) $0,103258$.

13.12. The mean load transported with a lórry on each trip is equal to m = 20. The mean absolute deviation of the above mentioned load is equal to $\sigma = 1$. Then

1) the probability that the weight of the load transmitted during 100 trips will be in interval $[1950; 2000]$, is equal to

a) $0,5$, b) $0,55$, c) $0,555$, d) $0,5555$;

2) the value which is grater with probability 0,95 than the weight of the load transmitted during 100 trips is equal to

a) 20164, b) 20264, c) 20364, d) 20464.

13.13. A mean weight of an apple is $m = 0,2$ kg. A mean of absolute deviation of the weight of accidentally chosen apple is $\sigma = 0,02$ kg. Then

1) the probability that the weight of the accidently chosen 49 apples will be in interval $[9,5; 10]$, is equal to

a) $0,44$; b) $0,88$; c) $0,178$; d) $0,356$;

2) the value which will be smaller then the weight of the accidentally chosen 100 apples with probability 0,95, is equal to

a) $16,672$, b) $17,672$, c) $18,672$, d) $19,672$.

13.14. The probability that túrner will make a standard detail is equal to 0,64. Then the probability that

1) 70 details, accidentally chosen from the complect of 100 details will be standard, is equal to

a) $0,6241$, b) $0,7241$, c) $0,8241$, d) $0,9241$;

2)the number of standard details in the accidentally chosen 100 details will be in interval [50,65], is equal to

a) $0,1108$, b) $0,1308$, c) $0,1508$, d) $0,1708$.

13.15. The factory sent 15000 standard details to the stórehouse. The probability that the detail will dámaged during transportation, is equal to 0,0002. Then the probability that

1)3 dámaged details will be brought to bring at stórehose, is equal to

a) $0,094042$, b) $0,114042$, c) $0,134042$, d) $0,154042$;

2) the number of dámaged details will be in interval [2,4], is equal to

a) $0,414114$, b) $0,515115$, c) $0,616116$, d) $0,717117$.

13.16. Suppose that 19% of all sales are for amounts greater that 1.000 dollars. In a random sample of 30 invoices, what is the probability that more than six of the invoices are for over 1.000 dollars?

a) $0,4443$, b) $0,9562$, c) $0,5678$, d) $0,5678$;

Chapter 14

Markov Chains

Let $(\Omega, \mathcal{F}, P)$ be a probability space. Assume that we have a physical system, which after each step changes its phase position. Assume that the number of possible positions $\varepsilon_1, \varepsilon_2, \cdots$. is finite or countable. Let $\xi_n(\omega)$ be a position of physical system after n steps $(\omega \in \Omega)$. Clearly, the chain of logical transitions

$$\xi_0(\omega) \to \xi_1(\omega) \to \cdots \quad (\omega \in \Omega)$$

depends on the chance factor. Assume that the following regularity condition is preserved: if after n steps the system is in position ε_i , then, independently of its early positions it will pass to position ε_j with probability P_{ij}, i.e.,

$$P_{ij} = P(\{\omega : \xi_{n+1}(\omega) = \varepsilon_j \mid \xi_n(\omega) = \varepsilon_i\}), i, j = 1, 2, \cdots .$$

The above described model is called Markov [1] homogeneous chain. Number P_{ij} is called the transition probability. Besides there is also given also the distributions of initial positions, i.e.,

$$P_i^{(0)} = P(\{\omega : \xi_0(\omega) = \varepsilon_i\}) \;\; i = 1, 2, \cdots$$

Here naturally arises the following problem:

what is the probability that the physic system will be in the position ε_i *after* n *steps ?*

Let denote this probability by $P_j(n)$, i. e.,

$$P_j(n) = P(\{\omega : \xi_n(\omega) = \varepsilon_j\}).$$

Note that after $n-1$ steps the physical system will be in one of the possible positions ε_k $(k = 1, 2, \cdots)$. The probability that the physical system will be in position ε_k is equal to $P_k(n-1)$. The probability that the physical system will occur in position ε_j after n steps if

[1] Markov, A (2(14).1856-20.7.1922) - Russian mathematician, the member of Petersburg Academy of Sciences (1890).

it is known that after $n-1$ steps it was in position ε_k is equal to transition probability P_{kj}. Using total probability formula we get

$$P(\{\omega : \xi_n(\omega) = \varepsilon_j\}) =$$

$$\sum_{k \in N} P(\{\omega : \xi_n(\omega) = \varepsilon_j\} | \{\omega : \xi_{n-1}(\omega) = \varepsilon_k\}) \cdot P(\{\omega : \xi_{n-1}(\omega) = \varepsilon_k\}).$$

The formula gives the following recurrent formula for calculation of the probability $P_j(n)$:

$$P_j(0) = P_j^{(0)}, \; P_j(n) = \sum_{k \in N} P_k(n-1) \cdot P_{kj} \;\; (j, n = 1, 2, \cdots).$$

In this case when the physical system at the initial moment is in position ε_i the initial distribution has the following form

$$P_i^{(0)} = 1, \; P_k^{(0)} = 0, \; k \neq i,$$

and probability $P_j(n)$ coincides with $P_{ij}(n)$, which is equal to transition probability from position ε_i to position ε_j after n steps, i. e.,

$$P_{ij}(n) = P(\{\omega : \xi_n(\omega) = \varepsilon_j | \{\omega : \xi_0(\omega) = \varepsilon_i\}\} \;\; i, j = 1, 2, \cdots .$$

In the case of the following initial distribution $P_i^{(0)} = 1, \; P_k^{(0)} = 0 \; (k \neq i)$ we get

$$P_{ij} = \begin{cases} 1, & \text{if } \; i = j \\ 0, & \text{if } \; i \neq j \end{cases} ,$$

Setting

$$\mathcal{P}(n) = (P_{ij}(n))_{i,j \in N} ,$$

we get

$$\mathcal{P}(0) = I, \; \mathcal{P}(1) = \mathcal{P}, \; \mathcal{P}(2) = \mathcal{P}(1) \cdot \mathcal{P} = \mathcal{P}^2, \cdots ,$$

where I is an infinite-dimensional unite matrix and $\mathcal{P}$ is the matrix of transition probabilities. It is evident that

$$\mathcal{P}(n) = \mathcal{P}^n \;\; (n = 1, 2, \cdots).$$

Let consider some examples.

Example 14.1 (Random roaming). Let consider random roaming connected with infinite number of Bernoulli independent experiments when the particle is roaming in the integer-valued points of the real axis such that if it is placed in the i-th position, then the transition probabilities to positions $i+1$ or $i-1$ are equal to p or $q = 1-p$, respectively $(0 < p < 1)$. If with ξ_n we denote the position of the particle after n steps, then sequence

$$\xi_0(\omega) \to \xi_1(\omega) \to \cdots \;\; (\omega \in \Omega)$$

will be the Markov chain, whose transition probabilities have the following form

$$P_{ij} = \begin{cases} p\,, & \text{if } j = i+1 \\ q\,, & \text{if } j = i-1 \end{cases} \quad .$$

Remark 14.1 In our case the physical system (i.e., the particle) has an infinite number of phase positions.

Example 14.2 Let consider a physical system which has three different possible positions $\varepsilon_1, \varepsilon_2, \varepsilon_3$. Assume that after one step matrix $\mathcal{P}$ of transition probabilities has the following form

$$\mathcal{P} = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 1. \end{pmatrix}.$$

In the present example position ε_3 has the property that if physical system will be placed in it, then it remains in this position with probability 1. Such position is called absorbable. If the particle is placed in some position and it remains in it with probability 0 then such position is called reflectable. If position ε_i is absorbable then $P_{ii} = 1$ and if the position ε_i is reflectable, then $P_{ii} = 0$.

If we know that before observation the physical system is placed in position ε_i $(1 \le i \le n)$, then using matrix $\mathcal{P}(m)$ we can find transition probability $P_{ij}(m)$ after m steps. In this case, when an initial position of physical system is not known, but we know probabilities $P_i^{(0)}$ that system is placed in position i, then using total probability formula we can calculate the probability that after m steps the physical system will be placed in position ε_j by the following formula

$$P_j(m) = \sum_{k=1}^{n} P_k^{(0)} \cdot P_{kj}(m).$$

The row-vector

$$P^{(0)} = (P_1^{(0)}, P_2^{(0)}, \cdots, P_n^{(0)})$$

is called the vector of initial distribution of the Markov chain and the row-vector

$$\mathcal{P}^{(m)} = (P_1^{(m)}, P_2^{(m)}, \cdots, P_n^{(m)})$$

is called the distribution vector after m steps for Markov chain. In our notations we get

$$\mathcal{P}^{(m)} = \mathcal{P}^{(0)} \cdot \mathcal{P}(m) = \mathcal{P}^{(0)} \cdot \mathcal{P}^m.$$

We present Markov theorem about limit probabilities without proof.

Theorem 14.1 *Let $(\varepsilon_i)_{1 \le i \le n}$ be the possible positions of a physical system. If the crossing probabilities $P_{ij}^{(m)}$ of the Markov chain are positive for arbitrary natural number m, then there exists a finite family of real numbers $(q_i)_{1 \le i \le n}$ such that*

$$(\forall i)(1 \le i \le n \to \lim_{m \to \infty} P_{ij}(m) = q_j) \ \ (1 \le j \le n).$$

The number q_j $(1 \le j \le n)$ can be considered as the probability of the occurrence in the j-th position of physical system for sufficiently large natural number m.

Tests

14.1. The matrix of the transition probabilities of the Markov chain is defined by

$$\mathcal{P} = \begin{pmatrix} 0 & 1 & 0 \\ 0,5 & 0,5 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

and the vector of initial probabilities coincides with $(0,2;\ 0,5;\ 0,3)$. Then the distribution vector after two steps will be equal to

a) $(0,125;0,475;0,4)$, b) $(0,225;0,475;0,3)$,
c) $(0,025;0,575;0,4)$, d) $(0,125;0,375;0,5)$.

14.2. The matrix of transition probabilities of the Markov chain is defined by

$$\mathcal{P} = \begin{pmatrix} 0,3 & 0,1 & 0,6 \\ 0 & 0,4 & 0,6 \\ 0,4 & 0,3 & 0,3 \end{pmatrix}.$$

Then matrix $\mathcal{P}(2)$ of transition probabilities of the Markov chain after 2-steps has the following form

a)

$$\begin{pmatrix} 0,25 & 0,15 & 0,6 \\ 0 & 0,3 & 0,7 \\ 0,4 & 0,3 & 0,3 \end{pmatrix},$$

b)

$$\begin{pmatrix} 0,33 & 0,21 & 0,46 \\ 0,4 & 0,3 & 0,3 \\ 0,24 & 0,13 & 0,33 \end{pmatrix}.$$

14.3. The matrix of crossing probabilities of the Markov chain is defined by

$$\mathcal{P} = \begin{pmatrix} 0,1 & 0,5 & 0,4 \\ 0 & 0 & 1 \\ 0,5 & 0,3 & 0,2 \end{pmatrix}.$$

Then transition probability from position ε_2 to position ε_3 after 3 steps $P_{23}(3)$ will be equal to

a) $0,125$, b) $0,225$, c) $0,54$, d) $0,375$.

14.4. The matrix of transition probabilities of the Markov chain is defined by

$$\mathcal{P} = \begin{pmatrix} 0,3 & 0,7 \\ 0,1 & 0,9 \end{pmatrix}$$

and the vector of initial probabilities coincides with $(0,2;\ 0,8)$. It is known that transition probability from any initial position ε_i to position ε_i after 2 steps is equal to 0,128. Then i is equal to

a) 1, b) 2.

Chapter 15

The Process of Brownian Motion

Let consider a little particle which is placed in homogeneous liquid. Since the particle undergoes chaotic collisions with molecules of liquid, it is in continuous chaotic (unordered) motion. A discrete analogue of this process is the following random roaming of the particle on the real axis: the particle changes its positions in such moments of times which are múltiple of $\Delta t(\Delta t > 0)$. If the particle is placed in point x then the transition probabilities to positions $x+\Delta x$ and $x-\Delta x$ are the same and are equal to $0,5$ (Here we consider one-dimensional random roaming). We assume that x is the same for arbitrary position x. In the limit, when $\Delta t \to 0$, $\Delta x \to 0$, with spatial law, a continuous random roaming is obtained which describes a model of the Brownian[1] physic process.

Let denote with $\xi_t(\omega)$ the position of the particle in moment t. Assume that the particle is placed in position $x=0$ in initial moment $t=0$. In this case of discrete roaming during time t this particle makes $n=\frac{t}{\Delta t}$ steps. If we denote with $S_n(\omega)$ the number of steps with Δx in positive direction, then the common shift in positive direction will be to equal to $S_n(\omega)\cdot\Delta x$, and the common shift in negative direction will be equal to $(n-S_n(\omega))\cdot\Delta x$. Hence, common shift $\xi_t(\omega)$ after time $t=n\Delta t$ is connected with $S_n(\omega)$ by the following equality

$$\xi_t(\omega)=[S_n(\omega)\Delta x-(n-S_n(\omega))\Delta x]=(2S_n(\omega)-n)\Delta x.$$

If we assume that $\xi_0(\omega)=0$, then

$$\xi_t(\omega)=(\xi_s(\omega)-\xi_0(\omega))+(\xi_t(\omega)-\xi_s(\omega))$$

for every $s\in[0,t]$. Clearly, in our model random variables $\xi_s-\xi_0$ and $\xi_t-\xi_s$ are independent. Since distribution functions of increases $\xi_t-\xi_s$ and $\xi_{t-s}-\xi_0$ are equal

we get that $\sigma^2(t)=D\xi_t$ satisfies the following condition

$$\sigma^2(t)=\sigma^2(s)+\sigma^2(t-s)\ \ (0\le s\le t).$$

[1] Brown; Robert (21.12.1773, - 10.6.1858) - English botanist who was the first to discover so called Brownian motion, which in the probability theory is also known as Wiener process.

It follows that $\sigma^2(t)$ linearly depends on t. It means that there exists positive real number σ^2, such that

$$D\xi_t = \sigma^2 \cdot t.$$

Number σ^2 is called a diffusion coefficient of the Brownian process. On the other hand, it is easy to show that variance of shift after time t (equivalently, after $n = \frac{t}{\Delta t}$ steps) is $D\xi_t = (\Delta x)^2 \cdot \frac{t}{\Delta t}$. Finally, we get following relation between values Δx and Δt :

$$\frac{(\Delta x)^2}{\Delta t} = \sigma^2.$$

Since particle transitions are independent, they can be considered as the Bernoulli experiment with success probability $p = \frac{1}{2}$. Then the number of steps in positive direction $S_n(\omega)$ will be equal to the number of successes in n independent Bernoulli experiments. In this case the position of particle $\xi_t(\omega)$ at moment t will be connected with normalized random variable $S_n^*(\omega) = \frac{1}{\sqrt{n}}(2S_n(\omega) - n)$ with the following equality

$$\xi_t(\omega) = S_n^*(\omega)\sqrt{n}\Delta x = S_n^*(\omega)\sqrt{t}\frac{\Delta x}{\sqrt{\Delta t}} = S_n^*(\omega)\sigma\sqrt{t}.$$

Using Theorem 13.10 , we deduce that the distribution function of random variable $\xi_t(\omega)$ in the case of one-dimensional Brownian process has the following form

$$P(\{\omega : x_1 \leq \frac{\xi_t(\omega)}{\sigma\sqrt{t}} \leq x_2\}) = \lim_{\Delta t \to 0} P(\{\omega : x_1 \leq S_n^*(\omega) \leq x_2\}) =$$

$$\lim_{n\to\infty} P(\{\omega : x_1 \leq \frac{S_n(\omega) - np}{\sqrt{npq}} \leq x_2\}) = \frac{1}{\sqrt{2\pi}} \int_{x_1}^{x_2} e^{-\frac{x^2}{2}} dx,$$

where $p = q = \frac{1}{2}$.

One can easily demonstrate the validity of the following formula

$$P(\{\omega : y_1 \leq \xi_t(\omega) \leq y_2\}) = (\Phi(\frac{y_2}{\sigma\sqrt{t}}) - \Phi(\frac{y_1}{\sigma\sqrt{t}}))\ (t > 0, y_1 < y_2)$$

Now we consider a problem of prognosis of the Brownian motion. Let $(\xi_t(\omega))t > 0 (\omega \in \Omega)$ be a Brownian process with unknown diffusion coefficient σ^2. Let $(\xi_{t_k}(\omega))_{1\leq k \leq n+1}$ be the result of observations on this process at moments $(t_k)_{1\leq k\leq n+1}$. Here we assume that $t_1 = 0, \xi_{t_1}(\omega) = 0$ and $t_k < t_{k+1}$. We set

$$X_k(\omega) = \frac{\xi_{t_{k+1}}(\omega) - \xi_{t_k}(\omega)}{\sqrt{t_{k+1} - t_k}}\ (1 \leq k \leq n).$$

It is clear that $(X_k(\omega))_{1\leq k\leq n}$ is a sequence of independent random variables normally distributed with parameters $(0;\sigma^2)$, where σ^2 is an unknown parameter. From the course of mathematical statistics it is known that statistics σ_n^2 defined with

$$\sigma_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i(\omega) - \frac{1}{n}\sum_{j=1}^{n} X_j(\omega)),$$

is a good estimation of unknown parameter σ^2.

The prognosis of the stochastic behavior of the Brownian motion in moment $t(t > t_{n+1})$ can be given with the following formula

$$P(\{\omega : y_1 \leq \xi_t(\omega) \leq y_2\}) = (\Phi(\frac{y_2}{\sigma_n\sqrt{t}}) - \Phi(\frac{y_1}{\sigma_n\sqrt{t}}))\ (t > 0, y_1 < y_2).$$

Remark 15.1. Using statistical functions NORMDIST and VAR (cf. p.134) the prognosis of the stochastic behavior of the Brownian motion in the moment $t(t > t_{n+1})$ can be given with the following formula

$$P(\{\omega : y_1 \leq \xi_t(\omega) \leq y_2\}) = \text{ NORMDIST}(\frac{y_2}{\sqrt{t\times\ \text{VAR}(x_1 : x_n)}};0;1;1)-$$
$$\text{NORMDIST}(\frac{y_1}{\sqrt{t\times\ \text{VAR}(x_1 : x_n)}};0;1;1),$$

where $x_k = X_k(\omega) for 1 \leq k \leq n$.

Remark 15.2. It is reasonable to note that the hypothesis about the form of the distribution function of one-dimensional Brownian motion belongs to eminent physician Albert Einstein. His conjecture was strongly proved by American mathematician Norbert Wiener to whom belongs the mathematical construction of the Brownian motion. Hence, in literature the Brownian process is mentioned also as Wiener process.

Tests

15.1. The change of the commodity price is the Brownian process with diffusion coefficient $\sigma^2 = 1$. At $t = 0$ the price of the commodity was equal to 9 lari. The probability that the price of the commodity will not increase at moment $t = 9$, is equal to

a) $0,4$, b) $0,5$, c) $0,6$, d) $0,3$.

15.2. The change of the commodity price is the Brownian process with diffusion coefficient $\sigma^2 = 1$. At $t = 0$ the price of the commodity was equal to 200 lari. The probability that the price of the commodity at the moment t = 9 will be

1) less than 190 lari, is equal to

a) $0,3064$, b) $0,3164$, c) $0,3264$, d) $0,3364$;

2) more than 210 lari, is equal to

a) $0,2864$, b) $0,3264$, c) $0,3464$, d) $0,3664$;

3) placed in interval [185 , 205], is equal to

a) $0,3027$, b) $0,3227$, c) $0,3527$, d) $0,3727$.

15.3. The change of a bonds price is the Brownian process with diffusion coefficient $\sigma^2 = 1$. The firm bought the bond for 3000 lari at the moment $t = 0$. The probability that

1) the pr0́fit obtained by buying the bond at moment t = 250000 will be more than 300 lari, is equal to

a) 0, b) 0,1, c) 0,2, d) 0,3;

2) the damage obtained by buying of bond at the moment t = 900 will be grater than 15 lari, is equal to

a) 0, b) 1, c) 0,3, d) 0,6.

15.4. The change of the goods price in the shop is the Brownian process with diffusion coefficient $\sigma^2 = 1$. At $t = 0$ the price of the goods was equal to 50 lari. The buyer is interested to buy the goods for no more than 55 lari. The shop stops selling the goods if its price decreases below 41 lari. The probability that the buyer bought the goods in moment $t = 1\frac{2}{3}$, is equal to

a) 0,2287, b) 0,3387, c) 0,4487, d) 0,5587.

Chapter 16

Mathematical Statistics

16.1 Introduction

Statistics [1] is the science of the collection, organization, and interpretation of data(see [1],[2]). It deals with all aspects of this, including the planning of data collection in terms of the design of surveys and experiments(see [1]).

A statistician is someone who is particularly well versed in the ways of thinking necessary for the successful application of statistical analysis. Such people have often gained this experience through working in any of a wide number of fields. There is also a discipline called mathematical statistics, which is concerned with the theoretical basis of the subject.

The word statistics, when referring to the scientific discipline, is singular, as in ”Statistics is an art”(see [3]) This should not be confused with the word statistic, referring to a quantity (such as mean or median) calculated from a set of data(see [4]), whose plural is statistics, e.g. ”This statistic seems wrong.” or ”These statistics are misleading.”

16.2 Scope

Statistics is considered by some to be a mathematical science pertaining to the collection, analysis, interpretation or explanation, and presentation of data,[5] while others consider it a branch of mathematics[6] concerned with collecting and interpreting data. Because of its empirical roots and its focus on applications, statistics is usually considered to be a distinct mathematical science rather than a branch of mathematics.[7][8]

Statisticians improve the quality of data with the design of experiments and survey sampling. Statistics also provides tools for prediction and forecasting using data and statistical models. Statistics is applicable to a wide variety of academic disciplines, including natural and social sciences, government, and business. Statistical consultants are available to provide help for organisations and companies without direct access to expertise relevant to their particular problems.

[1] This material is referred from cite http://en.wikipedia.org/wiki/Statistics

Statistical methods can be used to summarize or describe a collection of data; this is called descriptive statistics. This is useful in research, when communicating the results of experiments. In addition, patterns in the data may be modeled in a way that accounts for randomness and uncertainty in the observations, and are then used to draw inferences about the process or population being studied; this is called inferential statistics. Inference is a vital element of scientific advance, since it provides a prediction (based in data) for where a theory logically leads. To further prove the guiding theory, these predictions are tested as well, as part of the scientific method. If the inference holds true, then the descriptive statistics of the new data increase the soundness of that hypothesis. Descriptive statistics and inferential statistics (a.k.a., predictive statistics) together comprise applied statistics.[9]

Statistics is closely related to probability theory, with which it is often grouped; the difference, roughly, is that in probability theory one starts from given parameters of a total population, to deduce probabilities pertaining to samples, while statistical inference, moving in the opposite direction, is inductive inference from samples to the parameters of a larger or total population.

16.3 History

Some scholars pinpoint the origin of statistics to 1663, with the publication of Natural and Political Observations upon the Bills of Mortality by John Graunt.[10] Early applications of statistical thinking revolved around the needs of states to base policy on demographic and economic data, hence its stat- etymology. The scope of the discipline of statistics broadened in the early 19th century to include the collection and analysis of data in general. Today, statistics is widely employed in government, business, and the natural and social sciences.

Its mathematical foundations were laid in the 17th century with the development of probability theory by Blaise Pascal and Pierre de Fermat. Probability theory arose from the study of games of chance. The method of least squares was first described by Carl Friedrich Gauss around 1794. The use of modern computers has expedited large-scale statistical computation, and has also made possible new methods that are impractical to perform manually.

16.4 Overview

In applying statistics to a scientific, industrial, or societal problem, it is necessary to begin with a population or process to be studied. Populations can be diverse topics such as ”all persons living in a country” or ”every atom composing a crystal”. A population can also be composed of observations of a process at various times, with the data from each observation serving as a different member of the overall group. Data collected about this kind of ”population” constitutes what is called a time series.

For practical reasons, a chosen subset of the population called a sample is studied as opposed to compiling data about the entire group (an operation called census). Once a sample that is representative of the population is determined, data is collected for the sample members in an observational or experimental setting. This data can then be subjected to statistical analysis, serving two related purposes: description and inference.

* Descriptive statistics summarize the population data by describing what was observed in the sample numerically or graphically. Numerical descriptors include mean and standard deviation for continuous data types (like heights or weights), while frequency and percentage are more useful in terms of describing categorical data (like race).

* Inferential statistics uses patterns in the sample data to draw inferences about the population represented, accounting for randomness. These inferences may take the form of: answering yes/no questions about the data (hypothesis testing), estimating numerical characteristics of the data (estimation), describing associations within the data (correlation) and modeling relationships within the data (for example, using regression analysis). Inference can extend to forecasting, prediction and estimation of unobserved values either in or associated with the population being studied; it can include extrapolation and interpolation of time series or spatial data, and can also include data mining.

The concept of correlation is particularly noteworthy for the potential confusion it can cause. Statistical analysis of a data set often reveals that two variables (properties) of the population under consideration tend to vary together, as if they were connected. For example, a study of annual income that also looks at age of death might find that poor people tend to have shorter lives than affluent people. The two variables are said to be correlated; however, they may or may not be the cause of one another. The correlation phenomena could be caused by a third, previously unconsidered phenomenon, called a lurking variable or confounding variable. For this reason, there is no way to immediately infer the existence of a causal relationship between the two variables. (See Correlation does not imply causation.)

For a sample to be used as a guide to an entire population, it is important that it is truly a representative of that overall population. Representative sampling assures that the inferences and conclusions can be safely extended from the sample to the population as a whole. A major problem lies in determining the extent to which the sample chosen is actually representative. Statistics offers methods to estimate and correct for any random trending within the sample and data collection procedures. There are also methods for designing experiments that can lessen these issues at the outset of a study, strengthening its capability to discern truths about the population. Statisticians[citation needed] describe stronger methods as more ”robust”.(See experimental design.)

Randomness is studied using the mathematical discipline of probability theory. Probability is used in ”Mathematical statistics” (alternatively, ”statistical theory”) to study the sampling distributions of sample statistics and, more generally, the properties of statistical procedures. The use of any statistical method is valid when the system or population under consideration satisfies the assumptions of the method.

Misuse of statistics can produce subtle, but serious errors in description and interpretation subtle in the sense that even experienced professionals make such errors, and serious in the sense that they can lead to devastating decision errors. For instance, social policy, medical practice, and the reliability of structures like bridges all rely on the proper use of statistics. There is further discussion later. Even when statistical techniques are correctly applied, the results can be difficult to interpret for those lacking expertise. The statistical significance of a trend in the data which measures the extent to which a trend could be caused by random variation in the sample may or may not agree with an intuitive sense of its significance. The set of basic statistical skills (and skepticism) that people need to deal with information in their everyday lives properly is referred to as statistical literacy.

16.5 Statistical methods

16.5.1 Experimental and observational studies

A common goal for a statistical research project is to investigate causality, and in particular to draw a conclusion on the effect of changes in the values of predictors or independent variables on dependent variables or response. There are two major types of causal statistical studies: experimental studies and observational studies. In both types of studies, the effect of differences of an independent variable (or variables) on the behavior of the dependent variable are observed. The difference between the two types lies in how the study is actually conducted. Each can be very effective. An experimental study involves taking measurements of the system under study, manipulating the system, and then taking additional measurements using the same procedure to determine if the manipulation has modified the values of the measurements. In contrast, an observational study does not involve experimental manipulation. Instead, data are gathered and correlations between predictors and response are investigated.

16.5.2 Experiments

The basic steps of a statistical experiment are:

1. Planning the research, including finding the number of replicates of the study, using the following information: preliminary estimates regarding the size of treatment effects, alternative hypotheses, and the estimated experimental variability. Consideration of the selection of experimental subjects and the ethics of research is necessary. Statisticians recommend that experiments compare (at least) one new treatment with a standard treatment or control, to allow an unbiased estimate of the difference in treatment effects.

2. Design of experiments, using blocking to reduce the influence of confounding variables, and randomized assignment of treatments to subjects to allow unbiased estimates of treatment effects and experimental error. At this stage, the experimenters and statisticians write the experimental protocol that shall guide the performance of the experiment and that specifies the primary analysis of the experimental data.

3. Performing the experiment following the experimental protocol and analyzing the data following the experimental protocol.

4. Further examining the data set in secondary analyses, to suggest new hypotheses for future study.

5. Documenting and presenting the results of the study.

Experiments on human behavior have special concerns. The famous Hawthorne study examined changes to the working environment at the Hawthorne plant of the Western Electric Company. The researchers were interested in determining whether increased illumination would increase the productivity of the assembly line workers. The researchers first measured the productivity in the plant, then modified the illumination in an area of the plant and checked if the changes in illumination affected productivity. It turned out that productivity indeed improved (under the experimental conditions). However, the study is heavily criticized today for errors in experimental procedures, specifically for the lack of a control group and blindness. The Hawthorne effect refers to finding that an outcome (in this case, worker productivity) changed due to observation itself. Those in the Hawthorne study became more productive not because the lighting was changed but because they were being observed

16.5.3 Observational study

An example of an observational study is one that explores the correlation between smoking and lung cancer. This type of study typically uses a survey to collect observations about the area of interest and then performs statistical analysis. In this case, the researchers would collect observations of both smokers and non-smokers, perhaps through a case-control study, and then look for the number of cases of lung cancer in each group.

16.5.4 Levels of measurement

There are four main levels of measurement used in statistics: nominal, ordinal, interval, and ratio. Each of these have different degrees of usefulness in statistical research. Ratio measurements have both a meaningful zero value and the distances between different measurements defined; they provide the greatest flexibility in statistical methods that can be used for analyzing the data.[citation needed] Interval measurements have meaningful distances between measurements defined, but the zero value is arbitrary (as in the case with longitude and temperature measurements in Celsius or Fahrenheit). Ordinal measurements have imprecise differences between consecutive values, but have a meaningful order to those values. Nominal measurements have no meaningful rank order among values.

Because variables conforming only to nominal or ordinal measurements cannot be reasonably measured numerically, sometimes they are grouped together as categorical variables, whereas ratio and interval measurements are grouped together as quantitative or continuous variables due to their numerical nature.

16.5.5 Key terms used in statistics - Null hypothesis

Interpretation of statistical information can often involve the development of a null hypothesis in that the assumption is that whatever is proposed as a cause has no effect on the variable being measured.

The best illustration for a novice is the predicament encountered by a jury trial. The null hypothesis, H0, asserts that the defendant is innocent, whereas the alternative hypothesis, H1, asserts that the defendant is guilty.

The indictment comes because of suspicion of the guilt. The H0 (status quo) stands in opposition to H1 and is maintained unless H1 is supported by evidence beyond a reasonable doubt. However, failure to reject H0 in this case does not imply innocence, but merely that the evidence was insufficient to convict. So the jury does not necessarily accept H0 but fails to reject H0. While to the casual observer the difference appears moot, misunderstanding the difference is one of the most common and arguably most serious errors made by non-statisticians. Failure to reject the H0 does NOT prove that the H0 is true, as any crook with a good lawyer who gets off because of insufficient evidence can attest to. While one can not prove a null hypothesis one can test how close it is to being true with a power test, which tests for type II errors.

16.5.6 Key terms used in statistics - Error

Working from a null hypothesis two basic forms of error are recognized:

* Type I errors where the null hypothesis is falsely rejected giving a ”false positive”.

* Type II errors where the null hypothesis fails to be rejected and an actual difference between populations is missed.

Error also refers to the extent to which individual observations in a sample differ from a central value, such as the sample or population mean. Many statistical methods seek to minimize the mean-squared error, and these are called ”methods of least squares.”

Measurement processes that generate statistical data are also subject to error. Many of these errors are classified as random (noise) or systematic (bias), but other important types of errors (e.g., blunder, such as when an analyst reports incorrect units) can also be important.

16.5.7 Key terms used in statistics - Confidence intervals

Most studies will only sample part of a population and then the result is used to interpret the null hypothesis in the context of the whole population. Any estimates obtained from the sample only approximate the population value. Confidence intervals allow statisticians to express how closely the sample estimate matches the true value in the whole population. Often they are expressed as 95 confidence intervals. Formally, a 95 confidence interval of a procedure is a range where, if the sampling and analysis were repeated under the same conditions, the interval would include the true (population) value 95 of the time. This does not imply that the probability that the true value is in the confidence interval is 95. (From the frequentist perspective, such a claim does not even make sense, as the true value is not a random variable. Either the true value is or is not within the given interval.) One quantity that is in fact a probability for an estimated value is the credible interval from Bayesian statistics.

16.5.8 Key terms used in statistics - Significance

Statistics rarely give a simple Yes/No type answer to the question asked of them. Interpretation often comes down to the level of statistical significance applied to the numbers and often refer to the probability of a value accurately rejecting the null hypothesis (sometimes referred to as the p-value).

Referring to statistical significance does not necessarily mean that the overall result is significant in real world terms. For example, in a large study of a drug it may be shown that the drug has a statistically significant but very small beneficial effect, such that the drug will be unlikely to help the patient in a noticeable way.

16.5.9 Key terms used in statistics - Examples

Some well-known statistical tests and procedures are:
* Analysis of variance (ANOVA)
* Chi-square test
* Correlation
* Factor analysis
* MannWhitney U
* Mean square weighted deviation (MSWD)
* Pearson product-moment correlation coefficient
* Regression analysis
* Spearman's rank correlation coefficient
* Student's t-test
* Time series analysis

16.6 Application of Statistical Techniques

Statistical techniques are used in a wide range of types of scientific and social research, including: Biostatistics, Computational biology, Computational sociology, Network biology, Social science, Sociology and Social research. Some fields of inquiry use applied statistics so extensively that they have specialized terminology. These disciplines include:
* Actuarial science
* Applied information economics
* Biostatistics
* Business statistics
* Chemometrics (for analysis of data from chemistry)
* Data mining (applying statistics and pattern recognition to discover knowledge from data)
* Demography
* Econometrics
* Energy statistics
* Engineering statistics

* Epidemiology
* Geography and Geographic Information Systems, specifically in Spatial analysis
* Image processing
* Psychological statistics
* Reliability engineering
* Social statistics

16.6.1 Key terms used in statistics -Specialized disciplines

In addition, there are particular types of statistical analysis that have also developed their own specialized terminology and methodology:

* Bootstrap Jackknife Resampling
* Multivariate statistics
* Statistical classification
* Statistical surveys
* Structured data analysis (statistics)
* Survival analysis
* Statistics in various sports, particularly baseball and cricket

Statistics form a key basis tool in business and manufacturing as well. It is used to understand measurement systems variability, control processes (as in statistical process control or SPC), for summarizing data, and to make data-driven decisions. In these roles, it is a key tool, and perhaps the only reliable tool.

16.6.2 Key terms used in statistics -Statistical computing

The rapid and sustained increases in computing power starting from the second half of the 20th century have had a substantial impact on the practice of statistical science. Early statistical models were almost always from the class of linear models, but powerful computers, coupled with suitable numerical algorithms, caused an increased interest in nonlinear models (such as neural networks) as well as the creation of new types, such as generalized linear models and multilevel models.

Increased computing power has also led to the growing popularity of computationally intensive methods based on resampling, such as permutation tests and the bootstrap, while techniques such as Gibbs sampling have made use of Bayesian models more feasible. The computer revolution has implications for the future of statistics with new emphasis on ”experimental” and ”empirical” statistics. A large number of both general and special purpose statistical software are now available.

16.6.3 Key terms used in statistics -Misuse

There is a general perception that statistical knowledge is all-too-frequently intentionally misused by finding ways to interpret only the data that are favorable to the presenter. The famous saying, ”There are three kinds of lies: lies, damned lies, and statistics”.[11] which was popularized in the USA by Samuel Clemens and incorrectly attributed by him to Disraeli (18041881), has come to represent the general mistrust [and misunderstanding] of statistical science. Harvard President Lawrence Lowell wrote in 1909 that statistics, "...like veal pies, are good if you know the person that made them, and are sure of the ingredients."

If various studies appear to contradict one another, then the public may come to distrust such studies. For example, one study may suggest that a given diet or activity raises blood pressure, while another may suggest that it lowers blood pressure. The discrepancy can arise from subtle variations in experimental design, such as differences in the patient groups or research protocols, which are not easily understood by the non-expert. (Media reports usually omit this vital contextual information entirely, because of its complexity.)

By choosing (or rejecting, or modifying) a certain sample, results can be manipulated. Such manipulations need not be malicious or devious; they can arise from unintentional biases of the researcher. The graphs used to summarize data can also be misleading.

Deeper criticisms come from the fact that the hypothesis testing approach, widely used and in many cases required by law or regulation, forces one hypothesis (the null hypothesis) to be "favored," and can also seem to exaggerate the importance of minor differences in large studies. A difference that is highly statistically significant can still be of no practical significance. (See criticism of hypothesis testing and controversy over the null hypothesis.)

One response is by giving a greater emphasis on the p-value than simply reporting whether a hypothesis is rejected at the given level of significance. The p-value, however, does not indicate the size of the effect. Another increasingly common approach is to report confidence intervals. Although these are produced from the same calculations as those of hypothesis tests or p-values, they describe both the size of the effect and the uncertainty surrounding it.

16.6.4 Key terms used in statistics -Statistics applied to mathematics or the arts

Traditionally, statistics was concerned with drawing inferences using a semi-standardized methodology that was "required learning" in most sciences. This has changed with use of statistics in non-inferential contexts. What was once considered a dry subject, taken in many fields as a degree-requirement, is now viewed enthusiastically. Initially derided by some mathematical purists, it is now considered essential methodology in certain areas.

* In number theory, scatter plots of data generated by a distribution function may be transformed with familiar tools used in statistics to reveal underlying patterns, which may then lead to hypotheses.

* Methods of statistics including predictive methods in forecasting, are combined with chaos theory and fractal geometry to create video works that are considered to have great beauty.

* The process art of Jackson Pollock relied on artistic experiments whereby underlying distributions in nature were artistically revealed. With the advent of computers, methods of statistics were applied to formalize such distribution driven natural processes, in order to

make and analyze moving video art.

* Methods of statistics may be used predicatively in performance art, as in a card trick based on a Markov process that only works some of the time, the occasion of which can be predicted using statistical methodology.

* Statistics can be used to predicatively create art, as in the statistical or stochastic music invented by Iannis Xenakis, where the music is performance-specific. Though this type of artistry does not always come out as expected, it does behave in ways that are are predictable and tuneable using statistics.

References

1. Dodge, Y. (2003) The Oxford Dictionary of Statistical Terms, OUP. ISBN 0-19-920613-9

2. The Free Online Dictionary

3. ”Statistics”. Merriam-Webster Online Dictionary. http://www.merriam-webster.com/dictionary/statistics.

4. ”Statistic”. Merriam-Webster Online Dictionary. http://www.merriam-webster.com/dictionary/statistic.

5. Moses, Lincoln E. Think and Explain with statistics, pp. 13. Addison-Wesley, 1986.

6. Hays, William Lee, Statistics for the social sciences, Holt, Rinehart and Winston, 1973, p.xii, ISBN 978-0-03-077945-9

7. Moore, David (1992). ”Teaching Statistics as a Respectable Subject”. Statistics for the Twenty-First Century. Washington, DC: The Mathematical Association of America. pp. 1425.

8. Chance, Beth L.; Rossman, Allan J. (2005). ”Preface”. Investigating Statistical Concepts, Applications, and Methods. Duxbury Press. ISBN 978-0495050643. http://www.rossmanchance.com/iscam/preface.pdf.

9. Anderson, , D.R.; Sweeney, D.J.; Williams, T.A.. Statistics: Concepts and Applications, pp. 59. West Publishing Company, 1986.

10. Willcox, Walter (1938) The Founder of Statistics. Review of the International Statistical Institute 5(4):321328.

11. Leonard H.Courtney (18321918) in a speech at Saratoga Springs, New York, August 1895, in which this sentence appeared: After all, facts are facts, and although we may quote one to another with a chuckle the words of the Wise Statesman, Lies damned lies and statistics, still there are some easy figures the simplest must understand, and the astutest cannot wriggle out of., earliest documented use of exact phrase.

Chapter 17

Point, Well-Founded and Effective Estimations

Let $(\Omega, \mathcal{F}, P)$ be a probability space. Let $X_1, \cdots, X_n$ be a sequence of equally distributed random variables, i.e.,

$$(\forall y)(y \in \overline{R} \rightarrow F_{X_1}(y) = \cdots = F_{X_n}(y) = F(y)).$$

Suppose that results of n independent experiments is a vector $x = (x_1, \cdots, x_n)$, which can be considered as a partial realization of the random vector $X = (X_1, \cdots, X_n)$, i.e., there exists an elementary event $\omega \in \Omega$ such that

$$x = (x_1, \cdots, x_n) = (X_1(\omega), \cdots, X_n(\omega)) = X(\omega).$$

Definition 17.1 A vector $x = (x_1, \cdots, x_n)$ is called n-dimensional sample or a sample of size n.

Statistical Assumption 17. 1 Suppose that the probability measure P_F defined by F (see Section 5.3) belongs to the family of Borel probability measures $(P_\theta)_{\theta \in \Theta}$ defined on R, i.e. there is a parameter $\theta_0 \in \Theta$ such that $P_F = P_{\theta_0}$.

Definition 17.2 A triplet $(R^n, \mathcal{B}(R^n), P_\theta^n)_{\theta \in \Theta}$ is called probability-statistic model.

Under Statistical Assumption 17. 1, there is a parameter $\theta_0 \in \Theta$ such that $F_{X_1} = \cdots = F_{X_n} = F_{\theta_0}$, where F_{θ_0} is a distribution function generated by the Borel probability measure P_{θ_0}.

Statistical Assumption 17. 2Suppose that a parameter set Θ is a Borel subset of the real axis and is equipped with induced Borel σ-algebra.

Definition 17.3 A Borel mapping $\widehat{\theta} : R^n \rightarrow \Theta$ called a point estimator or statistic of size-n.

Definition 17.4 For a given sample x, a value $e(x) = \widehat{\theta}(x) - \theta$, is called the "error" of the estimator $\widehat{\theta}$ as where θ is the parameter being estimated.

Note that the error $e(x)$ depends not only on the estimator (the estimation formula or procedure), but on the sample.

Definition 17.5 The mean squared error of $\widehat{\theta}$ is defined as the expected value (probability-weighted average, over all samples) of the squared errors; that is,

$$\mathrm{MSE}(\widehat{\theta}) = \mathrm{E}[(\widehat{\theta}(X)-\theta)^2] = \int_{\Omega}[(\widehat{\theta}(X(\omega))-\theta)^2]dP(\omega) = \int_{R^n}(\widehat{\theta}(x_1,\cdots,x_n)-\theta)^2 dP_{\theta}^n(x_1,\cdots,x_n).$$

It is used to indicate how far, on average, the collection of estimates are from the single parameter being estimated. Consider the following analogy. Suppose the parameter is the bull's-eye of a target, the estimator is the process of shooting arrows at the target, and the individual arrows are estimates (samples). Then high MSE means the average distance of the arrows from the bull's-eye is high, and low MSE means the average distance from the bull's-eye is low. The arrows may or may not be clustered. For example, even if all arrows hit the same point, yet grossly miss the target, the MSE is still relatively large. Note, however, that if the MSE is relatively low, then the arrows are likely more highly clustered (than highly dispersed).

Definition 17.6 For a given sample x , the sampling deviation of the estimator $\widehat{\theta}$ is defined as

$$d(x) = \widehat{\theta}(x) - \mathrm{M}(\widehat{\theta}(X)) = \widehat{\theta}(x) - \mathrm{M}(\widehat{\theta}),$$

where $\mathrm{M}(\widehat{\theta}(X))$ is the mathematical expectation of the estimator.

Note that the sampling deviation $d(x)$ depends not only on the estimator, but on the sample.

Definition 17.7 The variance of $\widehat{\theta}$ is simply the expected value of the squared sampling deviations; that is,

$$\mathrm{var}(\widehat{\theta}) = \mathrm{M}[(\widehat{\theta} - \mathrm{M}(\widehat{\theta}))^2].$$

It is used to indicate how far, on average, the collection of estimates are from the expected value of the estimates. Note the difference between MSE and variance. If the parameter is the bull's-eye of a target, and the arrows are estimates, then a relatively high variance means the arrows are dispersed, and a relatively low variance means the arrows are clustered. Some things to note: even if the variance is low, the cluster of arrows may still be far off-target, and even if the variance is high, the diffuse collection of arrows may still be unbiased. Finally, note that even if all arrows grossly miss the target, if they nevertheless all hit the same point, the variance is zero.

Definition 17.8 The bias of $\widehat{\theta}$ is defined as

$$B(\widehat{\theta}) = \mathrm{M}(\widehat{\theta}) - \theta.$$

It is the distance between the average of the collection of estimates, and the single parameter being estimated. It also is the expected value of the error, since

$$\mathrm{M}(\widehat{\theta}) - \theta = \mathrm{M}(\widehat{\theta} - \theta).$$

If the parameter is the bull's-eye of a target, and the arrows are estimates, then a relatively high absolute value for the bias means the average position of the arrows is off-target, and a relatively low absolute bias means the average position of the arrows is on target. They may be dispersed, or may be clustered. The relationship between bias and variance is analogous to the relationship between accuracy and precision.

Definition 17.9 The estimator $\widehat{\theta}$ is an unbiased estimator of θ if and only if $B(\widehat{\theta}) = 0$, i.e.

$$(\forall\theta)(\theta \in \Theta \to \mathrm{M}_\theta \widehat{\theta} = \theta),$$

where

$$\mathrm{M}_\theta \widehat{\theta} = \int_{R^n} \widehat{\theta}(x_1, \cdots, x_n) dP_\theta^n(x_1, \cdots, x_n).$$

Note that bias is a property of the estimator, not of the estimate. Often, people refer to a "biased estimate" or an "unbiased estimate," but they really are talking about an "estimate from a biased estimator," or an "estimate from an unbiased estimator." Also, people often confuse the "error" of a single estimate with the "bias" of an estimator. Just because the error for one estimate is large, does not mean the estimator is biased. In fact, even if all estimates have astronomical absolute values for their errors, if the expected value of the error is zero, the estimator is unbiased. Also, just because an estimator is biased, does not preclude the error of an estimate from being zero (we may have gotten lucky). The ideal situation, of course, is to have an unbiased estimator with low variance, and also try to limit the number of samples where the error is extreme (that is, have few outliers). Yet unbiasedness is not essential. Often, if just a little bias is permitted, then an estimator can be found with lower MSE and or fewer outlier sample estimates.

Remark 17. 1 The MSE, variance, and bias, are related:

$$\mathrm{MSE}(\widehat{\theta}) = \mathrm{var}(\widehat{\theta}) + (B(\widehat{\theta}))^2,$$

i.e. mean squared error = variance + square of bias. In particular, for an unbiased estimator, the variance equals the MSE.

Definition 17. 10 The standard deviation of an estimator of θ (the square root of the variance), or an estimate of the standard deviation of an estimator of θ , is called the standard error of θ.

Definition 17. 11 An estimator $\widehat{\theta}$ is called well-founded estimator of θ if $\lim_{n\to n} M_\theta \widehat{\theta} = \theta$.

Definition 17. 12 An estimator $\widehat{\theta}$ is called Effective estimator of θ if

$$D_\theta \widehat{\theta} = \inf\{D_\theta T : T \in \mathcal{T}^n\},$$

where

$$D_\theta \widehat{\theta} = M_\theta(\widehat{\theta} - M_\theta \widehat{\theta})^2$$

and $\mathcal{T}^n$ denotes a class of all well-founded estimators of θ.

Chapter 18

Point Estimators of Average and Variance

Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a probability space. Let $X_1, \cdots, X_n$ be a sequence of equally distributed random variables with average μ and variance δ^2.

$(R^n, \mathcal{B}(R^n), P_\theta^n)_{\theta \in \Theta}$ is probability-statistic model such that $\Theta \subseteq R \times (0, +\infty)$ and there exists $(\theta_0^{(1)}, \theta_0^{(2)}) \in \Theta$ such that $(\theta_0^{(1)}, \theta_0^{(2)}) = (\mu, \delta^2)$.

Definition 18. 1 An estimator $\overline{X}_n : R^n \to R$, defined by

$$\overline{X}_n(x_1, \cdots, x_n) = \frac{1}{n} \sum_{k=1}^{n} x_k,$$

is called a sample average (or mean).

Remark 18. 1 "EXCEL's" statistical function AVERAGE$(x_1 : x_n)$ calculates a sample average $\overline{X}_n(x_1, \cdots, x_n) = \frac{1}{n} \sum_{k=1}^{n} x_k$.

Theorem 18.1 *A sample average estimator $\overline{X}_n$ is an unbiased dot estimator of the first coordinate of the parameter $(\theta^{(1)}, \theta^{(2)}) \in \Theta$.*

Proof. We have

$$(\forall \theta)(\theta = (\theta^{(1)}, \theta^{(2)}) \in \Theta \to \mathrm{M}_\theta \overline{X}_n =$$

$$\int_{R^n} \overline{X}_n(x_1, \cdots, x_n) dP_\theta^n(x_1, \cdots, x_n) =$$

$$\int_{R^n} \frac{1}{n} \sum_{k=1}^{n} x_k dP_\theta^n(x_1, \cdots, x_n) =$$

$$\frac{1}{n} \sum_{k=1}^{n} \int_{R^n} x_k dP_\theta^n(x_1, \cdots, x_n) = \frac{1}{n} \sum_{k=1}^{n} \int_R x_k dP_\theta(x_k) = \frac{1}{n} n\theta^{(1)} = \theta^{(1)}).$$

□

Definition 18. 2 An estimator $S_n^2 : R^n \to R$, defined by

$$S_n^2(x_1, \cdots, x_n) = \frac{1}{n}\sum_{k=1}^{n}(x_k - \overline{X}_n)^2$$

is called a sample variance.

Remark 18.2 "EXCEL's" statistical function VARP$(x_1 : x_n)$ calculates a sample variance $S_n^2(x_1, \cdots, x_n) = \frac{1}{n}\sum_{k=1}^{n}(x_k - \overline{X}_n)^2$.

Theorem 18.2 *For sample variance estimator S_n^2 we have*

$$(\forall\theta)(\theta = (\theta^{(1)}, \theta^{(2)}) \in \Theta \to \mathrm{M}_\theta S_n^2 = \frac{n-1}{n}\theta^{(2)}).$$

Proof.

$$(\forall\theta)(\theta = (\theta^{(1)}, \theta^{(2)}) \in \Theta \to \mathrm{M}_\theta S_n^2 =$$

$$\mathrm{M}_\theta(\sum_{k=1}^{n}\frac{1}{n}x_k^2 - \frac{2}{n}\sum_{k=1}^{n}x_k\overline{X}_n + \frac{1}{n}\sum_{k=1}^{n}\overline{X}_n^2) =$$

$$\mathrm{M}_\theta(\frac{1}{n}\sum_{k=1}^{n}x_k^2 - 2\overline{X}_n^2 + \overline{X}_n^2) =$$

$$\mathrm{M}_\theta(\sum_{k=1}^{n}\frac{1}{n}x_k^2 - \overline{X}_n^2) =$$

$$\mathrm{M}_\theta x_i^2 - \mathrm{M}_\theta \overline{X}_n^2 =$$

$$\mathrm{D}_\theta x_i + (\mathrm{M}_\theta x_i)^2 - (\mathrm{D}_\theta \overline{X}_n + (\mathrm{M}_\theta \overline{X}_n)^2) =$$

$$\theta^{(2)} + (\theta^{(1)})^2 - \frac{1}{n}\theta^{(2)} - (\theta^{(1)})^2 = \frac{n-1}{n}\theta^{(2)}.$$

□

Definition 18. 3 An estimator $S_n'^2 : R^n \to R$, defined by

$$S_n'^2(x_1, \cdots, x_n) = \frac{1}{n-1}\sum_{k=1}^{n}(x_k - \overline{X}_n)^2$$

is called a corrected sample variance.

Remark 18. 3 "EXCEL's" statistical function VAR$(x_1 : x_n)$ calculates a corrected sample variance $S_n'^2(x_1, \cdots, x_n) = \frac{1}{n-1}\sum_{k=1}^{n}(x_k - \overline{X}_n)^2$.

Theorem 18. 3 *A corrected sample variance estimator $S_n'^2$ is an unbiased dot estimator of the second coordinate of the parameter $(\theta^{(1)}, \theta^{(2)}) \in \Theta$.*

Proof. For $\theta = (\theta^{(1)}, \theta^{(2)}) \in \Theta$, we have

$$\mathrm{M}_\theta S_n'^2 = \mathrm{M}_\theta \frac{n}{n-1} S_n^2 = \frac{n}{n-1} \mathrm{M}_\theta S_n^2.$$

By Theorem 18. 2, we have

$$\frac{n}{n-1} \mathrm{M}_\theta S_n^2 = \frac{n}{n-1} \frac{n-1}{n} \theta^{(2)} = \theta^{(2)}.$$

This ends the proof of Theorem 18.3.

□

Corollary 18. 1 The estimator $\widehat{\theta} = (\overline{X}_n, S_n'^2)$ is an unbiased estimator of the parameter θ.

Exercises

18.1.1 Here are given annual salaries of 10 different secretaries (1000 American dollars corresponds to one unit)

35.0; 67.5; 51.5; 53; 38; 42; 29.5; 31.5; 45; 37.5.

Suppose that the population of secretaries' annual salaries are distributed normally. Find

a) an unbiased and effective estimate of an annual salary's average (Use "EXCEL's" statistical function AVERAGE$(x_1 : x_n)$).

b) an unbiased estimate of an annual salary's variance (Use "EXCEL's" statistical function VAR$(x_1 : x_n)$).

c) a well-founded estimate of an annual salary's variance (Use "EXCEL's" statistical function VARP$(x_1 : x_n)$).

18.1.2 Here are given weights of 10 accidentally chosen rolls (1 gram corresponds to one unit)

55.0; 57.5; 51.5; 53; 48; 49; 45.5; 45.4; 46; 47.5.

Suppose that the population of rolls' weights are distributed normally. Find

a) an unbiased and effective estimate of a roll's weight average (Use "EXCEL's" statistical function AVERAGE$(x_1 : x_n)$).

b) an unbiased estimate of a roll's weight variance (Use "EXCEL's" statistical function VAR$(x_1 : x_n)$).

c) a well-founded estimate of a roll's weight variance (Use "EXCEL's" statistical function VARP$(x_1 : x_n)$).

18.1.3 Here are given weights of 10 accidentally chosen apples (1 gram corresponds to one unit)

155.0; 157.5; 151.5; 153; 148; 149; 145.5; 145.4; 146; 147.5.

Suppose that the population of apples' weights are distributed normally. Find

a) an unbiased and effective estimate of an apple's weight average (Use "EXCEL's" statistical function AVERAGE$(x_1 : x_n)$).

b) an unbiased estimate of an apple's weight variance (Use "EXCEL's" statistical function VAR$(x_1 : x_n)$).

c) a well-founded estimate of an apple's weight variance (Use "EXCEL's" statistical function VARP$(x_1 : x_n)$).

18.1.4 Here are given lengths of 10 accidentally chosen pencils (1 mm corresponds to one unit)

150.0; 150.5; 150.5; 153; 149; 149; 140.5; 140.4; 140; 140.5.

Suppose that the population of pencils lengths are distributed normally. Find

a) an unbiased and effective estimate of a pencil's length average (Use "EXCEL's" statistical function AVERAGE$(x_1 : x_n)$).

b) an unbiased estimate of a pencil's length variance (Use "EXCEL's" statistical function VAR$(x_1 : x_n)$).

c) a well-founded estimate of a pencil's length variance (Use "EXCEL's" statistical function VARP$(x_1 : x_n)$).

18.1.5 Let the random variable ξ follow a normal distribution with a mean μ and a standard deviation σ. Let $\overline{X_{25}}$ be the mean of samples of sizes 25, randomly and independently selected from the population. Consider the following values: $A := P(\{\omega : \mu - 0,2\sigma < \overline{X_{25}}(\omega) < \mu + \sigma\})$; $B := P(\{\omega : \mu - \sigma < \overline{X_{25}}(\omega) < \mu + 0,2\sigma\})$; Which of the following statements is true?

a) $A = B$; b) $A > B$; c) $A < B$;
d) Unable to determine the relationship between values A and B.

18.1.6 In a recent survey of high school students. It was found that the average amount of money spent on entertainment each week was normally distributed with a mean of $52,30$ dollars and a standard deviation of $18,23$ dollars. Assuming these values are representative of all high school students, what is the probability that for a sample of 25, the average amount spent by each student exceeds 60 dollars?

a) $0,0174$; b) $0,0185$; c) $0,0195$; d) $0,0295$.

8.1.7 Let the random variable ξ follow a normal distribution with a mean μ and a standard deviation σ. Let $\overline{X_{16}}$ and $\overline{X_{25}}$ be the means of samples of sizes 16 and 25, respectively, randomly and independently selected from the population. Consider the following values : $A := P(\{\omega : \overline{X_{16}}(\omega) < \mu\})$; $B := P(\{\omega : \overline{X_{25}}(\omega) < \mu\})$; Which of the following statements is true?

a) $A = B$; b) $A > B$; c) $A < B$;
d) Unable to determine the relationship between values A and B.

Chapter 19

Interval Estimation. Confidence intervals. Credible intervals. Interval Estimators of Parameters of Normally Distributed Random Variable

In statistics, interval estimation is the use of sample data to calculate an interval of possible (or probable) values of an unknown population parameter, in contrast to point estimation, which is a single number. Neyman (1937) identified interval estimation ("estimation by interval") as distinct from point estimation ("estimation by unique estimate"). In doing so, he recognised that then-recent work quoting results in the form of an estimate plus-or-minus a standard deviation indicated that interval estimation was actually the problem statisticians really had in mind.

The most prevalent forms of interval estimation are:

a) confidence intervals (a frequentist method);

b) credible intervals (a Bayesian method).

In statistics, a confidence interval (CI) is a particular kind of interval estimate of a population parameter and is used to indicate the reliability of an estimate. It is an observed interval (i.e it is calculated from the observations), in principle different from sample to sample, that frequently includes the parameter of interest, if the experiment is repeated. How frequently the observed interval contains the parameter is determined by the confidence level or confidence coefficient.

A confidence interval with a particular confidence level is intended to give the assurance that, if the statistical model is correct, then taken over all the data that might have been obtained, the procedure for constructing the interval would deliver a confidence interval that included the true value of the parameter the proportion of the time set by the confidence level. More specifically, the meaning of the term ”confidence level” is that, if confidence intervals are constructed across many separate data analyses of repeated (and possibly different) experiments, the proportion of such intervals that contain the true value of the parameter will approximately match the confidence level; this is guaranteed by the reasoning underlying the construction of confidence intervals.

A confidence interval does not predict that the true value of the parameter has a particular probability of being in the confidence interval given the data actually obtained.(An interval intended to have such a property, called a credible interval, can be estimated using Bayesian methods; but such methods bring with them their own distinct strengths and weaknesses).

Interval estimates can be contrasted with point estimates. A point estimate is a single value given as the estimate of a population parameter that is of interest, for example the mean of some quantity. An interval estimate specifies instead a range within which the parameter is estimated to lie. Confidence intervals are commonly reported in tables or graphs along with point estimates of the same parameters, to show the reliability of the estimates.

For example, a confidence interval can be used to describe how reliable survey results are. In a poll of election voting-intentions, the result might be that 40 % of respondents intend to vote for a certain party. A 90%confidence interval for the proportion in the whole population having the same intention on the survey date might be 38% to 42%. From the same data one may calculate a 95% confidence interval, which might in this case be 36% to 44%. A major factor determining the length of a confidence interval is the size of the sample used in the estimation procedure, for example the number of people taking part in a survey.

Definition 19.1 Let $(R^n, \mathcal{B}(R^n), P^n_\theta)_{\theta\in\Theta}$ be a probability-statistic model and $T_n^{(1)} : R^n \to R$ and $T_n^{(2)} : R^n \to R$ be two statistics such that $T_n^{(1)} < T_n^{(2)}$.

A random interval $(T_n^{(1)}(x), T_n^{(2)})(x)$, based on a random sample $x \in R^n$ (or its observed value x) from that distribution is called a confidence interval with confidence level $\gamma(0 < \gamma < 1)$ for a parameter θ of a probability distribution if the following condition

$$(\forall\theta)(\theta \in \Theta \to P^n_\theta(\{x \in R^n : T_n^{(1)}(x) \le \theta \le T_n^{(2)})(x)\}) = \gamma)$$

holds.

The confidence interval is specified by the pair of statistics (i.e., observable random variables) $T_n^{(1)}(x)$ and $T_n^{(2)}$).

Remark 19.1 Let us discuss how we must interpret confidence intervals.

(see web site http://stattrek.com/AP-Statistics-4/Confidence-Interval.aspx?Tutorial=stat).

Consider the following confidence interval: We are 90% confident that the population mean is greater than 100 and less than 200.

Some people think this means there is a 90% chance that the population mean falls between 100 and 200. This is incorrect. Like any population parameter, the population mean is a constant, not a random variable. It does not change. The probability that a constant falls within any given range is always 0.00 or 1.00.

The confidence level describes the uncertainty associated with a sampling method. Suppose we used the same sampling method to select different samples and to compute a different interval estimate for each sample. Some interval estimates would include the true population parameter and some would not. A 90% confidence level means that we would expect 90% of the interval estimates to include the population parameter; A 95%confidence level means that 95% of the intervals would include the parameter; and so on.

How to Construct a Confidence Interval

There are four steps to constructing a confidence interval.

* Identify a sample statistic. Choose the statistic (e.g, mean, standard deviation) that you will use to estimate a population parameter.

* Select a confidence level. As we noted in the previous section, the confidence level describes the uncertainty of a sampling method. Often, researchers choose 90%, 95%, or 99% confidence levels; but any percentage can be used.

* Find the margin of error. If you are working on a homework problem or a test question, the margin of error may be given. Often, however, you will need to compute the margin of error, based on one of the following equations.

Margin of error = Critical value * Standard deviation of statistic

Margin of error = Critical value * Standard error of statistic

* Specify the confidence interval. The uncertainty is denoted by the confidence level. And the range of the confidence interval is defined by the following equation.

Confidence interval = sample statistic + Margin of error

The sample problem in the next section applies the above four steps to construct a 95% confidence interval for a mean score. The next few theorems discuss this topic in greater detail.

Theorem 19. 1 *Let $X_1, \cdots, X_n$ be a sequence of independent normally distributed real-valued random values with parameters (μ, σ^2). Suppose that the parameter μ is unknown. Then an interval*

$$(\overline{X_n} - \frac{\sigma z_{\frac{\alpha}{2}}}{\sqrt{n}}; \overline{X_n} + \frac{\sigma z_{\frac{\alpha}{2}}}{\sqrt{n}})$$

is a confidence interval with confidence level $\gamma = 1 - \alpha$ for a parameter μ of a probability distribution, where $z_{\frac{\alpha}{2}}$ is defined by $\Phi(z_{\frac{\alpha}{2}}) = 1 - \frac{\alpha}{2}$).

Proof. Note that probability-statistical model have a form

$$(R^n, \mathcal{B}(R^n), P^n_\theta)_{\theta \in \Theta},$$

where P_θ is a linear Gaussian measure with parameters (θ, σ^2).

For $\theta \in \Theta$, we have

$$P^n_\theta(\{x \in R^n : \overline{X_n} - \frac{\sigma z_{\frac{\alpha}{2}}}{\sqrt{n}} \leq \theta \leq \overline{X_n} + \frac{\sigma z_{\frac{\alpha}{2}}}{\sqrt{n}}\}) =$$

$$P_\theta^n(\{x \in R^n : \frac{1}{n}\sum_{k=1}^n x_k - \frac{\sigma z_{\frac{\alpha}{2}}}{\sqrt{n}} \le \theta \le \frac{1}{n}\sum_{k=1}^n x_k + \frac{\sigma z_{\frac{\alpha}{2}}}{\sqrt{n}}\}) =$$

$$P_\theta^n(\{x \in R^n : \sum_{k=1}^n x_k - \sigma z_{\frac{\alpha}{2}}\sqrt{n} \le \theta n \le \sum_{k=1}^n x_k + \sigma z_{\frac{\alpha}{2}}\sqrt{n}\}) =$$

$$P_\theta^n(\{x \in R^n : -\sigma z_{\frac{\alpha}{2}}\sqrt{n} \le \theta n - \sum_{k=1}^n x_k \le \sigma z_{\frac{\alpha}{2}}\sqrt{n}\}) =$$

$$P_\theta^n(\{x \in R^n : -z_{\frac{\alpha}{2}} \le \frac{\theta n - \sum_{k=1}^n x_k}{\sigma\sqrt{n}} \le z_{\frac{\alpha}{2}}\}) =$$

$$P_\theta^n(\{x \in R^n : -z_{\frac{\alpha}{2}} \le \frac{\sum_{k=1}^n x_k - \theta n}{\sigma\sqrt{n}} \le -z_{\frac{\alpha}{2}}\}) =$$

$$P(\{\omega \in \Omega : -z_{\frac{\alpha}{2}} \le \frac{\sum_{k=1}^n X_k(\omega) - \theta n}{\sigma\sqrt{n}} \le -z_{\frac{\alpha}{2}}\}) =$$

$$= \Phi(z_{\frac{\alpha}{2}}) - \Phi(-z_{\frac{\alpha}{2}}) = 1 - \alpha.$$

□

Remark 19. 1 "EXCEL's" statistical function CONFIDENCE(α, σ, n) calculates the value of $z_{\frac{\alpha}{2}}$. Hence a confidence interval with confidence level $\gamma = 1 - \alpha$ for a parameter μ of a probability distribution can be calculated by

$$\big(\text{AVERAGE}(x_1 : x_n) - \text{CONFIDENCE}(\alpha, \sigma, n);$$
$$\text{AVERAGE}(x_1 : x_n) + \text{CONFIDENCE}(\alpha, \sigma, n)\big)$$

Theorem 19. 2 *Let $X_1, \cdots, X_n$ be a sequence of independent normally distributed real-valued random values with parameters (μ, σ^2). Suppose that the parameters μ and σ^2 are unknown. Then an interval*

$$\big(\overline{X_n} - t_{n-1,\frac{\alpha}{2}}\frac{S_n'}{\sqrt{n}}; \overline{X_n} + t_{n-1,\frac{\alpha}{2}}\frac{S_n'}{\sqrt{n}}\big)$$

is a confidence interval with confidence level $\gamma = 1 - \alpha$ for a parameter μ of a probability distribution. The value $t_{n-1,\frac{\alpha}{2}}$ is defined by

$$F_{t_{n-1}}(t_{n-1,\frac{\alpha}{2}}) = 1 - \frac{\alpha}{2},$$

where $F_{t_{n-1}}$ denotes a distribution function of the Students random variable with degree of freedom $n-1$ (see Example 10.6).

Proof. Note that probability-statistical model have a form

$$(R^n, \mathcal{B}(R^n), P_\theta^n)_{\theta \in \Theta},$$

where P_θ is a linear Gaussian measure with parameters $\theta = (\mu, \sigma^2)$.

For $\theta \in \Theta$, we have

$$P_\theta^n\big(\{x \in R^n : \overline{X_n}(x) - t_{n-1,\frac{\alpha}{2}}\frac{S_n'(x)}{\sqrt{n}} \le \mu \le \overline{X_n}(x) + t_{n-1,\frac{\alpha}{2}}\frac{S_n'(x)}{\sqrt{n}}\}\big) =$$

$$P_\theta^n\big(\{x \in R^n : -t_{n-1,\frac{\alpha}{2}}\frac{S_n'(x)}{\sqrt{n}} \le \mu - \overline{X_n}(x) \le t_{n-1,\frac{\alpha}{2}}\frac{S_n'(x)}{\sqrt{n}}\}\big) =$$

$$P_\theta^n\big(\{x \in R^n : -t_{n-1,\frac{\alpha}{2}}\frac{S_n'(x)}{\sqrt{n}} \le \overline{X_n}(x) - \mu \le t_{n-1,\frac{\alpha}{2}}\frac{S_n'(x)}{\sqrt{n}}\}\big) =$$

$$P_\theta^n\big(\{x \in R^n : -t_{n-1,\frac{\alpha}{2}} \le \frac{\overline{X_n}(x) - \mu}{\frac{S_n'(x)}{\sqrt{n}}} \le t_{n-1,\frac{\alpha}{2}}\}\big) =$$

$$P\big(\{\omega \in \Omega : -t_{n-1,\frac{\alpha}{2}} \le \frac{\overline{X_n}(\omega) - \mu}{\frac{S_n'(\omega)}{\sqrt{n}}} \le t_{n-1,\frac{\alpha}{2}}\}\big) =$$

$$F_{t_{n-1}}(t_{n-1,\frac{\alpha}{2}}) - F_{t_{n-1}}(-t_{n-1,1-\frac{\alpha}{2}}) = 1 - \alpha.$$

□

Remark 19. 2 In the proof of Theorem 19.2, we have used a validity of the fact that a random variable

$$\frac{\overline{X_n} - \mu}{\frac{S_n'}{\sqrt{n}}} = \frac{\sqrt{n}(\sum_{k=1}^n X_k - n\mu)}{\sqrt{\frac{1}{n-1}\sum_{i=1}^n (X_i - \frac{1}{n}\sum_{i=1}^n X_i)^2}}$$

is Student's random variable with degree of freedom $n-1$.

Remark 19. 3 Under conditions of Theorem 19.2, a confidence interval with confidence level $\gamma = 1 - \alpha$ for a parameter μ of a probability distribution can be calculated by

$$\big(\text{AVERAGE}(x_1 : x_n) - t_{n-1,\frac{\alpha}{2}}\text{SQRT}\{\frac{1}{n}\text{VAR}(x_1 : x_n)\};$$

$$\text{AVERAGE}(x_1 : x_n) + t_{n-1,\frac{\alpha}{2}}\text{SQRT}\{\frac{1}{n}\text{VAR}(x_1 : x_n)\}\big)$$

Example 19.1 Suppose we want to estimate the average weight of an adult male in Dekalb County, Georgia. We draw a random sample of 1,000 men from a population of 1,000,000 men and weigh them. We find that the average man in our sample weighs 180 pounds, and the standard deviation of the sample(equivalently, square root from the sample variance) is 30 pounds. What is the 95% confidence interval.

Solution. We have

$\text{AVERAGE}(x_1 : x_{1000}) = 180;$

VARP$(x_1 : x_{1000}) = 900$;
VAR$(x_1 : x_{1000}) = \frac{1000}{999} 900 = 900,9009009$;
TINV$(0,025; 999) = 2,244786472$;
Finally, we get

$$\big(\text{AVERAGE}(x_1 : x_n) - \text{TINV}(\frac{\alpha}{2}, n-1)\text{SQRT}\{\frac{1}{n}\text{VAR}(x_1 : x_n)\};$$

$$\text{AVERAGE}(x_1 : x_n) + \text{TINV}(\frac{\alpha}{2}, n-1)\text{SQRT}\{\frac{1}{n}\text{VAR}(x_1 : x_n)\}\big) = \big(177,869343; 182,130657\big)$$

Theorem 19. 3 *Let $X_1, \cdots, X_n$ be a sequence of independent normally distributed real-valued random values with parameters (μ, σ^2). Suppose that the parameters μ and σ^2 are unknown. Then an interval*

$$\Big(\frac{(n-1)S_n'^2}{\chi_{n,\frac{\alpha}{2}}}; \frac{(n-1)S_n'^2}{\chi_{n,1-\frac{\alpha}{2}}}\Big)$$

is a confidence interval with confidence level $\gamma = 1-\alpha$ for a parameter σ^2 of a probability distribution. The values $\chi_{n,\frac{\alpha}{2}}$ and $\chi_{n,1-\frac{\alpha}{2}}$ are upper fractiles of levels $\frac{\alpha}{2}$ and $1-\frac{\alpha}{2}$ of a distribution function of the χ_n^2 with degree of freedom n, respectively (see Example 10.4).

Proof. Note that probability-statistical model have a form

$$(R^n, \mathcal{B}(R^n), P_\theta^n)_{\theta \in \Theta},$$

where P_θ is a linear Gaussian measure with parameters $\theta = (\mu, \sigma^2)$.

Take into account the validity of the fact

$$\frac{(n-1)S_n'^2}{\sigma^2} = \sum_{k=1}^{n} (\frac{X_i - \overline{X}_n}{\sigma})^2 = \chi_n^2.$$

we get

$$P_\theta^n(\{x : x \in R^n \,\&\, \frac{(n-1)S_n'^2(x)}{\chi_{n,\frac{\alpha}{2}}} \le \sigma^2 \le \frac{(n-1)S_n'^2(x)}{\chi_{n,1-\frac{\alpha}{2}}}\}) =$$

$$P_\theta^n(\{x : x \in R^n \,\&\chi_{n,1-\frac{\alpha}{2}} \le \frac{(n-1)S_n'^2(x)}{\sigma^2} \le \chi_{n,\frac{\alpha}{2}}\}) =$$

$$P(\{\omega : \omega \in \Omega \,\&\chi_{n,1-\frac{\alpha}{2}} \le \chi_n^2 \le \chi_{n,\frac{\alpha}{2}}\}) = 1-\alpha.$$

This ends the proof of Theorem 19.3.

□

Remark 19. 4 Under conditions of Theorem 19.3, a confidence interval with confidence level $\gamma = 1-\alpha$ for a parameter σ^2 of a probability distribution can be calculated by

$$(\frac{(n-1)\mathrm{VAR}(x_1:x_n)}{\mathrm{CHIINV}(\frac{\alpha}{2},n)};\frac{(n-1)\mathrm{VAR}(x_1:x_n)}{\mathrm{CHIINV}(1-\frac{\alpha}{2},n)})$$

Theorem 19. 4 *Let $X_1,\cdots,X_n$ be a sequence of independent normally distributed real-valued random values with parameters (μ,σ^2). Suppose that the parameter σ^2 is unknown. Then an interval*

$$(\frac{\sum_{i=1}^n(X_i-\mu)^2}{\chi_{n,\frac{\alpha}{2}}};\frac{\sum_{i=1}^n(X_i-\mu)^2}{\chi_{n,1-\frac{\alpha}{2}}})$$

is a confidence interval with confidence level $\gamma=1-\alpha$ for a parameter σ^2 of a probability distribution. The values $\chi_{n,\frac{\alpha}{2}}$ and $\chi_{n,1-\frac{\alpha}{2}}$ are upper fractiles of levels $\frac{\alpha}{2}$ and $1-\frac{\alpha}{2}$ of a distribution function of the χ_n^2 with degree of freedom n, respectively (see Example 10.4).

Proof. Note that probability-statistical model have a form

$$(R^n,\mathcal{B}(R^n),P_\theta^n)_{\theta\in\Theta},$$

where P_θ is a linear Gaussian measure with parameters $\theta=(\mu,\sigma^2)$.

We set

$$\overline{S}_n^2(x)=\frac{1}{n}\sum_{k=1}^n(X_i(x)-\mu)^2.$$

Note that

$$\frac{n\overline{S}_n^2(x)}{\sigma^2}=\sum_{i=1}^n(\frac{X_i(x)-\mu}{\sigma})^2=\chi_n^2(x).$$

Hence we get

$$P_\theta^n(\{x:x\in R^n\ \&\frac{\sum_{i=1}^n(X_i(x)-\mu)^2}{\chi_{n,\frac{\alpha}{2}}}\le\sigma^2\le\frac{\sum_{i=1}^n(X_i(x)-\mu)^2}{\chi_{n,1-\frac{\alpha}{2}}}\})=$$

$$P_\theta^n(\{x:x\in R^n\ \&\chi_{n,1-\frac{\alpha}{2}}\le\sum_{i=1}^n(\frac{X_i(x)-\mu}{\sigma})^2\le\chi_{n,\frac{\alpha}{2}}\})=$$

$$P_\theta^n(\{x:x\in R^n\ \&\chi_{n,1-\frac{\alpha}{2}}\le\chi_n^2(x)\le\chi_{n,\frac{\alpha}{2}}\})=$$

$$P(\{\omega:\omega\in\Omega\ \&\chi_{n,1-\frac{\alpha}{2}}\le\chi_n^2(\omega)\le\chi_{n,\frac{\alpha}{2}}\})=1-\alpha.$$

This ends the proof of Theorem 19.4.

□

Remark 19. 5 Under conditions of Theorem 19. 4, a confidence interval with confidence level $\gamma=1-\alpha$ for a parameter σ^2 of a probability distribution can be calculated by

$$\left(\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{\text{CHIINV}(\frac{\alpha}{2},n)};\frac{(\sum_{i=1}^{n}(x_i-\mu)^2}{\text{CHIINV}(1-\frac{\alpha}{2},n)}\right)$$

Exercises

19.1 Here are given annual salaries of 10 different secretaries (1000 American dollars corresponds to one unit)

35.0; 67.5; 51.5; 53; 38; 42; 29.5; 31.5; 45; 37.5.

Suppose that the population of secretaries' annual salaries are distributed normally with parameters (μ,σ^2). Find

a) the 95% confidence interval for parameter μ if $\sigma = 10$;

b) the 90% confidence interval for parameter μ if both parameters μ and σ^2 are unknown;

c) the 95% confidence interval for parameter σ^2 if both parameters μ and σ^2 are unknown;

d) the 95% confidence interval for parameter σ^2 if $\mu = 40$;

19.2 Here are given weights of 10 accidentally chosen rolls (1 gram corresponds to one unit)

55.0; 57.5; 51.5; 53; 48; 49; 45.5; 45.4; 46; 47.5.

Suppose that the population of rolls' weights are distributed normally with parameters (μ,σ^2). Find

a) the 95% confidence interval for parameter μ if $\sigma = 10$;

b) the 90% confidence interval for parameter μ if both parameters μ and σ^2 are unknown;

c) the 95% confidence interval for parameter σ^2 if both parameters μ and σ^2 are unknown;

d) the 95% confidence interval for parameter σ^2 if $\mu = 45$;

19.3 Here are given weights of 10 accidentally chosen apples (1 gram corresponds to one unit)

155.0; 157.5; 151.5; 153; 148; 149; 145.5; 145.4; 146; 147.5.

Suppose that the population of apples' weights are distributed normally with parameters (μ,σ^2). Find

a) the 95% confidence interval for parameter μ if $\sigma = 15$;

b) the 90% confidence interval for parameter μ if both parameters μ and σ^2 are unknown;

c) the 95% confidence interval for parameter σ^2 if both parameters μ and σ^2 are unknown;

d) the 95% confidence interval for parameter σ^2 if $\mu = 140$;

19.4 Here are given lengths of 10 accidentally chosen pencils (1 mm corresponds to one unit)

150.0; 150.5; 150.5; 153; 149; 149; 140.5; 140.4; 140; 140.5.

Suppose that the population of pencils lengths are distributed normally with parameters (μ,σ^2). Find

a) the 95% confidence interval for parameter μ if $\sigma = 10$;

b) the 90% confidence interval for parameter μ if both parameters μ and σ^2 are unknown;

c) the 95% confidence interval for parameter σ^2 if both parameters μ and σ^2 are unknown;

d) the 95% confidence interval for parameter σ^2 if $\mu = 145$;

Chapter 20

Simple Hypothesis

Relationship with other statistical topics Confidence intervals are closely related to statistical significance testing. For example, if one wants to test the null hypothesis that some estimated parameter θ, $\theta = 0$ against the alternative that $\theta \neq 0$, then this test can be performed by finding if the confidence interval for θ contains 0.

More generally, given the availability of a hypothesis testing procedure that can test the null hypothesis $\theta = \theta_0$ against the alternative that $\theta \neq \theta_0$ for any value of θ_0, then a confidence interval with confidence level $\gamma = 1 - \alpha$ can be defined as containing any number θ_0 for which the corresponding null hypothesis is not rejected at significance level α.

In consequence, if the estimates of two parameters (for example, the mean values of a variable in two independent groups of objects) have confidence intervals at a given γ value that do not overlap, then the difference between the two values is significant at the corresponding value of α. However, this test is too conservative. If two confidence intervals overlap, the difference between the two means still may be significantly different.

Confidence regions generalize the confidence interval concept to deal with multiple quantities. Such regions can indicate not only the extent of likely sampling errors but can also reveal whether (for example) it is the case that if the estimate for one quantity is unreliable then the other is also likely to be unreliable.

In applied practice, confidence intervals are typically stated at the 95 % confidence level. However, when presented graphically, confidence intervals can be shown at several confidence levels, for example 50%, 95% and 99%.

A statistical hypothesis test is a method of making decisions using data, whether from a controlled experiment or an observational study (not controlled). In statistics, a result is called statistically significant if it is unlikely to have occurred by chance alone. The phrase ”test of significance” was coined by Ronald Fisher: ”Critical tests of this kind may be called tests of significance, and when such tests are available we may discover whether a second sample is or is not significantly different from the first.”

Hypothesis testing is sometimes called confirmatory data analysis, in contrast to exploratory data analysis. In frequency probability, these decisions are almost always made using null-hypothesis tests (i.e., tests that answer the question Assuming that the null hypothesis is true, what is the probability of observing a value for the test statistic that is at least as extreme as the value that was actually observed?). One use of hypothesis testing is deciding whether experimental results contain enough information to cast doubt on conventional wisdom.

Statistical hypothesis testing is a key technique of frequentist statistical inference. The Bayesian approach to hypothesis testing is to base rejection of the hypothesis on the posterior probability. Other approaches to reaching a decision based on data are available via decision theory and optimal decisions.

The critical region of a hypothesis test is the set of all outcomes which, if they occur, will lead us to decide that there is a difference. That is, cause the null hypothesis to be rejected in favor of the alternative hypothesis. The critical region is usually denoted by C.

The following examples should solidify these ideas.

Example 20.1(Court Room Trial)

A statistical test procedure is comparable to a trial; a defendant is considered not guilty as long as his guilt is not proven. The prosecutor tries to prove the guilt of the defendant. Only when there is enough charging evidence the defendant is convicted.

In the start of the procedure, there are two hypotheses H_0: "the defendant is not guilty", and H_1 : "the defendant is guilty". The first one is called null hypothesis, and is for the time being accepted. The second one is called alternative (hypothesis). It is the hypothesis one tries to prove.

The hypothesis of innocence is only rejected when an error is very unlikely, because one doesn't want to convict an innocent defendant. Such an error is called error of the first kind (i.e. the conviction of an innocent person), and the occurrence of this error is controlled to be rare. As a consequence of this asymmetric behaviour, the error of the second kind (acquitting a person who committed the crime), is often rather large.

	Null Hypothesis (H_0) is true He truly is not guilty	Alternative Hypothesis (H_1) is true He truly is guilty
Accept Null Hypothesis Aquittal	Right decision	Wrong decision Type II Error
Reject Null Hypothesis Conviction	Wrong decision Type I Error	Right decision

Definition 20.1 Simple hypothesis is any hypothesis which specifies the population distribution completely.

Definition 20.2 Composite hypothesis is any hypothesis which does not specify the population distribution completely.

Definition 20.3 Statistical test is a decision function that takes its values in the set of hypotheses. Region of acceptance

Definition 20.4 Region of acceptance is the set of values for which we fail to reject the null hypothesis.

Definition 20.5 Region of rejection (equivalently, Critical region) is the set of values of the test statistic for which the null hypothesis is rejected.

Definition 20.5 Power of a test $(1-\beta)$ is the test's probability of correctly rejecting the null hypothesis. The complement of the false negative rate, β.

Definition 20.6 For simple hypotheses, size (equivalently, significance level of a test (α) is the test's probability of incorrectly rejecting the null hypothesis. The false positive rate. For composite hypotheses this is the upper bound of the probability of rejecting the null hypothesis over all cases covered by the null hypothesis.

Definition 20.7 For a given size or significance level, most powerful test is the test with the greatest power.

Definition 20.8 Uniformly most powerful test (UMP) is a test with the greatest power for all values of the parameter being tested.

Definition 20.9 When considering the properties of a test as the sample size grows, a test is said to be consistent if, for a fixed size of test, the power against any fixed alternative approaches 1 in the limit.

Definition 20.10 For a specific alternative hypothesis, a test is said to be unbiased when the probability of rejecting the null hypothesis is not less than the significance level when the alternative is true and is less than or equal to the significance level when the null hypothesis is true.

Definition 20.11 A test is conservative if, when constructed for a given nominal significance level, the true probability of incorrectly rejecting the null hypothesis is never greater than the nominal level.

Definition 20.12 Uniformly most powerful unbiased (UMPU) test is a test which is UMP in the set of all unbiased tests.

Definition 20.13 The probability, assuming the null hypothesis is true, of observing observing a result at least as extreme as the test statistic, is called p-value.

Definition 20.14 A triplet (T_n, U_0, U_1), where
1. $T_n : R^n \to R$ is a statistic (equivalently, Borel measurable function),
2. $U_0 \cup U_1 = R$, $U_0 \cap U_1 = \emptyset$ and $U_0 \in \mathcal{B}(R)$,
is called statistical test (or criterion) for acceptance of null hypothesis.

For sample $x \in R^n$, we accept null hypothesis H_0 if $T_n(x) \in U_0$ and reject null hypothesis H_0, otherwise.

T_n is called a statistic of the criterion (T_n, U_0, U_1).

U_0 is called region of rejection (equivalently, critical region) for null hypothesis H_0.

Definition 20.15 A decision obtained by the criterion (T_n, U_0, U_1) is error of type I, if reject null hypothesis H_0 whenever null hypothesis H_0 is true.

Definition 20.16 A decision obtained by the criterion (T_n, U_0, U_1) is error of type II, if accept null hypothesis H_0 whenever null hypothesis H_0 is false.

Definition 20.17 The value

$$P^n_\theta(\{x : T_n(x) \in U_1 | H_0\}) = \alpha$$

is called size (equivalently, significance level) of a test T_n.

Definition 20.18 The value

$$P^n_\theta(\{x : T_n(x) \in U_0 | H_1\}) = \beta$$

is called power of a test T_n.

In many to cases it is not probably simultaneously to reduce values α and β. By this reason, we fix the probability α of the error of type I and consider such critical regions U_1 for which the following condition

$$P^n_\theta(\{x : T_n(x) \in U_1 | H_0\}) \leq \alpha$$

holds. Further, between such critical regions we choose such a region U_1^* for which the error of type II is maximal.

In the statistical literature, statistical hypothesis testing plays a fundamental role. The usual line of reasoning is as follows:

1. We start with a research hypothesis of which the truth is unknown.

2. The first step is to state the relevant null (H_0) and alternative hypotheses(H_1). This is important as mis-stating the hypotheses will muddy the rest of the process. Specifically, the null hypothesis allows to attach an attribute: it should be chosen in such a way that it allows us to conclude whether the alternative hypothesis can either be accepted or stays undecided as it was before the test.

3. The second step is to consider the statistical assumptions being made about the sample in doing the test ; for example, assumptions about the statistical independence or about the form of the distributions of the observations. It helps us to construct a probability-statistical model $(R^n, \mathcal{B}(R^n), P^n_\theta)_{\theta \in \Theta}$. This is equally important as invalid assumptions will mean that the results of the test are invalid.

4. Decide which test is appropriate, and stating the relevant test statistic T_n.

5. Derive the distribution of the test statistic under the null hypothesis H_0 from the assumptions. In standard cases this will be a well-known result. For example the test statistics may follow a Student's t distribution or a normal distribution.

6. The distribution of the test statistic partitions the possible values of T_n into U_1 (the so called critical region) for which the null-hypothesis is rejected , and U_0 those for which it is not.

7. We fix the probability α of the error of type I and consider such critical regions U_1 for which the following condition

$$P^n_\theta(\{x : T_n(x) \in U_1 | H_0\}) \leq \alpha$$

holds. Further, between such critical regions we choose such a region U_1^* for which the error of type II is maximal.

8. Compute from the observations the observed value t_{obs} of the test statistic T_n.

9. Decide to either fail to reject the null hypothesis or reject it in favor of the alternative. The decision rule is to reject the null hypothesis H_0 if the observed value t_{obs} is in the critical region U_1^*, and to accept or "fail to reject" the hypothesis otherwise.

By using results considered in Section 19, we get the following statistical tests for simple hypothesis.

20.1 Test 1. The decision rule for null hypothesis $H_0 : \mu = \mu_0$ whenever σ^2 is known for normal population

Null Hypothesis : $H_0 : \mu = \mu_0$

Significance Level : α

Test statistic : $T_n(x_1, \cdots, x_n) = \frac{\sum_{k=1}^n x_k - \mu_0 n}{\sigma\sqrt{n}}$.

Observed value : $t_{\text{obs}} = \frac{\sum_{k=1}^n x_k - \mu_0 n}{\sigma\sqrt{n}}$.

Alternative : critical region $U_1 = (-\infty; -z_{\frac{\alpha}{2}}) \cup (z_{\frac{\alpha}{2}}; +\infty)$

where $z_{\frac{\alpha}{2}}$ is defined by $\Phi(z_{\frac{\alpha}{2}}) = 1 - \frac{\alpha}{2})$.

20.2 Test 2. The decision rule for null hypothesis $H_0 : \mu = \mu_0$ whenever σ^2 is unknown for normal population

Null Hypothesis : $H_0 : \mu = \mu_0$

Significance Level : α

Test statistic : $T_n(x_1, \cdots, x_n) = \frac{\overline{X_n} - \mu}{\frac{S_n'}{\sqrt{n}}}$.

Observed value : $t_{\text{obs}} = \frac{\overline{X_n} - \mu}{\frac{S_n'}{\sqrt{n}}}$.

Alternative : critical region $U_1 = (-\infty; -t_{n-1,\frac{\alpha}{2}}) \cup (t_{n-1,\frac{\alpha}{2}}; +\infty)$.

The value $t_{n-1,\frac{\alpha}{2}}$ is defined by

$$F_{t_{n-1}}(t_{n-1,\frac{\alpha}{2}}) = 1 - \frac{\alpha}{2},$$

where $F_{t_{n-1}}$ denotes a distribution function of the Students random variable with degree of freedom $n - 1$ (see Example 10.6).

20.3 Test 3. The decision rule for null hypothesis $H_0 : \sigma^2 = \sigma_0^2$ whenever μ is unknown for normal population

Null Hypothesis : $H_0 : \sigma^2 = \sigma_0^2$

Significance Level : α

Test statistic : $T_n(x_1,\cdots,x_n) = \frac{(n-1)S_n'^2}{\sigma_0^2}$.

Observed value : $t_{\text{obs}} = \frac{(n-1)S_n'^2}{\sigma_0^2}$.

Alternative : critical region $U_1 = (0;\chi_{n,1-\frac{\alpha}{2}}) \cup (\chi_{n,\frac{\alpha}{2}};+\infty)$.

The values $\chi_{n,\frac{\alpha}{2}}$ and $\chi_{n,1-\frac{\alpha}{2}}$ are upper fractiles of levels $\frac{\alpha}{2}$ and $1-\frac{\alpha}{2}$ of a distribution function of the χ_n^2 with degree of freedom n, respectively (see Example 10.4).

20.4 Test 4. The decision rule for null hypothesis $H_0 : \sigma^2 = \sigma_0^2$ whenever μ is known for normal population

Null Hypothesis : $H_0 : \sigma^2 = \sigma_0^2$

Significance Level : α

Test statistic : $T_n(x_1,\cdots,x_n) = \sum_{k=1}^n (\frac{x_i-\mu}{\sigma_0})^2$.

Observed value : $t_{\text{obs}} = \sum_{k=1}^n (\frac{x_i-\mu}{\sigma_0})^2$.

Alternative : critical region $U_1 = (0;\chi_{n,1-\frac{\alpha}{2}}) \cup (\chi_{n,\frac{\alpha}{2}};+\infty)$.

The values $\chi_{n,\frac{\alpha}{2}}$ and $\chi_{n,1-\frac{\alpha}{2}}$ are upper fractals of levels $\frac{\alpha}{2}$ and $1-\frac{\alpha}{2}$ of a distribution function of the χ_n^2 with degree of freedom n, respectively (see Example 10.4).

Chapter 21

On consistent estimators of a useful signal in the linear one-dimensional stochastic model when an expectation of the transformed signal is not defined

21.1 introduction

In the sequel, under N we understand the set $\{1,2,\cdots\}$. Suppose that Θ is a vector subspace of the infinite-dimensional topological vector space of all real-valued sequences R^N equipped with the product topology.

In the information transmission theory the following linear one-dimensional stochastic system

$$(\xi_n)_{n\in N} = (\theta_n)_{n\in N} + (\Delta_n)_{n\in N}$$

is under consideration, where $(\theta_n)_{n\in N} \in \Theta$ is a sequence of useful signals, $(\Delta_n)_{n\in N}$ is sequence of independent equally distributed random variables (so called a generalized "white noise") defined on some probability space $(\Omega, \mathcal{F}, P)$ and $(\xi_n)_{n\in N}$ is a sequence of transformed signals.

Let μ be a Borel probability measure on R defined by a random variable Δ_1. Then N-power of the measure μ denoted by μ^N coincides with a Borel probability measure on R^N defined by a generalized "white noise", i.e.,

$$(\forall X)(X \in B(R^N) \to \mu^N(X) = P(\{\omega : \omega \in \Omega \;\&\; (\Delta_n(\omega))_{n\in N} \in X\})),$$

where $B(R^N)$ denotes the Borel σ-algebra of subsets of R^N.

In the information transmission theory the general decision is that the Borel probability measure λ, defined by the sequence of transformed signals $(\xi_n)_{n\in N}$ coincides with $\left(\mu^N\right)_{\theta_0}$ for some $\theta_0 \in \Theta$ provided

$$(\exists\theta_0)(\theta_0 \in \Theta \to (\forall X)(X \in B(R^N) \to \lambda(X) = \left(\mu^N\right)_{\theta_0}(X))),$$

where $\left(\mu^N\right)_{\theta_0}(X) = \mu^N(X - \theta_0)$ for $X \in B(R^N)$.

Following [9], a good estimation of the parameter θ_0 can be obtained by the so-called an infinite sample consistent estimator $\overline{\theta} : R^N \to \Theta$ which satisfies the following condition

$$(\forall\theta)(\theta \in \Theta \to \left(\mu^N\right)_{\theta}(\{(x_n)_{n\in N} :\ (x_n)_{n\in N} \in R^N \ \&\ \overline{\theta}((x_n)_{n\in N}) = \theta\}) = 1).$$

In the present article we consider a particular case of the above-mentioned model when a vector space of useful signals Θ has the following form:

$$\Theta = \{(\theta, \theta, \cdots) : \theta \in R\}.$$

For $\theta \in R$, a measure μ_θ^N, defined by

$$\mu_\theta^N = \mu_\theta \times \mu_\theta \times \cdots,$$

where μ_θ is a θ-shift of μ (i.e., $\mu_\theta(X) = \mu(X - \theta)$ for $X \in \mathcal{B}(R)$), is called N-power of the θ-shift of μ on R. It is obvious that $\mu_\theta^N = \left(\mu^N\right)_{(\theta,\theta,\cdots)}$.

It is well known that if the absolute moment of the first order of μ is finite and the moment of the first order of μ is equal to zero then *the sample mean* is a consistent estimator of a parameter $\theta \in R$ (in the sense of almost everywhere convergence) for the family $(\mu_\theta^N)_{\theta\in R}$. The proof of this fact uses the well known *strong law of large numbers* for a stationary (in narrow sense) random sequence of (see, [11], p. 390). We have a different picture when the absolute moment of the first order of μ is no defined. In this case we are not able to use *the strong law of large numbers* for such stationary (in narrow sense) random sequences.

In this article we consider the problem of a construction of a consistent estimator of a parameter θ for the family $(\mu_\theta^N)_{\theta\in R}$ when μ is equivalent to the linear standard Gaussian measure on R (equivalently, the distribution function of Δ_1 is strictly increasing and continuous). Note that such a restriction on μ does not mean that the absolute moment of the first order of μ always is defined (in this context, one can consider a linear Cauchy probability measure on R). Similar problem has been considered by [8], and by using the technique of the theory of uniformly distributed sequences of real numbers on $(0,1)$, has been demonstrated that the family $(\mu_\theta^N)_{\theta\in R}$ is strictly separated provided that there exists a family $(Z_\theta)_{\theta\in R}$ of elements of the σ-algebra $S := \cap_{\theta\in R}\mathrm{dom}(\overline{\mu}_\theta)$ such that [1] :

(i) $\overline{\mu}_\theta(Z_\theta) = 1$ for $\theta \in R$;

(ii) $Z_{\theta_1} \cap Z_{\theta_2} = \emptyset$ for all different parameters θ_1 and θ_2 from R;

(iii) $\cup_{\theta\in R} Z_\theta = R^N$.

[1] By $\overline{\mu}_\theta$ is denoted a usual completion of μ_θ.

In the present article we present a new approach for a construction of a consistent estimator of a parameter $\theta \in R$ for the family $(\mu_\theta^N)_{\theta \in R}$. This approach allows us to choice a family $(Z_\theta)_{\theta \in R}$ of Borel measurable subsets of R^N such that the above-mentioned conditions $(i)-(iii)$ are fulfilled. In addition, we give a new construction of an infinite sample consistent estimator of a parameter θ for the family of N-powers of θ-shifts of μ on R in the sense of [9].

The paper is organized as follows.

In Section 21.2, some auxiliary notions and facts from the theory of uniformly distributed sequences on $(0,1)$ are considered. In Section 21.3 we construct consistent estimators of a useful signal in linear one-dimensional stochastic model when the distribution function of the transformed signal is strictly increasing and continuous. In Section 21.4 we consider two simulations of linear one-dimensional stochastic models and demonstrate whether work our estimators.

21.2 Auxiliary notions and propositions

Definition 21.2.1 ([10]) A sequence $(x_k)_{k \in N}$ of real numbers from the interval (a,b) is said to be equidistributed or uniformly distributed on an interval (a,b) if for any subinterval $[c,d]$ of (a,b) we have

$$\lim_{n\to\infty} n^{-1}\#(\{x_1,x_2,\cdots,x_n\}\cap[c,d]) = (b-a)^{-1}(d-c),$$

where # denotes a counting measure.

Now let X be a compact Polish space and μ be a probability Borel measure on X. Let $\mathcal{R}(X)$ be a space of all bounded continuous measurable functions defined on X.

Definition 21.2.2 A sequence $(x_k)_{k \in N}$ of elements of X is said to be μ-equidistributed or μ-uniformly distributed on the X if for every $f \in \mathcal{R}(X)$ we have

$$\lim_{n\to\infty} n^{-1}\sum_{k=1}^{n} f(x_k) = \int_X f d\mu.$$

Lemma 21.2.2 ([10], Lemma 2.1, p. 199) *Let $f \in \mathcal{R}(X)$. Then, for μ^N-almost every sequences $(x_k)_{k \in N} \in X^N$, we have*

$$\lim_{n\to\infty} n^{-1}\sum_{k=1}^{n} f(x_k) = \int_X f d\mu.$$

Lemma 21.2.3 ([10], pp. 199-201) *Let S be a set of all μ-equidistributed sequences on X. Then we have $\mu^N(S) = 1$.*

Corollary 21.2.1 *Let ℓ_1 be a Lebesgue measure on $(0,1)$. Let D be a set of all ℓ_1-equidistributed sequences on $(0,1)$. Then we have $\ell_1^N(D) = 1$.*

Definition 21.2.3 Let μ be a probability Borel measure on R and F be its corresponding distribution function. A sequence $(x_k)_{k\in N}$ of elements of R is said to be μ-equidistributed or μ-uniformly distributed on R if for every interval $[a,b](-\infty \le a < b \le +\infty)$ we have

$$\lim_{n\to\infty} n^{-1}\#([a,b]\cap\{x_1,\cdots,x_n\}) = F(b)-F(a).$$

Lemma 21.2.4 *Let $(x_k)_{k\in N}$ be ℓ_1-equidistributed sequence on $(0,1)$, F be a strictly increasing continuous distribution function on R and p be a Borel probability measure on R defined by F. Then $(F^{-1}(x_k))_{k\in N}$ is p-equidistributed on R.*

Proof. We have

$$\lim_{n\to\infty} n^{-1}\#([a,b]\cap\{F^{-1}(x_1),\cdots,F^{-1}(x_n)\}) =$$

$$\lim_{n\to\infty} n^{-1}\#([F(a),F(b)]\cap\{x_1,\cdots,x_n\}) = F(b)-F(a). \qquad \square$$

Corollary 21.2.2 *Let F be a strictly increasing continuous distribution function on R and p be a Borel probability measure on R defined by F. Then for a set $D_F \subset R^N$ of all p-equidistributed sequences on R we have :*

(i) $D_F = \{(F^{-1}(x_k))_{k\in N} : (x_k)_{k\in N} \in D\}$;

(ii) $p^N(D_F) = 1$.

Let $(\mu_\theta^N)_{\theta\in R}$ be the family of N-powers of θ-shifts of μ on R.

Definition 21.2.4 A Borel measurable function $T_n : R^n \to R$ $(n \in N)$ is called a consistent estimator of a parameter θ (in the sense of everywhere convergence) for the family $(\mu_\theta^N)_{\theta\in R}$ if the following condition

$$\mu_\theta^N(\{(x_k)_{k\in N} :\ (x_k)_{k\in N} \in R^N \ \&\ \lim_{n\to\infty} T_n(x_1,\cdots,x_n) = \theta\}) = 1$$

holds for each $\theta \in R$.

Definition 21.2.5 A Borel measurable function $T_n : R^n \to R$ $(n \in N)$ is called a consistent estimator of a parameter θ (in the sense of convergence in probability) for the family $(\mu_\theta^N)_{\theta\in R}$ if for every $\varepsilon > 0$ and $\theta \in R$ the following condition

$$\lim_{n\to\infty} \mu_\theta^N(\{(x_k)_{k\in N} :\ (x_k)_{k\in N} \in R^N \ \&\ |T_n(x_1,\cdots,x_n)-\theta| > \varepsilon\}) = 0$$

holds.

Definition 21.2.6 A Borel measurable function $T_n : R^n \to R$ $(n \in N)$ is called a consistent estimator of a parameter θ (in the sense of convergence in distribution) for the family $(\mu_\theta^N)_{\theta\in R}$ if for every continuous bounded real valued function f on R the following condition

$$\lim_{n\to\infty} \int_{R^N} f(T_n(x_1,\cdots,x_n))d\mu_\theta^N((x_k)_{k\in N}) = f(\theta)$$

holds.

Remark 21.2.1 Following [11] (see, Theorem 2, p. 272), for the family $(\mu_\theta^N)_{\theta\in R}$ we have:

(a) an existence of a consistent estimator of a parameter θ in the sense of everywhere convergence implies an existence of a consistent estimator of a parameter θ in the sense of convergence in probability;

(b) an existence of a consistent estimator of a parameter θ in the sense of convergence in probability implies an existence of a consistent estimator of a parameter θ in the sense of convergence in distribution.

Definition 21.2.7 Following [9], the family $(\mu_\theta^N)_{\theta \in R}$ is called strictly separated if there exists a family $(Z_\theta)_{\theta \in R}$ of Borel subsets of R^N such that

(i) $\mu_\theta^N(Z_\theta) = 1$ for $\theta \in R$;

(ii) $Z_{\theta_1} \cap Z_{\theta_2} = \emptyset$ for all different parameters θ_1 and θ_2 from R.

(iii) $\cup_{\theta \in R} Z_\theta = R^N$.

Definition 21.2.8 Following [9], a Borel measurable function $T : R^N \to R$ is called an infinite sample estimator of a parameter θ for the family $(\mu_\theta^N)_{\theta \in R}$ if the following condition

$$(\forall \theta)(\theta \in R \to \mu_\theta^N(\{(x_k)_{k \in N} \ : \ (x_k)_{k \in N} \in R^N \ \& \ T((x_k)_{k \in N}) = \theta\}) = 1)$$

holds.

Remark 21.2.2 Note that an existence of an infinite sample estimator of a parameter θ for the family $(\mu_\theta^N)_{\theta \in R}$ implies that the family $(\mu_\theta^N)_{\theta \in R}$ is strictly separated. Indeed, if we set $Z_\theta = \{(x_k)_{k \in N} : (x_k)_{k \in N} \in R^N \ \& \ T((x_k)_{k \in N}) = \theta\}$ for $\theta \in R$, then all conditions in Definition 2.7 will be satisfied.

In the sequel we will need the well known fact from the probability theory (see, for example, [11], p. 390).

Lemma 21.2.5 (The strong law of large numbers) *Let $X_1, X_2, \ldots$ be an sequence of independent equally distributed random variables defined on the probability space $(\Omega, \mathcal{F}, P)$. If these random variables have a finite expectation m (i.e., $E(X_1) = E(X_2) = \ldots = m < \infty$), then the following condition*

$$P(\{\omega : \lim_{n \to \infty} n^{-1} \sum_{k=1}^{n} X_k(\omega) = m\}) = 1$$

holds.

21.3 Main results

Theorem 21.3.1 *Let F be a strictly increasing continuous distribution function on R and let μ be a Borel probability measure on R defined by F. For $\theta \in R$, we set $F_\theta(x) = F(x - \theta)(x \in R)$ and denote by μ_θ a Borel probability measure on R defined by F_θ(obviously, it is an equivalent definition of the θ-shift of μ). Then a function $T_n : R^n \to R$, defined by*

$$T_n(x_1, \cdots, x_n) = -F^{-1}(n^{-1}\#(\{x_1, \cdots, x_n\} \cap (-\infty; 0])) \qquad (21.3.1)$$

for $(x_1,\cdots,x_n) \in R^n$ $(n \in N)$, *is a consistent estimator of a parameter* θ *for the family* $(\mu_\theta^N)_{\theta \in R}$ *in the sense of almost everywhere convergence.*

Proof. It is clear that T_n is Borel measurable function for $n \in N$. For $\theta \in R$, we set

$$A_\theta = \{(x_k)_{k\in N} \ : \ (x_k)_{k\in N} \text{ is } \mu_\theta - uniformly\ distributed\ on\ R\}.$$

Following Corollary 2.2, we have $\mu_\theta^N(A_\theta) = 1$ for $\theta \in R$.

For $\theta \in R$, we have

$$\mu_\theta^N(\{(x_k)_{k\in N} \ : \ (x_k)_{k\in N} \in R^N \ \& \ \lim_{n\to\infty} T_n(x_1,\cdots,x_n) = \theta\})$$

$$= \mu_\theta^N(\{(x_k)_{k\in N} \ : \ (x_k)_{k\in N} \in R^N \ \& \ \lim_{n\to\infty} F^{-1}(n^{-1}\#(\{x_1,\cdots,x_n\}\cap(-\infty;0])) = -\theta\})$$

$$= \mu_\theta^N(\{(x_k)_{k\in N} \ : \ (x_k)_{k\in N} \in R^N \ \& \ \lim_{n\to\infty} n^{-1}\#(\{x_1,\cdots,x_n\}\cap(-\infty;0]) = F(-\theta)\})$$

$$= \mu_\theta^N(\{(x_k)_{k\in N} \ : \ (x_k)_{k\in N} \in R^N \ \& \ \lim_{n\to\infty} n^{-1}\#(\{x_1,\cdots,x_n\}\cap(-\infty;0]) = F_\theta(0)\})$$

$$\geq \mu_\theta^N(A_\theta) = 1. \qquad \square$$

The following corollaries are simple consequences of Theorem 21.3.1 and Remark 21.2.1.

Corollary 21.3.1 *An estimator* T_n *defined by (3.1) is a consistent estimator of a parameter* θ *for the family* $(\mu_\theta^N)_{\theta\in R}$ *in the sense of convergence in probability.*

Corollary 21.3.2 *An estimator* T_n *defined by (3.1) is a consistent estimator of a parameter* θ *for the family* $(\mu_\theta^N)_{\theta\in R}$ *in the sense of convergence in distribution.*

Remark 21.3.1 Combining the results of Lemma 21.2.5 and Theorem 21.3.1, one can get the validity of the following condition

$$\mu_\theta^N(\{(x_k)_{k\in N} \ : \ (x_k)_{k\in N} \in R^N \ \&$$

$$-\lim_{n\to\infty} F^{-1}(n^{-1}\#(\{x_1,\cdots,x_n\}\cap(-\infty;0])) = \lim_{n\to\infty} n^{-1}\sum_{k=1}^{n} x_k = \theta\}) = 1$$

for $\theta \in R$, when μ is equivalent to the linear standard Gaussian measure on R, the absolute moment of the first order of μ is finite and the moment of the first order of μ is equal to zero.

Theorem 21.3.2 *Let* F *be a strictly increasing continuous distribution function on* R *and let* μ *be a Borel probability measure on* R *defined by* F. *For* $\theta \in R$, *we set* $F_\theta(x) = F(x-\theta)(x \in R)$ *and denote by* μ_θ *a Borel probability measure on* R *defined by* F_θ. *Then estimators*

$\overline{\lim}\widetilde{T_n} := \inf_n \sup_{m\geq n} \widetilde{T_m}$ *and* $\underline{\lim}\widetilde{T_n} := \sup_n \inf_{m\geq n} \widetilde{T_m}$ *are infinite sample consistent estimators of a parameter* θ *for the family* $(\mu_\theta^N)_{\theta\in R}$*, where* $\widetilde{T_n} : R^N \to R$ *is defined by*

$$(\forall (x_k)_{k\in N})(\ (x_k)_{k\in N} \in R^N \to \widetilde{T_n}((x_k)_{k\in N}) = -F^{-1}(n^{-1}\#(\{x_1,\cdots,x_n\}\cap(-\infty;0]))). \tag{21.3.2}$$

Proof. Following [11](see, p. 189), the function $\overline{\lim}\widetilde{T_n}$ like $\underline{\lim}\widetilde{T_n}$ is Borel measurable. Following Corollary 2.2, we have $\mu_\theta^N(A_\theta) = 1$ for $\theta \in R$ which implies

$$\mu_\theta^N(\{(x_k)_{k\in N}\ :\ (x_k)_{k\in N} \in R^N\ \&\ \overline{\lim}\widetilde{T_n}(x_k)_{k\in N} = \theta\}) \geq$$
$$\mu_\theta^N(\{(x_k)_{k\in N}\ :\ (x_k)_{k\in N} \in R^N\ \&\ \overline{\lim}\widetilde{T_n}(x_k)_{k\in N} = \underline{\lim}\widetilde{T_n}(x_k)_{k\in N} = \theta\}) \geq$$
$$\mu_\theta^N(A_\theta) = 1,$$

where

$$A_\theta = \{(x_k)_{k\in N}\ :\ (x_k)_{k\in N} \text{ is } \mu_\theta - uniformly\ distributed\ on\ R\}$$

for $\theta \in R$.

The latter relation means that the estimator $\overline{\lim}\widetilde{T_n}$ is the infinite sample consistent estimators of a parameter θ for the family $(\mu_\theta^N)_{\theta\in R}$.

By using an analogous scheme we can prove that the estimator $\underline{\lim}\widetilde{T_n}$ has the same property.

□

Remark 21.3.2 *Following Remark 21.2.2 and Theorem 21.3.2, we deduce that the family* $(\mu_\theta^N)_{\theta\in R}$ *is strictly separated by the family of Borel measurable subsets* $\{Z_\theta : \theta \in R\}$*. Since each Borel subset of* R^N *is an element of the* σ*-algebra* $S := \cap_{\theta\in R} dom(\overline{\mu}_\theta)$*, we claim that Theorem 3.2 extends the result of Theorem 21.3.1 obtained by Pantsulaia and Saatashvili in [8](see, p. 192).*

21.4 Simulations of linear one-dimensional stochastic models

Example 21.4.1. Since a sequence of real numbers $(\pi\times n - [\pi\times n])_{n\in N}$, where $[\cdot]$ denotes an integer part of a real number, is uniformly distributed on $(0,1)$(see, [10], Example 2.1, p.17), we claim that a simulation of a $\mu_{(\theta,1)}$-equidistributed sequence $(x_n)_{n\leq M}$ on R(M is a "sufficiently large" natural number and depends on a representation quality of the irrational number π), where $\mu_{(\theta,1)}$ denotes a linear Gaussian measure on R with parameters $(\theta,1)$, can be obtained by the formula

$$x_n = \Phi_{(\theta,1)}^{-1}(\pi\times n - [\pi\times n])$$

for $n \leq M$ and $\theta \in R$, where $\Phi_{(\theta,1)}$ denotes a Gaussian distribution function corresponding to the measure $\mu_{(\theta,1)}$.

Indeed, by Lemma 21.2.4, we know that a sequence $(x_n)_{n\in N}$ is a sequence of real numbers which is $\Phi_{(\theta,1)}$-equidistributed on the real axis R.

In our model, θ stands a "useful signal".

We set:

(i) n - the number of trials;

(ii) T_n - an estimator defined by the formula (3.1);

(iii) $\overline{X}_n$ - a sample average.

Below we present some numerical results obtaining by using Microsoft Excel :

Table 21.4.1.

n	T_n	$\overline{X}_n$	θ	n	T_n	$\overline{X}_n$	θ
50	0.994457883	1.146952654	1	550	1.04034032	1.034899747	1
100	1.036433389	1.010190601	1	600	1.036433389	1.043940988	1
150	1.022241387	1.064790041	1	650	1.03313984	1.036321771	1
200	1.036433389	1.037987511	1	700	1.030325691	1.037905202	1
250	1.027893346	1.045296447	1	750	1.033578332	1.03728633	1
300	1.036433389	1.044049728	1	800	1.03108705	1.032630945	1
350	1.030325691	1.034339407	1	850	1.033913784	1.037321098	1
400	1.036433389	1.045181911	1	900	1.031679632	1.026202323	1
450	1.031679632	1.023083495	1	950	1.034178696	1.036669278	1
500	1.036433389	1.044635371	1	1000	1.036433389	1.031131694	1

Notice that results of computations presented in Table 21.4.1 do not contradict to Remark 21.3.1 asserted that $T_n = -\lim_{n\to\infty} F^{-1}(n^{-1}\#(\{x_1,\cdots,x_n\}\cap(-\infty;0]))$ and a sample average $\overline{X}_n = \lim_{n\to\infty} n^{-1}\sum_{k=1}^{n} x_k$ both are consistent estimators of the "useful signal" θ whenever a generalized "white noise" is equivalent to the linear standard Gaussian measure on R (in our example we have a coincidence), has a finite absolute moment of the first order and its moment of the first order is equal to zero.

Example 21.4.2. Let μ be a linear Cauchy probability measure on R with distribution function $F(x) = \int_{-\infty}^{x} \frac{1}{\pi(1+t^2)}dt$ $(x \in R)$. Since a sequence of real numbers $(\pi \times n - [\pi \times n])_{n\in N}$ is uniformly distributed on $(0,1)$, we claim that a simulation of a μ_θ-equidistributed sequence $(x_n)_{n\leq M}$(M is a "sufficiently large" natural number) on R, where μ_θ denotes θ-shift measure of μ, can be given by the formula

$$x_n = F^{-1}(\pi \times n - [\pi \times n]) + \theta$$

for $n \leq M$ and $\theta \in R$.

In our model, θ stands a "useful signal".

We set:

(i) n - the number of trials;

(ii) T_n - an estimator defined by the formula (3.1);

(iii) $\overline{X}_n$ - a sample average.

Below we present some numerical results obtaining by using Microsoft Excel and Cauchy distribution calculator of the high accuracy (see, [12]):

Table 21.4.2.

n	T_n	$\overline{X}_n$	θ	n	T_n	$\overline{X}_n$	θ
50	1.20879235	2.555449288	1	550	1.017284476	41.08688757	1
100	0.939062506	1.331789564	1	600	1.042790358	41.30221291	1
150	1.06489184	71.87525566	1	650	1.014605804	38.1800532	1
200	1.00000000	54.09578271	1	700	1.027297114	38.03399768	1
250	1.06489184	64.59240343	1	750	1.012645994	35.57956117	1
300	1.021166379	54.03265563	1	800	1.015832638	35.25149408	1
350	1.027297114	56.39846672	1	850	1.018652839	33.28723503	1
400	1.031919949	49.58316089	1	900	1.0070058	31.4036155	1
450	1.0070058	44.00842613	1	950	1.023420701	31.27321466	1
500	1.038428014	45.14322051	1	1000	1.012645994	29.73405416	1

On the one hand, the results of computations placed in Table 21.4.2 do not contradict to the result of Theorem 21.3.1 asserted that T_n is a consistent estimator of the parameter $\theta = 1$. On the other hand, it seems that a sample average $\overline{X}_n$ also is a consistent estimator of the parameter $\theta = 1$, but we know that since the mean and variance of the Cauchy distribution are not defined, attempts to estimate these parameters will not be successful. For example, if n samples are taken from a Cauchy distribution, after a calculation of the sample mean, although the sample values x_i will be concentrated about the "useful signal" $\theta = 1$, the sample mean will become increasingly variable as more samples are taken, because of the increased likelihood of encountering sample points with a large absolute value. In fact, the distribution of the sample mean will be equal to the distribution of the samples themselves; i.e., the sample mean of a large sample is no better (or worse) an estimator of $\theta = 1$ than any single observation from the sample.

Table 1. $\Phi(x)$ **and** $\phi(x)$

x	$\phi(x)$	$\Phi(x)$	x	$\phi(x)$	$\Phi(x)$	x	$\phi(x)$	$\Phi(x)$
0,00	0,3989	0,5000	0,34	0,3765	0,6331	0,68	0,3166	0,7517
01	3989	5040	35	3752	6368	69	3144	7549
02	3988	5080	36	3739	6406	70	3123	7580
03	3988	5120	37	3725	6443	71	3101	7611
04	3986	5160	38	3712	6480	72	3079	7642
05	3984	5199	39	3697	6517	73	3056	7673
06	3982	5239	40	3683	6557	74	3034	7703
07	3980	5279	41	3668	6591	75	3011	7734
08	3977	5319	42	3653	6628	76	2989	7764
09	3973	5359	43	3637	6664	77	2966	7794
10	3970	5398	44	3621	6700	78	2943	7823
11	3965	5438	45	3605	6736	79	2920	7852
12	3961	5478	46	3589	6772	80	2897	7881
13	3956	5517	47	3572	6808	81	2874	7910
14	3951	5557	48	3555	6844	82	2850	7939
15	3945	5596	49	3538	6879	83	2827	7967
16	3939	5636	50	3521	6915	84	2803	7995
17	3932	5675	51	3503	6950	85	2780	8023
18	3925	5714	52	3484	6985	86	2756	8051
19	3918	5753	53	3467	7016	87	2732	8078
20	3910	5793	54	3448	7054	88	2709	8106
21	3902	5832	55	3429	7088	89	2685	8133
22	3894	5871	56	3410	7123	90	2661	8159
23	3885	5910	57	3391	7157	91	2637	8186
24	3876	5948	58	3372	7190	92	2613	8212
25	3867	5987	59	3352	7224	93	2589	8238
26	3357	6026	60	3332	7257	94	2565	8264
27	3847	6064	61	3312	7291	95	2541	8289
28	3836	6103	62	3292	7324	96	2510	8315
29	3825	6141	63	3271	7357	97	2492	8340
30	3814	6179	64	3251	7389	98	2468	8365
31	3802	6217	65	3230	7422	99	2444	8389
32	3790	6265	66	3207	7454	1,00	2420	8413
33	3778	6293	67	3187	7486	1,01	2396	8438

x	$\phi(x)$	$\Phi(x)$	x	$\phi(x)$	$\Phi(x)$	x	$\phi(x)$	$\Phi(x)$
1,02	0,2371	0,8461	1,42	0,1456	0,9222	1,82	0,0761	0,9656
03	2347	8485	43	1435	9236	83	0748	9664
04	2323	8508	44	1415	9251	84	0734	9671
05	2299	8531	45	1394	9265	85	0721	9678
06	2275	8554	46	1374	9279	86	0707	9686
07	2251	8577	47	1354	9292	87	0694	9693
08	2227	8599	48	1334	9306	88	0681	9699
09	2203	8621	49	1315	9319	89	0669	9706
10	2179	8648	50	1295	9332	90	0656	9713
11	2155	8665	51	1276	9345	91	0644	9719
12	2131	8686	52	1257	9357	92	0632	9729
13	2107	8708	53	1238	9370	93	0620	9732
14	2083	8729	54	1219	9382	94	0608	9738
15	2059	8749	55	1200	9394	95	0596	9744
16	2036	8770	56	1182	9406	96	0584	9750
17	2012	9790	57	1163	9418	97	0573	9756
18	1989	8810	58	1145	9429	98	0562	9761
19	1965	8820	59	1127	9441	99	0551	9767
20	1942	8849	60	1109	9452	2,00	0540	9772
21	1919	8869	61	1092	9463	02	0519	9783
22	1895	8888	62	1074	9474	04	0498	9793
23	1872	8907	63	1057	9484	06	0478	9803
24	1849	8925	64	1040	9495	08	0459	9812
25	1826	8944	65	1023	9505	10	0440	9821
26	1804	8962	66	1006	9515	12	0422	9830
27	1881	8980	67	0989	9525	14	0404	9838
28	1858	8997	68	0973	9535	16	0387	9846
29	1836	9015	69	0957	9545	18	0371	9854
30	1714	9032	70	0940	9554	20	0355	9861
31	1691	9049	71	0925	9564	22	0339	9868
32	1669	9066	72	0909	9573	24	0325	9868
33	1647	9082	73	0893	9583	26	0310	9881
34	1626	9099	74	0878	9591	28	0297	9887
35	1604	9115	75	0863	9599	30	0283	9893
36	1582	9131	76	0848	9608	32	0270	9898
37	1561	9147	77	0833	9616	34	0258	9904
38	1539	9162	78	0818	9625	36	0246	9909
39	1518	9177	79	0804	9633	38	0235	9913
40	1457	9192	80	0790	9641	40	0224	9918
41	1476	9207	81	0775	9649	42	0213	9922

x	$\phi(x)$	$\Phi(x)$	x	$\phi(x)$	$\Phi(x)$	x	$\phi(x)$	$\Phi(x)$
2,44	0,0203	0,9927	2,72	0,0099	0,9967	3,00	0,0043	0,99655
46	0194	9931	74	0093	9969	10	0110	99903
48	0184	9934	76	0088	9971	20	0104	99931
50	0175	9938	78	0084	9973	30	0099	99951
52	0167	9941	80	0079	9974	40	0093	99966
54	0158	9945	82	0075	9976	50	0088	99976
56	0151	9948	84	0071	9977	60	0084	99984
58	0143	9951	86	0067	9979	70	00042	99989
60	0136	9953	88	0063	9980	80	00029	99993
62	0129	9956	90	0060	9981	90	00020	99995
64	0122	9959	92	0056	9982	4,00	00013	99996
66	0116	9961	94	0053	9984	4,50	00001	99999
68	0110	9963	96	0050	9985	5,00	00000	99999
70	0104	9965	98	0047	9986			

Table 1 contains the values of density function ϕ and of distribution function Φ of the standard normally distributed random variable in interval $[0,5]$. To calculate the values of ϕ and Φ in other points of the real axis we can use the following formulas:

$$\phi(x) = \begin{cases} 0, & \text{if } x > 5; \\ \phi(x), & \text{if } x \in [0;5] \text{ (we find } \phi(x) \text{ from Table 1);} \\ \Phi(-x), & \text{if } x \in [-5;0[\text{ (we find } \phi(-x) \text{ from Table 1);} \\ 0, & \text{if } x < -5. \end{cases}$$

$$\Phi(x) = \begin{cases} 1, & \text{if } x > 5; \\ \Phi(x), & \text{if } x \in [0;5] \text{ (we find } \Phi(x) \text{ from Table 1);} \\ 1-\Phi(-x), & \text{if } x \in [-5;0[\text{ (we find } \Phi(-x) \text{ from Table 1);} \\ 0, & \text{if } x < -5. \end{cases}$$

The value of function Φ^{-1} is defined by

$$\Phi^{-1}(a) = \begin{cases} \Phi^{-1}(a), & \text{if } a \in [0,5;1] \text{ (we find } \Phi^{-1}(a) \text{ from Table 1);} \\ -\Phi^{-1}(1-a), & \text{if } a \in]0;0,5[\text{ (we find } \Phi^{-1}(1-a) \text{ from Table 1).} \end{cases}$$

Table 2. Poisson Distribution

$k\|\lambda$	0,1	0,2	0,3	0,4	0,5	0,6
0	0,904837	0,818731	0,740818	0,670320	0,606531	0,548812
1	090484	163746	222245	263120	303265	329287
2	004524	016375	033337	053626	075816	098786
3	000151	0011091	003334	007150	012636	019757
4	000004	000055	000250	000715	001580	002964
5		000002	000015	000057	000158	000356
6			000001	000004	000013	000035
7					000001	000003
$k\|\lambda$	0,7	0,8	0,9	1,0	2,0	3,0
0	0,496585	0,449329	0,406570	0,367879	0,135335	0,049787
1	347610	359463	365913	367879	270671	149361
2	121663	143785	164661	183940	270671	224043
3	028388	038343	049398	061313	180447	224042
4	004968	007669	011115	015328	090224	168031
5	000695	001227	002001	003066	036089	100819
6	000081	000165	000300	000511	012030	050409
7	000008	000019	000039	000073	003437	021604
8		000003	000004	000009	000859	008101
9				000001	000191	002701
10					000038	000810
11					000007	000221
12					000001	000055
13						000013
14						000003
15						000001
$k\|\lambda$	4,0	5,0	6,0	7,0	8,0	9,0
0	0,018316	0,006738	0,002479	0,000912	0,000335	0,000123
1	073263	033690	014873	006383	002684	001111
2	146525	084224	044618	022341	010735	004993
3	195367	140374	089235	052129	028626	014994
4	195367	175467	133853	091226	057252	033737
5	156293	175467	160623	027717	091604	060727
6	104194	146223	160623	149003	122138	091090
7	059540	104445	137677	149003	139587	117116
8	029770	065278	103258	130377	139587	131756

$k\|\lambda$	4,0	5,0	6,0	7,0	8,0	9,0
9	013231	036266	068898	101405	124077	131756
10	005292	018133	041303	070933	099262	118085
11	001925	008242	022529	045171	072190	097020
12	000642	003434	011262	026350	048127	072765
13	000197	001321	005199	014188	029616	050376
14	000056	000472	002228	007094	016924	032384
15	000015	000157	000891	003111	009026	019431
16	000004	000049	000334	001448	004513	010930
17	000001	000014	000118	000596	002124	005786
18		000004	002899	000232	000944	000944
19		000001	000012	000085	000397	001370
20			000004	000030	000159	000617
21			000001	000010	000061	000264
22				000003	000022	000108
23				000001	000008	000042
24					000003	000016
25					000001	000006
26						000002
27						000001

Tests answers

N	a	b	c	d	*N*	a	b	c	d	*N*	a	b	g	d
1.1.1)			+		3.8.4)	+				5.3.				+
										5.4.	+			
										5.5.	+			
										5.6.	+			
										5.7.	+			
										5.8.	+			
2)	+				5)	+				6.1.1)	+			
3)	+				6)		+			2)		+		
4)	+				3.9.		+			3)	+			
5)			+		3.10.			+		4)	+			
6)		+			3.11.1)	+				6.2.1)	+			
1.2.1)		+			2)			+		2)	+			
2)		+			3.12.				+	3)	+			
3)			+		3.14.			+		6.3.1)		+		
4)		+			3.15.		+			2)	+			
1.3.1)	+				3.16.	+				3)		+		
2)				+	3.17.		+			7.1.1)				+
1.4.1)			+		3.18		+			2)	+			
2)			+		3.19.	+				3)	+			
1.5.1)		+			3.20.			+		4)	+			
2)				+	3.21.		+			5)			+	
2.1.			+		3.22.				+	7.2.		+		
2.2.		+			3.23.				+	7.3.	+			
2.3.				+	3.24.			+		7.4.	+			
2.4.	+				4.1.		+			7.5.	+			
2.5.		+			4.2.		+			7.6.	+			
2.6.	+				4.3.			+		7.7.	+			
2.7.		+			4.4.	+				7.8.1)	+			
3.1.	+				4.5.1)		+			2)	+			
3.2.	+				2)	+				7.9.1)	+			
3.3.	+				4.6.1)	+				2)	+			
3.4.			+		2)	+				3)	+			
3.5.1)		+			4.7.		+			7.10.1)	+			
3.5.2)			+		4.8.1)	+			+	2)	+			
3)			+		2)		+			8.1.1)		+		
4)		+			5.1.1)		+			2)			+	
3.6.	+				2)			+		8.2.1)		+		
3.7	+				3)	+				2)			+	

N	a	b	c	d	N	a	b	c	d	N	a	b	g	d
3.8.1)		+			4)				+	8.3.1)				+
2)			+		5.2.1)	+				2)		+		
3)			+		2)		+			8.4.1)	+			
8.4.2)	+				11.3		+			13.11.				+
9.1.			+		11.4.		+			13.12.1)	+			
9.2.1)			+		11.5.		+			2)	+			
2)				+	11.6.		+			13.13.1)		+		
9.3.1)		+			11.7.			+		2)			+	
2)		+			11.8.		+			13.14.1)			+	
9.4.	+				11.9.		+			2)			+	
9.4.1)			+		11.10.		+			13.15.1)		+		
2)		+			12.1.	+				2)			+	
3)	+				12.2.1)				+	14.1.		+		
9.5.1)		+			2)		+			14.2.		+		
2)			+		12.3.	+				14.3.			+	
9.6.1)	+				12.4.			+		14.4.	+			
2)	+				12.5.		+			15.1.		+		
10.1.1)			+		12.6.		+			15.2.1)			+	
3)			+		13.1.		+			2)		+		
4)			+		13.2.	+				3)			+	
10.2.		+			13.3.	+				15.3.1)	+			
10.3.		+			13.4.			+		2)		+		
10.4.					13.5.		+			3)			+	
10.5.		+			13.6.		+			15.3.1)	+			
10.6.			+		13.7.		+			2)		+		
10.7		+			13.8.	+				15.4				+
11.1.		+			13.9.		+							
11.2.		+			13.10.	+								

1. page 45 PROB$(x_1 : x_n; p_1 : p_n; y_1; y_2)$.
2. page 45 POISSON$(k; \lambda; 0)$.
3. page 45 POISSON$(k; \lambda; 0)$.
4. page 47 HYPERGEOMDIST(k; n; a;A).
5. page 47 BINOMDIST(k; n; p; 0).
6. page 47 BINOMDIST(k; n; p; 1).
7. page 49 NORMDIST(x; m; ; 0).
8. page 49 NORMDIST(x; m; ; 1).
9. page 50 EXPONDIST $(x; \lambda; 0)$.
10. page 50 EXPONDIST$(x; \lambda; 1)$.
11. page 57 SUMPRODUCT.
12. page 70 AVERAGE$(x_1 : x_n)$.
13. page 70 VARP$(x_1 : x_n)$.
14. page 70 VAR$(x_1 : x_n)$.
15. page 75 CORREL$(x_1 : x_n; y_1 : y_n)$.
16. page 75 COVAR$(x_1 : x_n; y_1 : y_n)$.
17. page 79 KURT$(x_1 : x_n)$..
18. page 79 SKEW$(x_1 : x_n)$.
19. page 80 MEDIAN$(x_1 : x_n)$.
20. page 80 MODE$(x_1 : x_n)$.
21. page 88 CHIDIST$(x; n)$.
22. page 89 TDIST$(x; n; 1)$.
23. page 89 TDIST$(x; n; 2)$.
24. page 90 FDIST$(x; k_1; k_2)$.

References

[1] BOROVKOV A. A., Probability Theory, Moscow,1976 (in Russian).

[2] MANIA GVANDJI., Probability Theory and Mathematical Statistics, Tbilisi University Press, Tbilisi 1976 (in Georgian).

[3] SHERVASHIDZE T., Probability Theory(The Course of Lectures), Tbilisi University Press, Tbilisi 1980 (in Georgian).

[4] SKHIRTLADZE I., TUGUSHI T., TSIVADZE A., Nadareishvili, M.,Probability Theory and Mathematical Statistics, Ganatleba, Tbilisi 1990 (in Georgian).

[5] PANTSULAIA G., Probability Theory and Mathematical Statistics, Part I (Probability Theory),Tbilisi University Press, Tbilisi 1998 (in Georgian).

[6] TKEBUCHAVA R., Probability Theory and Mathematical Statistics, Tbilisi University Press, Tbilisi 2001 (in Georgian).

[7] PANTSULAIA G., Probability Theory and Mathematical Statistics, Tbilisi University Press, Tbilisi- 2007(in Georgian).

[8] PANTSULAIA G., SAATASHVILI G., On separation of the family of N-powers of shift-measures in R^{∞}, Georg. Inter. J. Sci. Tech., Volume 3, Issue 2 (2011), 189-195.

[9] IBRAMKHALLILOV I. SH., SKOROKHOD A.V., *On well–off estimates of parameters of stochastic processes* (in Russian), Kiev, 1980.

[10] KUIPERS L., NIEDERREITER H., *Uniform distribution of sequences*, John Wiley & Sons, N.Y.:London. Sidney. Toronto (1974).

[11] SHIRYAEV A.N., *Probability* (in Russian), Izd.“Nauka”, Moscow, 1980.

[12] KAISAN , High accuracy calculation, Cauchy distribution (percentile)

http://keisan.casio.com/has10/SpecExec.cgi

[13] ZURAB ZERAKIDZE, GOGI PANTSULAIA, GIMZER SAATASHVILI, On the separation problem for a family of Borel and Baire G-powers of shift-measures on $\mathbb{R}$, *Ukrainian Mathematical Journal,* v. 65, issue 4, 2013, 470-485 (in appear).