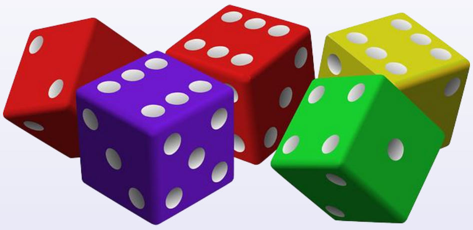
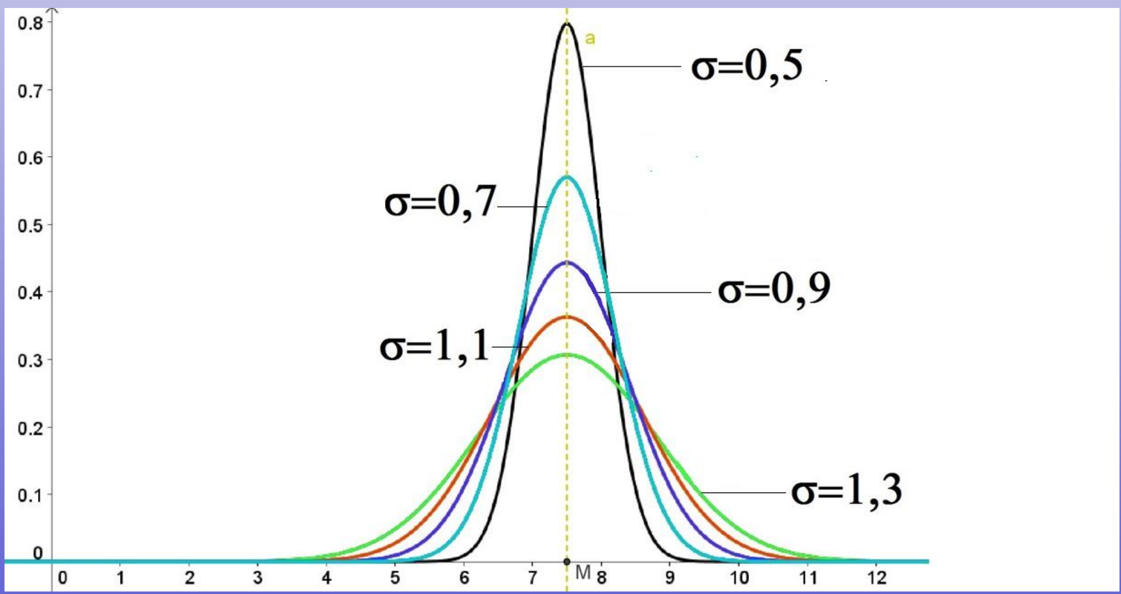
UNIVERSITATEA PEDAGOGICĂ DE STAT “ION CREANGĂ”

NEAGU Natalia



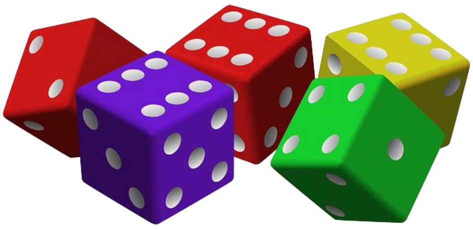
TEORIA PROBABILITĂȚILOR
ȘI
STATISTICA MATEMATICĂ



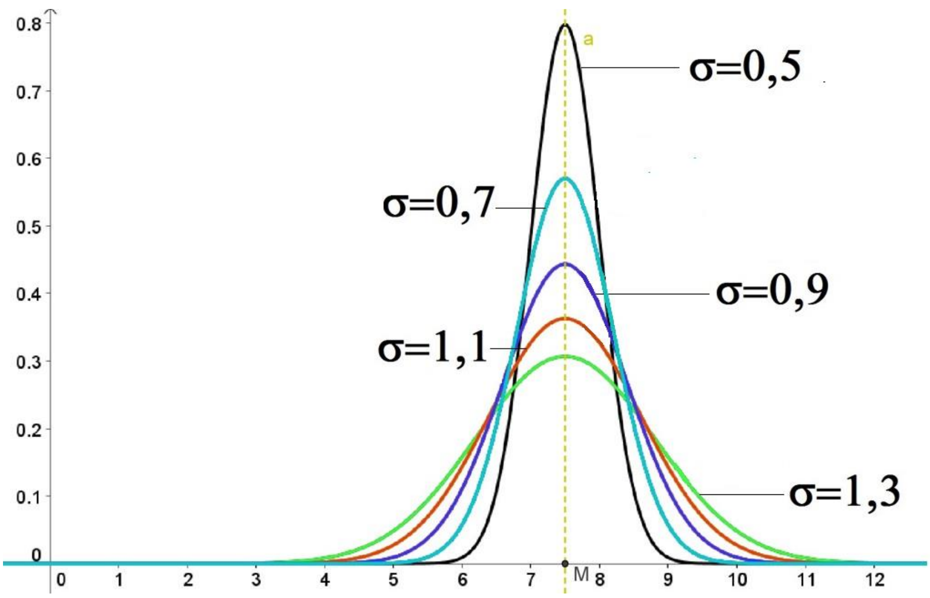
Chișinău, 2022

UNIVERSITATEA PEDAGOGICĂ DE STAT “ION CREANGĂ”

NEAGU Natalia



TEORIA PROBABILITĂȚILOR
ȘI
STATISTICA MATEMATICĂ



Chișinău, 2022

Aprobată de Senatul Universităţii Pedagogice de Stat „Ion Creangă” din Chișinău

SUPORT DE CURS
TEORIA PROBABILITĂȚILOR ȘI STATISTICA MATEMATICĂ

Recenzenți:

1. *Port Sergiu*, doctor, conferențiar universitar, Universitatea Pedagogică de Stat “Ion Creangă”
2. *Pricop Victor*, doctor, conferențiar universitar, Universitatea Tehnică a Moldovei

Machetarea computerizată: Neagu Natalia, doctor, lector universitar

Design copertă: Neagu Natalia, doctor, lector universitar

Neagu, Natalia
Teoria probabilităților și statistica matematică : Suport de curs / Neagu Natalia ; Universitatea Pedagogică de Stat "Ion Creangă". – Chişinău : S. n., 2023 (CEP UPSC). – 137 p. : fig., tab.

Bibliogr.: p. 131-133 (30 tit.). – 100 ex.

ISBN 978-9975-46-685-1.
519.2(075.8)
N 31

Centrul Editorial-Poligrafic al Universității Pedagogice de Stat „Ion Creangă”

CUPRINS

INTRODUCERE

Calculul probabilităților a apărut pe la mijlocul secolului al XVII-lea ca urmare a problemelor ce țin de jocurile de noroc.

Primele definiții fiind introduse de iluștrii savanți B. Pascal (1623-1662), P. de Fermat (1601-1665), Ch. Huygens (1629-1695) și în special de J. Bernoulli (1654-1705). De asemenea un rol important în dezvoltarea metodelor analitice ale teoriei probabilităților, l-au avut A. de Moivre (1667-1754), P. S. Laplace (1749-1827), K. F. Gauss (1777-1855) și S. D. Poisson (1781-1840).

Definiția clasică a probabilității a fost introdusă de Laplace (1812) în tratatul *Teoria analitică a probabilităților* și a fost într-atât de reușită încât a fost luată ca bază de toți cercetătorii [1].

O altă etapă importantă în dezvoltarea teoriei probabilității au fost datorită metodelor efective de demonstrare a teoremelor limită pentru sumele de variabile aleatoare independente, create de P. L. Cebîșev (1821-1894) și elevii săi A. A. Markov (1856-1922) și A. M. Lyapunov (1857-1918).

O dată cu dezvoltarea teoriei probabilităților, au început să apară probleme pentru rezolvarea cărora noțiunea clasică a probabilității a devenit insuficientă. Însă, în anii treizeci ai secolului XX, A. N. Kolmogorov (1903-1987), în lucrarea *Bazele teoriei probabilităților*, a reușit să realizeze o axiomatizare a noțiunilor de bază, astfel că astăzi sunt acceptate unanim. Iar bazele statisticii matematice au fost puse de K. Pearson (1857-1936) și K. A. Fisher (1890-1962) [2].

Evident, teoria probabilităţilor şi statistica matematică se aplică în diverse domenii ale ştiinţei, începând cu ştiinţele exacte şi inginereşti şi finalizând cu ştiinţele socio-economice, în special în domeniile unde întâlnim condiţii de risc şi incertitudine şi unde este necesară luarea unor decizii riguros argumentate.

Fiind o ramură important a matematicii, a fost introdusă ca obiect de studiu începând cu clasele liceale, apoi continuând cu studiile superioare de specialitate.

Prezenta lucrare cuprinde introducere, patru capitole, bibliografie și anexe, în care sunt incluse tabelele de valori pentru diverse funcții.

În primul capitol sunt aduse datele teotetice și exemple pentru fiecare tip de probabilitate (clasică, geometrică, condiționată și totală), inclusiv formula lui Bayes și schemele clasice de probabilitate.

Capitolul doi este consacrat variabilelor aleatoare: operații cu variabile aleatoare discrete, funcția de repartiție, densitatea de probabilitate, valoarea medie, dispersia, momente de ordinul *k*, coeficientul de asimetrie și boltire și legile de repartiție a variabilelor aleatoare discrete și continue.

În capitolul trei sunt descrise elementele de statistică matematică: ce reprezintă o selecție, cum se înregistrează datele, se grupează, se reprezintă grafic, care sunt indicatorii seriei statistice și este expusă detaliat cercetarea statistică realizată prin sondaj.

În ultimul capitol sunt aduse câteva simulări, personale, de probleme din probabilitate, care au fost prezentate anterior la Conferința "Probleme ale ştiinţelor socioumanistice şi modernizării învăţământului", UPSC (2016, 2017 și 2022).

De asemenea, în majoritatea paragrafelor sunt incluse exemple rezolvate, iar la finele acestora sunt propuse exerciții și probleme pentru aprofundarea cunoștințelor.

CAPITOLUL I. TEORIA PROBABILITĂȚILOR

§1.1. Experiențe și evenimente. Operații cu evenimente

Teoria probabilităţilor operează cu o serie de noţiuni specifice care, în mod succint, pot fi prezentate astfel [1-17]:

Definiția 1. *Realizarea practică a unui ansamblu de condiţii și resurse bine precizate, conform unui anumit criteriu de cercetare, poartă numele de experienţă sau probă.*

Definiția 2. *Prin eveniment vom înţelege orice rezultat al unei experienţe despre care putem spune că s-a realizat sau că nu s-a realizat, după efectuarea experimentului considerat.*

Probabilitatea unui eveniment este măsura numerică a posibilităţii de realizare a evenimentului cercetat. Din punct de vedere al posibilităţilor de realizare, evenimentele se pot clasifica în:

- ✓ evenimente sigure;
- ✓ evenimente imposibile;
- ✓ evenimente aleatoare;
- ✓ evenimente contrarii.

Definiția 3. *Evenimentul sigur este evenimentul care, se produce în mod obligatoriu, la efectuarea unui experiment şi se notează cu* Ω.

De exemplu, la aruncarea unui zar să obținem fața cu un număr mai mic ca 6.

Definiția 4. *Evenimentul imposibil este evenimentul care în mod obligatoriu nu se produce în cadrul unui experiment și se notează cu Ø.*

De exemplu, la aruncarea unui zar să obținem fața cu un număr mai mare ca 6.

Definiția 5. *Evenimentul aleator (întâmplător) este evenimentul care poate, sau nu, să se realizeze într-un experiment, şi se notează în general prin literele mari ale alfabetului latin A, B, C, etc., sau prin literele mari urmate de indici* $A_1, A_2, A_3, \dots, A_n$.

Definiția 6. *Prin eveniment contrar al evenimentului A, vom înțelege evenimentul* $\bar{A}$ *care nu se produce dacă s-a produs evenimentul A și invers.*

Deci

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

Din definiția 6 rezultă

$$\bar{\Omega} = \emptyset \quad \text{și} \quad \bar{\emptyset} = \Omega.$$

În dependență de numărul de probe din experiment, avem două tipuri de evenimente:

Definiția 7. *Un eveniment se numeşte:*

1) elementar, dacă se realizează ca rezultatul unei singure probe și se notează cu ω;

2) compus, dacă apare ca rezultatul a două sau mai multe probe.

În sens tehnic, evenimentul legat de un element component simplu se poate numi eveniment simplu, iar evenimentele referitoare la sistemele tehnice sunt evenimente compuse.

O analogie între evenimente şi mulţimi, permite o scriere şi, în general, o exprimare mai comodă ale unor idei şi rezultate, legate de conceptul de eveniment. Astfel, vom înţelege evenimentul sigur ca mulţimea tuturor evenimentelor elementare, adică

$$\Omega = \{\omega_1, \omega_2, ..., \omega_n\}$$

şi orice eveniment compus ca o submulţime a lui Ω.

Definiţia 8. *Mulţimea tuturor evenimentelor elementare generate de un experiment aleator, se numeşte spaţiul evenimentelor elementare (spaţiul de selecţie) şi se notează cu Ω.*

Menționăm că spaţiul evenimentelor elementare poate fi finit sau infinit.

În cazul aruncării unei monede, o singură dată, spaţiul evenimentelor elementare este finit

$$\Omega = \{1,2\},$$

iar în cazul aruncării unei monede până la apariția stemei

$$\Omega = \left\{s, bs, bbs, bbbs, \dots, \underbrace{bb \dots b}_{n\ ori}\, s, \dots\right\}$$

este o mulțime infinită și nu putem preciza un număr exact de aruncări în cadrul cărora să apară cu siguranță stema.

Exemplul 1. Se aruncă două zaruri, să se scrie mulţimea evenimentelor elementare.

Rzolvare. Fiecare zar are 6 feţe, iar pe fiecare faţă sunt indicate punctele de la 1 la 6. În total mulţimea evenimentelor elementare este formată din $6 \cdot 6 = 36$ elemente, deci

$$\begin{gathered}\{(1;1),(1;2),(1;3),(1;4),(1;5),(1;6),\\ (2;1),(2;2),(2;3),(2;4),(2;5),(2;6),\\ (3;1),(3;2),(3;3),(3;4),(3;5),(3;6),\\ (4;1),(41;2),(4;3),(4;4),(4;5),(4;6),\\ (5;1),(5;2),(5;3),(5;4),(5;5),(5;6),\\ (6;1),(6;2),(6;3),(6;4),(6;5),(6;6)\}.\end{gathered}$$

În cazul unui experiment se pot produce simultan mai multe evenimente aleatoare diferite.

Definiţia 9. *Prin reunirea evenimentelor A şi B, vom înţelege evenimentul notat* $A \cup B$*, care constă în realizarea a cel puţin unuia dintre evenimentele A şi B.*

Deoarece evenimentele A şi B sunt submulţimi formate din evenimentele elementare ale spaţiului Ω, rezultă că reunirea evenimentelor este

$$A \cup B = \{\omega \in \Omega | \omega \in \mathrm{A} \text{ sau } \omega \in \mathrm{B}\}.$$

Se cunosc următoarele proprietăți ale reuniunii evenimentelor [2]:

1. $A \cup B = B \cup A$ (legea comutativă);
2. $(A \cup B) \cup C = A \cup (B \cup C)$ (legea asociativă);
3. Dacă $A \subset B \Longrightarrow A \cup B = B$;
4. $A \cup A = A$; $A \cup \Omega = \Omega$, $A \cup \emptyset = \mathrm{A}$, $\emptyset \cup \Omega = \Omega$ şi $A \cup \overline{\mathrm{A}} = \Omega$.

Definiţia 10. *Evenimentul care prezintă realizarea simultană a evenimentelor A și B, se numește intersecția evenimentelor A și B, notat*

$$A \cap B = \{\omega: \omega \in A \text{ și } \omega \in B\}.$$

În cazul intersecţiei, au loc următoarele proprietăți:

1. $A \cap B = B \cap A$ (legea comutativă);
2. $(A \cap B) \cap C = A \cap (B \cap C)$ (asociativă);
3. Dacă $A \subset B \Longrightarrow A \cap B = A$;
4. $A \cap A = A$; $A \cap \Omega = \mathrm{A}$, $A \cap \emptyset = \emptyset$, $\emptyset \cap \Omega = \emptyset$ şi $A \cap \mathrm{A} = \mathrm{A}$);
5. Pentru orice eveniment A, există $\overline{\mathrm{A}}$ astfel încât $A \cap \overline{\mathrm{A}} = \emptyset$.

Definiţia 11. *Evenimentul care se realizează dacă și numai dacă se realizează evenimentul A, dar nu se realizează evenimentul B, se numește diferența evenimentelor A și B, notat*

$$A - B = \{\omega: \omega \in A \text{ și } \omega \notin B\}.$$

Definiţia 12. *Evenimentele A și B, se numesc compatibile, dacă ele se pot produce simultan în cadrul unui experiment, adică* $A \cap B \neq \emptyset$.

De exemplu, la un echipament electric pus în funcţiune, după un anumit timp pot să apară concomitent diverse defecte (evenimente), astfel de evenimente sunt compatibile.

Definiţia 13. *Evenimentele A și B se numesc incompatibile, dacă ele nu se produc simultan în cadrul unui experiment, adică* $A \cap B = \emptyset$.

Fie evenimentul A care constă din funcţionarea unui calculator, iar nefuncţionarea lui este evenimentul B, deoarece calculatorul în același timp nu poate să funcționeze și să nu funcționeze, rezută că evenimentele A și B sunt incompatibile.

Evenimentele contrarii (A și $\overline{\mathrm{A}}$) sunt incompatibile, în timp ce evenimentele incompatibile nu sunt în mod obligatoriu contrarii.

Definiţia 14. *Evenimentul A implică evenimentul B şi vom scrie* $A \subset B$*, dacă realizarea evenimentul A atrage după sine şi realizarea evenimentului B.*

Dacă $A \subset B$ şi $B \subset C$, atunci $A \subset C$ reprezintă proprietatea de tranzitivitate a relaţiei de implicare.

Definiţia 15. *Evenimentele A şi B sunt echivalente (egale), dacă avem simultan $A \subset B$ şi $B \subset A$ și vom scrie $A = B$.*

Operațiile cu evenimente aleatoare și relațiile dintre ele pot fi ilustrate cu ajutorul diagramelor Venn (fig. 1).

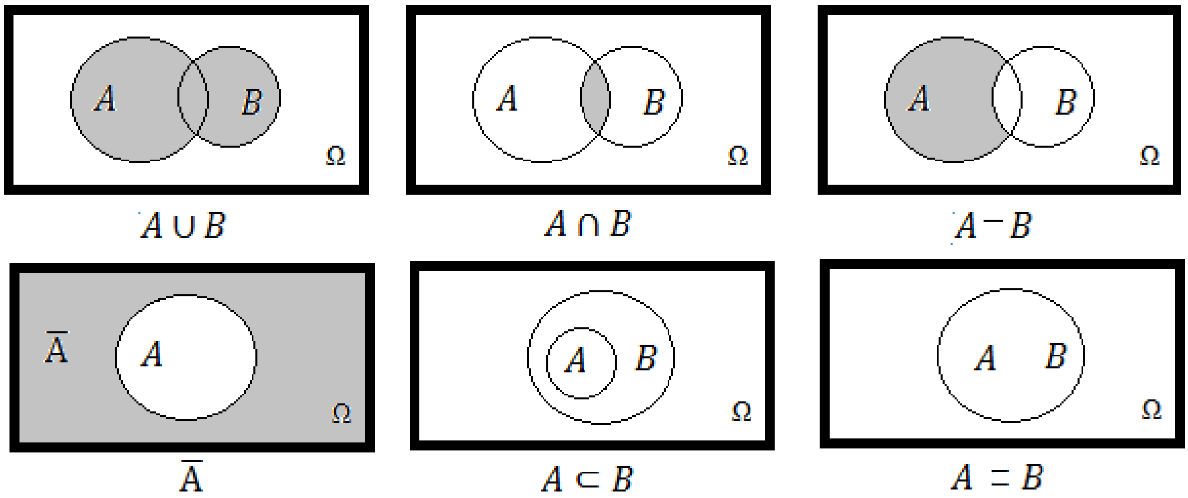


Fig. 1. Ilustrarea geometrică a operațiilor cu evenimente aleatoare și relațiile dintre ele

Exerciții și probleme propuse

1. *Fie evenimentul A - cel puțin o piesă este rebut din trei piese verificate, B - toate piesele sunt bune. Să se determine evenimentele $A \cup B$, $A \cap B$, $\bar{A}$, $\bar{B}$, $A \cup \bar{B}$.*
2. *Să se demonstreze echivalența evenimentelor $\overline{\overline{\bar{A} \cap B} \cup A}$ și $B - A$.*
3. *Să se demonstreze compatibilitate/incompatibilitatea evenimentelor A și $\overline{A \cup B}$.*

§1.2. Câmp de evenimente. Câmp de probabilitate. Frecvenţa relativă

O mulţime K de evenimente formează un câmp dacă au loc următoarele axiomele:

1) $\forall A \in K \implies \exists \bar{A} \in K$;

2) $\forall A, B \in K \implies A \cup B \in K$ *și* $A \cap B \in K$.

Definiția 1. *Cuplul (Ω, K) se numeşte spaţiu măsurabil sau câmp finit de evenimente, dacă K este un corp.*

Observaţia 1. *I. Într-un câmp finit de evenimente (Ω, K), sunt adevărate afirmaţiile*

a) $\forall A, B \in K \implies A - B \in K$;

b) *evident* $\emptyset \in K$ *şi* $\Omega \in K$;

c) *dacă* $A, B \in K$, atunci $A \cap B \in K$.

II. *Dacă mulţimea evenimentelor elementare este numărabilă, o mulţime* $K \subseteq P(\Omega)$ *se numeşte corp borelian pe* Ω, *în condiţiile*

- ✓ $\forall A \in K \Rightarrow \bar{A} \in K$;
- ✓ *dacă* $I \subseteq N, I \neq \emptyset$ *şi* $A_i \in K, \forall i \in I$, *atunci*

$$\bigcup_{i \in I} A_i \in K\,;$$

- ✓ $\Omega \in K$.

Definiția 2. *Perechea* (Ω, K) *în care* K *este un corp borelian se numeşte câmp borelian (câmp infinit) de evenimente.*

Definiţia 3. *Într-un câmp finit de evenimente* (Ω, K), *evenimentele* $A_i \in K$, $i = \overline{1, n}$, *formează un sistem complet de evenimente (sau o partiţie a câmpului), dacă*

- ➢ $\bigcup_{i \in I}^{n} A_i \in \Omega\,;$
- ➢ $A_i \cap A_j = \emptyset, \quad \forall i \neq j, \quad i, j = \overline{1, n}.$

Definiţia 4. *Fie câmpul de evenimente* (Ω, K), *evenimentele* $A_1, A_2, \dots, A_n$ *formează un sistem complet de evenimente, dacă*

$$\bigcup_{i \in 1}^{n} A_i = \Omega \ \text{ și } A_i \cap A_j = \emptyset;\ \forall i \neq j;\ i, j = \overline{1, n}.$$

Propoziţia 1. *Dacă mulţimea* Ω *este format din evenimentele*

$$\Omega = \{\omega_1, \omega_2, \dots, \omega_n\},$$

atunci câmpul (Ω, K) *conţine* 2^n *evenimente.*

Demonstraţie. Pentru un experiment din *n* rezultate elementare şi prin urmare, pentru un eveniment sigur compus din *n* evenimente elementare, vom avea diverse evenimente compuse din acestea, după cum urmează:

- evenimente compuse din câte zero evenimente elementare sunt C_n^0;
- evenimente compuse din câte un eveniment elementar sunt C_n^1;
- evenimente compuse din câte două evenimente elementare sunt C_n^2;
- -
- evenimente compuse din câte *n* evenimente elementare sunt C_n^n.

Prin urmare, numărul total de evenimente ale câmpului *K* este egal cu

$$C_n^0 + C_n^1 + C_n^2 + \cdots + C_n^n = 2^n.$$

Definiţia 5. *Fie un experiment ce se repetă de* n *ori, considerăm un eveniment A ca un rezultat al acestor experimente şi fixăm de câte ori s-a repetat acest eveniment, atunci prin frecvenţă relativă vom înţelege raportul* $\frac{k}{n}$ *care se notează*

$$f_k(A) = \frac{k}{n},$$

unde k este numărul de evenimente care se produc, iar n *este numărul total de evenimente.*

Proprietăţile frecvenţei relative

1) $0 \leq f_k(A) \leq 1$ pentru orice eveniment A;
2) $f_k(\Omega) = 1, \quad f_k(\emptyset) = 0$;
3) Dacă A şi B sunt incompatibile, atunci

$$f(A \cup B) = f(A) + f(B), \qquad B \subset A,$$

iar în caz general

$$f(A \cup B) = f(A) + f(B) - f(A \cap B);$$

4) În cazul evenimentelor arbitrare

$$f(A \backslash B) = f(A \cap \bar{B}) = f(A) - f(A \cap B),$$

iar când $B \subset A$

$$f(A \backslash B) = f(A \cap \bar{B}) = f(A) - f(B);$$

5) $f(\bar{A}) = 1 - f(A)$.

În secolul al XVIII-lea, naturalistul Buffon și, la începutul secolului XX, statisticianul Pearson, au trecut rezultatele unui experiment ce se repetă în aceleași condiții practic identice (aruncarea unei monede simetrice și omogene), în următorul tabel [1]

Tabelul1. Rezultatele experimentelor naturalistului Buffon și statisticianului Pearson

Experimentatorul	**Numărul aruncărilor**	**Frecvența absolută a stemei**	**Frecvența relativă**
Buffon	4040	2048	0,5069
Pearson	12000	6019	0,5016
Pearson	24000	12012	0,5005

Comte de Buffon (1707-1788)

Karl Pearson (1857-1936)

S-a observat, conform datelor din tabel, că frecvența relativă capătă o anumită stabilitate, oscilând în jurul unui număr constant p. Asemenea proprietăți pot fi observate și în alte experimente repetate de nu număr suficient de mare și anume aceste proprietăți ne sugerează următoarea definiție:

Definiția empirică sau statistică a probabilității [1]. *Numărul în jurul căruia se concentrează (oscilează) frecvențele relative ale unui eveniment A, în serii mari de efectuări ale experimentului, îl vom numi probabilitatea evenimentului A care se notează cu $P(A)$.*

§1.3. Definiția axiomatică și definiția clasică a probabilității

A defini o probabilitate în raport cu o experienţă, înseamnă a asocia fiecărui eveniment A un număr $P(A)$.

Definiţia 1 (definiția axiomatică a probabilității) [2]. *Fie (Ω, K) un câmp finit de evenimente, se numeşte probabilitate pe câmpul considerat, o funcţie*

$$P: K \longrightarrow R$$

care satisface axiomele

1) $P(A) \geq 0, \forall A \in K,$
2) $P(\Omega) = 1,$
3) $P(A \cup B) = P(A) + P(B),\ \forall A, B \in K$ şi $A \cap B = \emptyset.$

Tripletul $\{\Omega, K, P\}$ se numeşte *câmp finit de probabilitate*.

În cazul în care câmpul de evenimente (Ω, K) este infinit, atunci probabilitatea trebuie să satisfacă axiomele

1) $P(A) \geq 0, \forall A \in K;$
2) $P(\Omega) = 1;$
3) $P(\bigcup_i A_i) = \sum_i P(A_i),\quad A_i \cap A_j = \emptyset,\ i \neq j.$

În conformitate cu [2], avem următoarea regulă de adunare a probabilităților: dacă evenimentele $A_i \in K$ $(i = 1,2, \dots, n)$, sunt evenimente incompatibile două câte două, adică

$$A_i \cap A_j = \emptyset\ (i \neq j),$$

atunci

$$P(A_1 \cup A_2 \cup \dots \cup A_n) = P(A_1) + P(A_2) + \dots + P(A_n), \tag{1}$$

iar dacă evenimentele $A_i \in K$ $(i = 1,2, \dots, n)$ sunt compatibile, atunci

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum\nolimits_{i=1}^{n} P(A_i) - \sum\nolimits_{i,j=1}^{n} P(A_i \cap A_j) +$$

$$+ \sum\nolimits_{i,j,k=1}^{n} P(A_i \cap A_j \cap A_k) + \dots + (-1)^{n-1} P\left(\bigcap_{i=1}^{n} A_i\right)$$

$$(i \neq j \neq k).$$

Fie mulțimea finită de evenimente elementare $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$, atunci, conform regulii de adunarea a probabilităților (1), avem

$$P(\Omega) = P(\omega_1) + P(\omega_2) + \dots + P(\omega_n) = 1. \qquad (2)$$

Definiție 2. *Evenimentele se numesc echiprobabile, dacă au aceeași probabilitate, adică*

$$P(\omega_1) = P(\omega_2) = \dots = P(\omega_n).$$

Din condițiile definiției de mai sus se deduce probabilitatea pentru oricare eveniment elementar ω_i

$$P(\omega_i) = \frac{1}{n} \ (i = 1..n).$$

Conform relației (2) avem

$$P(A) = \underbrace{\frac{1}{n} + \frac{1}{n} + \dots + \frac{1}{n}}_{m\ ori} = \frac{m}{n},$$

unde m este numărul de evenimente elementare favorabile evenimentului A, deci $m = card\ A$, iar n este numărul total de evenimente elementare echiprobabile ($n = card\ \Omega$).

Definiția 3 (definiția clasică a probabilității) [3]. *Probabilitatea unui eveniment A este egală cu raportul dintre numărul de cazuri favorabile la numărul total de cazuri posibile*

$$P(A) = \frac{card\ A}{card\ \Omega}. \qquad (3)$$

Observația 1. *Conform definiției de mai sus, nu putem stabili probabilitatea unui eveniment dintr-un câmp infinit, deoarece ne întâlnim cu imposibilitatea stabilirii numărului de cazuri favorabile și a numărului total de cazuri.*

Examinăm aruncarea unei monede, aceasta poate înregistra 2 evenimente: căderea banului (*b*) sau a stemei (*s*), deci mulțimea de evenimente elementare este

$$\Omega = \{s, b\},$$

iar probabilitatea căderea banului (*b*), sau a stemei (s) este

$$P(s) = P(b) = \frac{1}{2}.$$

Din definiția probabilității, operațiile dintre evenimente și relațiile dintre acestea, rezultă următoarele proprietăți ale probabilității:

a) $P(\emptyset) = 0, P(\Omega) = 1;$

b) $P(\bar{A}) = 1 - P(A);$

c) $P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i),\ A_i \cap A_j = \emptyset,\ \ i \neq j,\ \ i,j = \overline{1,n};$

d) $P(A \backslash B) = P(A \cap \bar{B}) = P(A) - P(B),\ \ dacă\ B \subset A;$

$P(A \backslash B) = P(A \cap \bar{B}) = P(A) - P(A \cap B);$

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

f) $P(\emptyset) = 1 - P(\Omega)$, $P(\Omega) = 1 - P(\emptyset)$.

Exemplul 1. Considerăm aruncarea unui zar. Să se determine probabilitatea căderii unei feţe cu un număr impar de puncte.

Rezolvare. Mulţimea evenimentelor elementare este

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

Notăm prin A – evenimentul căderii feţei cu un număr impar de puncte, deci

$$A = \{1; 3; 5\}.$$

În conformitate cu definiția clasică a probabilității (definiția 3), formula (3), obținem

$$P(A) = \frac{3}{6} = \frac{1}{2} = 0{,}5.$$

Răspuns: $P(A) = 0{,}5$

Exemplul 2. Se aruncă o monedă de 3 ori, să se determine mulţimea de evenimente elementare și probabilitatea căderii a 2 steme.

Rezolvare. Notăm:

evenimentul A – căderea a 2 steme din 3 aruncări a monedei;

evenimentul b – căderea banului;

evenimentul s – căderea stemei.

În baza proprietăţii de mai sus, rezultă că mulţimea de evenimente elementare este alcătuită din 2^3=8 elemente, deci

$$\Omega = \{(bbb), (bbs), (bsb), (bss), (sbb), (sbs), (ssb), (sss)\}.$$

Din mulţimea de evenimente elementare determinăm cazurile favorabile

$$A = \{(bss), (sbs), (ssb)\},$$

deci avem 3 cazuri favorabile.

Probabilitatea se calculează în baza formulei (3)

$$P(A) = \frac{3}{8} = 0{,}375.$$

Răspuns: $\Omega = \{(bbb), (bbs), (bsb), (bss), (sbb), (sbs), (ssb), (sss)\}$,
$P(A) = 0{,}375$

Exemplul 3. Fie date două urne cu bile albe şi negre. Prima urnă conţine 4 bile albe şi 6 bile negre, iar a doua 3 bile albe şi 5 bile negre. Se extrag la întâmplare câte o bilă din fiecare urnă. Să se determine probabilitatea că:

1) ambele bile sunt albe (evenimentul A_1);

2) cel puţin o bilă este albă (evenimentul A_2);

3) din prima – albă, iar din a doua – neagră (evenimentul A_3).

Rezolvare. Notăm: evenimentul A – bila extrasă din prima urnă este albă și evenimentul B – bila extrasă din a doua urnă este albă.

Probabilitatea evenimetelor A și B se calculează în baza formulei (3)

$$P(A) = \frac{4}{10} = \frac{2}{5}$$

și

$$P(B) = \frac{3}{8}.$$

1) Probabilitatea că ambele bile să fie albe reprezintă probabilitatea intersecției evenimetelor A și B

$$P(A_1) = P(A \cap B) = P(A) \cdot P(B) = \frac{2}{5} \cdot \frac{3}{8} = \frac{3}{20}.$$

2) Probabilitatea că cel puțin o bilă să fie albă reprezintă probabilitatea reuniunii evenimetelor A și B

$$P(A_2) = P(A \cup B) = P(A) + P(B) - P(A \cap B) =$$
$$= \frac{2}{5} + \frac{3}{8} - \frac{3}{20} = \frac{16 + 15 - 6}{40} = \frac{25}{40} = \frac{5}{8}.$$

3) Probabilitatea că din prima urnă să extragem o bilă albă, iar din a doua o bilă neagră, reprezintă probabilitatea intersecției evenimetelor A și $\bar{B}$

$$P(A_3) = P(A \cap \bar{B}) = P(A) - P(A \cap B) = \frac{2}{5} - \frac{3}{20} = \frac{8 - 3}{20} = \frac{1}{4}.$$

Răspuns: $P(A_1) = \frac{3}{20}; P(A_2) = \frac{5}{8}; P(A_3) = \frac{1}{4}$

Exemplul 4. Un aparat este format din 3 piese. Aparatul nu funcționează dacă 2 din piesele sale sunt nestandarte. Să se determine probabilitatea că aparatul va funcționa dacă muncitorul are la dispoziție 10 piese dintre care 3 sunt nestandarte.

Rezolvare. Notăm evenimentul A – aparatul funcţionează.

Numărul total de posibilităţi de selectare a pieselor se calculează cu ajutorul combinărilor

$$C_{10}^3 = \frac{10!}{3! \cdot 7!} = 120.$$

Conform condiţiilor problemei, aparatul funcţionează dacă cel puţin 2 piese sunt bune, deci cazurile favorabile sunt:

$$N = C_7^3 + C_7^2 \cdot C_3^1;$$

- 3 piese bune

$$C_7^3 = \frac{7!}{3! \cdot 4!} = 35;$$

- 2 piese bune și una nestandartă

$$C_7^2 \cdot C_3^1 = \frac{7!}{2! \cdot 5!} \cdot 3 = 63.$$

Obținem

$$N = 35 + 63 = 98.$$

În baza formulei (3), avem

$$P(A) = \frac{98}{120} = \frac{49}{60}.$$

Răspuns: $P(A) = \frac{49}{60}$

Exerciții și probleme propuse

1. *Se aruncă o monedă de 5 ori. Să se determine probabilitatea că*
   a) *stema cade de 2 ori;*
   b) *stema cade cel puțin o dată.*
2. *Se aruncă un zar simetric. Să se determine probabilitatea căderii*
   a) *unui număr par de puncte;*
   b) *unui număr divizibil cu 3;*
   c) *unui număr nu mai mic ca 5.*
3. *Să se determine probabilitatea ca cifrele, dintr-un număr de trei cifre alese la întâmplare, să fie consecitive?*
4. *Într-o urnă sunt cartonașe numerotate de la 1 la 10000, fără repetare. Să se determine probabilitatea că un cartonaș extras la întâmplare să conțină cifra 7.*
5. *Într-o urnă sunt 6 bile albe, 2 negre și 8 roșii. Se extrag la întâmplare 2 bile. Să se determine probabilitatea că ambele bile sunt de aceeași culoare (evenimentul* A*) și probabilitatea că ambele bile sunt de culoare roșie (evenimentul* B*).*
6. *Fie o clădire cu 10 etaje. În accensorul acestei clădiri urcă 4 persoane. Să se determine probabilitatea că persoanele au coborât la etaje diferite*
7. *La o fabrică se produc piese pentru aparate foto, se știe că 2% din ele nu corespund standartului din cauza componentelor, iar 4% din cauza montării. Să se determine probabilitatea că o piesă extrasă la întâmplare să fie nestandartă.*
8. *Dintr-o urnă cu 100 bile numerotate de la 1 la 100, fără repetare, se extrage o bilă la întâmplare. Să se determine probabilitatea*
   a) *obţinerii unei bile cu un număr prim;*
   b) *obţinerii unei bile cu un număr par;*
   c) *obţinerii unei bile cu un număr divizibil cu 3.*

§1.4. Definiția geometrică a probabilității

Considerăm în $\mathbb{R}^1$, un domeniu Ω, reprezentat printr-un segment $[PQ]$ de lungimea L, iar A $(A \subset \Omega)$ un segment $[RS] \in [PQ]$, de lungimea $l < L$ (fig. 2).

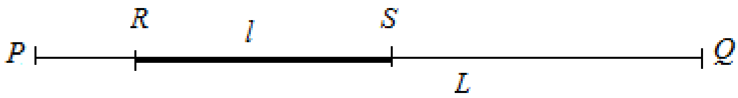


Fig. 2. Reprezentarea în $\mathbb{R}^1$ *a unui domeniu* $([PQ])$ *și a unei părți din acest domeniu* $([RS])$

Pe segmentul L se ia la întâmplare un punct, să se determine probabilitatea că punctul aparţine și segmentului l.

Dacă considerăm că probabilitatea de a nimeri în segmental l este direct proporțională cu lungimea lui, atunci probabilitatea că punctul dat nimerește pe segmentul l este raportul

$$P(A) = \frac{lung\ l}{lung\ L}. \tag{4}$$

Exemplul 1. Se dă un segment $[AB]$ *de lungimea m, se ia la întâmplare două puncte din acest segment C și D. Să se determine probabilitatea că segmentul* $[CD]$ *are lungimea* $\frac{1}{3}m$ *(fig. 3).*

Fig. 3. Reprezentarea segmentelor $[AB]$ și $[CD]$

Rezolvare. Aplicăm probabilitatea geometrică în $\mathbb{R}^1$, formula (4), obținem

$$P(A) = \frac{[CD]}{[AB]} = \frac{\frac{1}{3}m}{m} = \frac{1}{3}.$$

În $\mathbb{R}^2$, domeniu Ω este reprezentat printr-o figură plană G, iar A $(A \subset \Omega)$ o altă figură plană g, încât $g \in G$ (fig. 4).

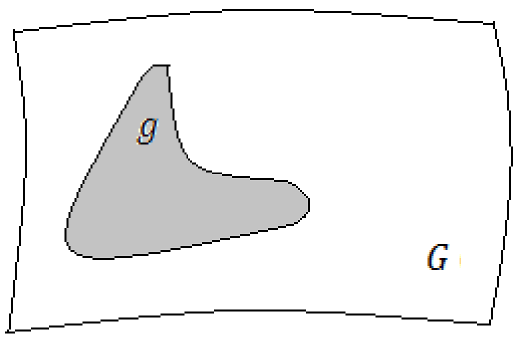


Fig. 4. Reprezentarea în $\mathbb{R}^2$ *a unui plan G și a unei părți din acest plan g*

Din figura G se ia la întâmplare un punct, să se determine probabilitatea că punctul aparţine și figurii g.

În conformitate cu literatura de specialitate [1-17], probabilitatea se exprimă ca raportul dintre ariile figurilor respective

$$P(A) = \frac{aria\ g}{aria\ G}. \quad (5)$$

Exemplul 2. Un ceas se opreşte la întâmplare. Să se determine probabilitatea că acul minutarul se va opri între orele 5 şi 8.

Rezolvare. Fie evenimentul A – acul minutarului se oprește între 5 și 8

Geometric problema dată poate fi ilustrată în 2 moduri:

a) Ca lungime (fig. 5)

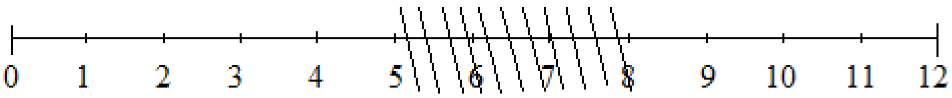


Fig. 5. Reprezentarea condițiilor exemplului 2 ca lungime

În acest caz, probabilitatea se calculează ca raportul dintre lungimile segmentelor respective (formula (4))

$$P(A) = \frac{lung\ l}{lung\ L} = \frac{3}{12} = \frac{1}{4} = 0{,}25.$$

b) Ca arie (fig. 6)

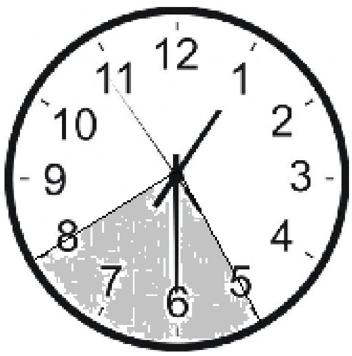


Fig. 6. Reprezentarea condițiilor exemplului 2 ca arie

Pentru simplitate vom împarte suprafața ceasului în 12 părți egale (fig. 7).

Fig. 7. O parte din cele 12 părți egale, obținute la divizarea ariei ceasului

Aria ceasului *G* este egală cu *12 g* (12 părți egale), iar de la 5 la 8 sunt 3 părți egale, adică *3g*. Deci, probabilitatea se va calcula ca raportul dintre ariile suprafețelor respective (formula (5))

$$P(A) = \frac{3g}{12g} = 0{,}25.$$

Răspuns: $P(A) = 0{,}25$

Exemplul 3. Fie două ceruri C_1 și C_2 de raze $r_1 = 200$ și $r_2 = 100,$ cu centrele cercurilor în acelaşi punct $O(x, y)$. Se ia la întâmplare un punct din cercul C_1. Să se determine probabilitatea că acest punct nu aparține cercului C_2.

Rezolvare. Fie evenimentul A - un punct luat la întâmplare din cercul C_1 să nu aparțină cercului C_2.

Reprezentăm geometric cercurile C_1 și C_2 cu centrele în același punct $O(x, y)$ (fig. 8).

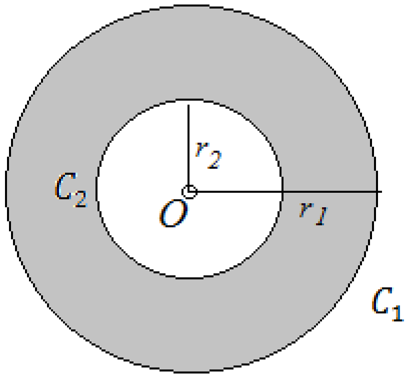


Fig. 8. Cercurile C_1 și C_2 cu centrele în acelaşi punct $O(x, y)$

Discul format de cercurile C_1 și C_2, reprezintă suprafața cazurilor favorabile, aria căruia se calculează ca diferenţa dintre aria cercului mare şi aria cercului mic. Determinăm ariile cercurilor C_1 și C_2

$$A_{c1} = \pi r_1^2 = 40000\,\pi\ (u.p)$$

și

$$A_{c2} = \pi r_2^2 = 10000\,\pi\ (u.p),$$

deci aria discului A_h este

$$A_h = \pi r_1^2 - \pi r_2^2 = (r_1^2 - r_2^2)\pi = (40000 - 10000)\pi = 30000\,\pi\ (u.p).$$

În baza formulei (5), calculăm probabilitatea

$$P(A) = \frac{A_h}{A_{c1}} = \frac{30000\,\pi}{40000\,\pi} = 0{,}75.$$

Răspuns: $P(A) = 0{,}75$

Exemplul 4. Doi studenți au hotărât să se întâlnească între orele 18^{00} și 19^{00}, au decis să se aștepte 15 min. Să se determine probabilitatea că studenții se vor întâlni [1].

Rezolvare. Notăm evenimentul *x* ora când vine studentul I şi *y* ora când vine studentul II.

Această problem poate fi ilustrată geometric în felul următor (fig. 9):

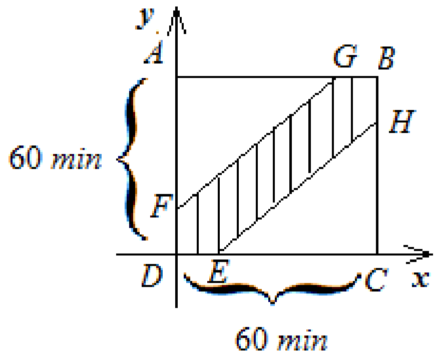


Fig. 9. Reprezentarea geometrică a întâlnirii studenţilor

unde *ABCD* reprezintă un pătrat cu latura egală cu 60 (între 18^{00} *și* 19^{00} sunt 60 de minute), iar *DE=DF=BH=GB=15* (timpul așteptării). Cazurile favorabile se determină ca aria figurii haşurate (*EDFGBH*), iar probabilitatea se determină ca raportul dintre arii (formula (5))

$$P(A) = \frac{A_{EDFGBH}}{A_{ABCD}}.$$

Observăm, mai simplu este să calculăm ariile triunghiurilor ΔAGF și ΔHCE, apoi din aria pătratului să scădem ariile acestor triunghiuri. Mai mult ca atât, aceste triunghiuri sunt asemenea și coincid

ΔAGF este triunghi dreptunghic și ΔHCE este triunghi dreptunghic

și

$$AF = AG = CE = CH = 45.$$

De aici rezultă că suma ariilor acestor triunghiuri reprezintă aria unui pătrat cu latura $AF = 45$. Deci

$$A_{EDFGBH} = A_{ABCD} - (A_{AGF} + A_{HCE}) = 3600 - 2025 = 1575\ (u.p)$$

În baza formulei (5), obținem

$$P(A) = \frac{1575}{3600} = 0{,}4375.$$

Răspuns: $P(A) = 0{,}4375$

În $\mathbb{R}^3$, domeniu Ω, este reprezentat printr-un corp V și fie v un alt corp, încât $v \in V$ (fig. 10).

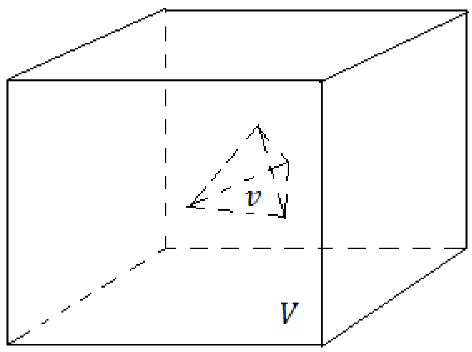


Fig. 10. Reprezentarea în $\mathbb{R}^3$ *a corpurilor* v *și* V $(v \subset V)$

Din corpul V se ia la întâmplare un punct, să se determine probabilitatea că punctul aparţine și corpului v.

În conformitate cu literatura de specialitate [3], probabilitatea se exprimă ca raportul dintre volumele corpurilor respective

$$P(A) = \frac{vol\ v}{vol\ V}. \tag{6}$$

Exerciții și probleme propuse

1. Să se determine probabilitatea că un punct luat la întâmplare din segmentul $[0,10]$, *să aparțină segmentului* $[0,1]$.

2. Fie un segment $[AB]$, *pe acest segment se i-au la întâmplare două puncte, formând 3 segmente. Să se determine probabilitatea că segmentele obținute satisfac proprietății triunghiului.*

3. Fie dat un segment $[AB]$ *de lungimea* t, *se iau la întâmplare două puncte* $L \in [AB]$ *și* $M \in [AB]$. *Să se determine probabilitatea că punctul* L *să fie mai aproape de* M *decât de* A.

4. Să se determine probabilitatea că un punct luat la întâmplare, din sistemul rectangular cartezian de coordonate, să aparțină cadranului IV.

5. O barcă trece un râu, dintr-un mal până la celălalt, timp de o oră. Să se determine probabilitatea că un vapor va fi observat, dacă se știe că barca vede vaporul cu 20 de minute înaite de ciocnire, după sau contra cursului vaporului.

§1.5. Probabilitatea condiționată

Fie (X, Ω, P) un câmp de probabilitate, considerăm evenimentele A și B, încât $P(B) > 0$.

Definiţia 1. *Definim* $P_B: \Omega \to R$ *(sau* $P(\cdot\,|B)$*) prin*

$$P(A/B) = P_B(A) = \frac{P(A \cap B)}{P(B)} \tag{7}$$

şi o numim probabilitatea lui A *condiţionată de* B.

În mod analog

$$P(B/A) = P_A(B) = \frac{P(A \cap B)}{P(A)}. \tag{8}$$

Teorema 1 [3]. *În condiţiile ultimii egalităţi,* P_B *este o probabilitate şi în plus*

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

Demonstraţie. Deoarece

$$\emptyset \subset A \cap B \subset B,$$

atunci

$$0 \leq P_B(A) \leq 1.$$

De asemenea

$$P_B(\emptyset) = 0 \quad \text{şi} \quad P_B(X) = 1$$

sunt evidente.

Dacă $A_1 \cap A_2 = \emptyset$, atunci şi

$$(A_1 \cap B) \cap (A_2 \cap B) = \emptyset,$$

deci

$$P_B(A_1 \cup A_2) = \frac{P\big((A_1 \cup A_2) \cap B\big)}{P(B)} = \frac{P((A_1 \cap B) \cup (A_2 \cap B))}{P(B)} =$$

$$= \frac{P(A_1 \cap B) + P(A_2 \cap B)}{P(B)} = P_B(A_1) + P_B(A_2).$$

Definiţia 2. *Două evenimente A şi B, se numesc independente, dacă probabilitatea efectuării unui din ele nu este influențată de efectuarea sau nu a celuilalt*

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

altfel spus

$$P(A) = P_B(A).$$

Propoziţia 1. *Dacă evenimentele A şi B sunt independente, atunci şi perechile de evenimente* $(A, \bar{B})$, $(\bar{A}, B)$, $(\bar{A}, \bar{B})$ *sunt independente.*

Demonstraţie. Prin definiţie avem

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

Deoarece

$$A = (A \cap B) \cup (A \cap \bar{B}) \text{ (reuniune disjunctă),}$$

avem

$$P(A) = P(A \cap B) + P(A \cap \bar{B}) = P(A)P(B) + P(A \cap \bar{B}),$$

de unde rezultă

$$P(A \cap \bar{B}) = P(A) - P(A) \cdot P(B) = P(A) \cdot \big(1 - P(B)\big) = P(A) \cdot P(\bar{B}).$$

Analog se demonstrează independenţa şi în celelalte cazuri.

Fie dată o urnă cu bile albe și negre. Notăm evenimentul A – extragerea unei bile albe, și evenimentul B – extragerea unei bile negre după ce a fost extrasă o bilă care nu s-a întors înapoi până nu s-a extras a doua.

Notăm bilele albe prin *a*, iar cele negre prin *b*, atunci

$$P(A) = \frac{a}{a+b},$$

dacă se extrage bila albă din prima oară, atunci

$$P(B) = \frac{b}{a+b-1},$$

însă, dacă se extrage din prima oară o bilă neagră, atunci avem o altă probabilitate a evenimentului *B* și anume

$$P(B) = \frac{b-1}{a+b-1}.$$

În concluzie putem spune că evenimentul B depinde de evenimentul A.

Fie două evenimente dependente A și B, vom determina care este probabilitatea producerii simultane a acestor două evenimente (probabilitatea intersecției).

Într-o operație în masă avem următoarele cazuri posibile:

- se produc evenimentele A și B, $A \cap B$, în m_1 cazuri favorabile;
- se produce primul eveniment, dar nu se produce al doilea, $A \cap \bar{B}$, în m_2 cazuri;
- nu se produce primul eveniment, dar se produce al doilea, $\bar{A} \cap B$, în m_3 cazuri;
- nu se produce nici un eveniment, $\bar{A} \cap \bar{B}$, în m_4 cazuri.

În total avem

$$m_1 + m_2 + m_3 + m_4 = n$$

cazuri posibile, deci

$$P(A \cap B) = \frac{m_1}{n},$$

$$P(A) = \frac{m_1 + m_2}{2}$$

și

$$P(B) = \frac{m_1 + m_3}{2}.$$

Evenimentele A și B sunt dependente, adică probabilitatea evenimentului B este influiențată de evenimentul A și se notează $P_A(B)$ care, din condițiile exemplului este

$$P_A(B) = \frac{m_1}{m_1 + m_3}.$$

În caz general, când evenimentele sunt dependente, obținem egalitatea

$$P_B(A) \cdot P(B) = P(A) \cdot P_A(B).$$

Exemplul 1. Într-o ladă sunt 40 de piese, dintre care 37 sunt bune, iar 3 defecte. Vasile a extras două piese. Să se determine probabilitatea că ambele sunt bune.

Rezolvare. Notăm evenimentul A_1 prima piesă extrasă este bună, evenimentul A_2 a doua piesă extrasă este bună și evenimentul A ambele piese extrase sunt bune.

Probabilitatea evenimentului A_1 este

$$P(A_1) = \frac{37}{40},$$

însă probabilitatea evenimentului A_2 este dependentă de calitatea primei piese extrase.

Dacă s-a extras o piesă bună, atunci

$$P(A_2) = \frac{36}{39},$$

iar dacă s-a extras o piesă cu defect, atunci

$$P(A_2) = \frac{37}{39}.$$

Deci probabilitatea că ambele piese să fie bune, este

$$P(A) = P(A_1 \cap A_2) = P(A_1) \cdot P_{A_1}(A_2) = \frac{37}{40} \cdot \frac{36}{39} = 0{,}85.$$

Răspuns: $P(A) = 0{,}85$

Dacă evenimentele sunt independente, probabilitatea se calculează în baza formulei

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

În caz general, probabilitatea mai multor evenimente dependente este

$$P(A_1 \cap A_2 \cap A_3 \cap \ldots \cap A_n) =$$
$$= P(A_1) \cdot P_{A_1}(A_2) \cdot P_{A_1 \cap A_2}(A_3) \cdot \ldots \cdot P_{A_1 \cap A_2 \cap A_3 \cap \ldots \cap A_{n-1}}(A_n).$$

Exemplul 2. Fie dată o urnă cu bile de două culori 6 albe și 4 negre. Să se determine probabilitatea că la două extrageri succesive a câte o bilă, cu întoarcere, să obținem o bilă albă şi una neagră.

Rezolvare. Notăm evenimentul A_1 – extragerea unei bile albe, evenimentul A_2 – extragerea unei bile negre și evenimentul A – extragerea unei bile albe şi apoi a unei bile negre. Deci

$$P(A_1) = \frac{6}{10} = \frac{3}{5} = 0{,}6$$

și

$$P(A_2) = \frac{4}{10} = \frac{2}{5} = 0{,}4.$$

Deoarece evenimentele A_1 și A_2 sunt independente, avem

$$P(A) = P(A_1 \cap A_2) = P(A_1) \cdot P(A_2) = 0{,}6 \cdot 0{,}4 = 0{,}24.$$

Răspuns: $P(A) = 0{,}24$

Dacă evenimentele sunt independente, atunci probabilitatea evenimentului A_2 în dependenţă de A_1 coincide cu $P(A_2)$

$$P_{A_1}(A_2) = \frac{P(A_1 \cap A_2)}{P(A_1)} = \frac{P(A_1) \cdot P(A_2)}{P(A_1)} = P(A_2)$$

și invers

$$P_{A_2}(A_1) = P(A_1).$$

Exerciții și probleme propuse

1. *Fie două urne cu bile albe și negre. Prima urnă conține 5 bile albe și 7 negre, iar a doua 3 bile albe și 3 negre. Dintr-o urnă luată la întâmplare se extrage o bilă, să se determine probabilitatea că bila extrasă este albă.*
2. *Două mașini produc una și aceeași piesă. Prima mașină produce piese bune cu probabilitatea de 0,96, iar a doua mașină cu probabilitatea de 0,93. Se ia pentru probă câte o piesă de la fiecare mașină, să se determine probabilitatea că piesele sunt bune.*
3. *Se ia la întâmplare un număr întreg pozitiv. Să se determine probabilitatea că acest număr se împarte la 2 sau la 3 fără rest.*
4. *O bunicuță vinde la bucată 15 trandafiri albi și 5 roșii. Să se determine probabilitatea că printre primii patru trandafiri vânduți să nu fie nici unul roșu.*
5. *Într-o urnă sunt 22 de bile dintre care 10 bile albe, 7 bile negre și 5 bile roșii. Se extrag succesiv, fără întoarcere, trei bile. Să de determine probabilitatea că:*
   a) *cele trei bile sunt albe;*
   b) *numai una este albă;*
   c) *bilele sunt extrase în ordinea alb, negru și roșu.*

§1.6. Probabilitatea totală

Fie un sistem complet de evenimente $A_1, A_2, A_3, \dots, A_n$. Evenimentul A nu se produce sigur, dar se produce împreună cu unul din celelalte evenimente, adică

$$A = (A_1 \cap A) \cup (A_2 \cap A) \cup \dots \cup (A_n \cap A).$$

Probabilitatea evenimentului *A* poate fi calculată aplicând probabilitatea la fiecare termen din ultima expresie

$$P(A) = P(A_1 \cap A) + P(A_2 \cap A) + \cdots + P(A_n \cap A),$$

sau

$$P(A) = P(A_1)P_{A_1}(A) + P(A_2)P_{A_2}(A) + \cdots + P(A_n)P_{A_n}(A). \quad (9)$$

Ultima egalitate reprezintă *formula de calculare a propabilităţii totale*.

Teorema 1. *Probabilitatea evenimentului A, care poate să se producă condiţionată de unul din evenimentele $A_1, A_2, A_3, \dots, A_n$, care este un sistem complet de evenimente, este egală cu suma produselor dintre probabilităţile acestor evenimente şi probabilităţile condiţionate ale evenimentului A.*

Deoarece $A_1, A_2, A_3, \dots, A_n$ este un sistem complet de evenimente, rezultă

$$A = (A_1 \cap A) \cup (A_2 \cap A) \cup \dots \cup (A_n \cap A).$$

Evenimentele din paranteze sunt incompatibile, adică se exclud unele pe altele

$$A_i \cap A_j = \emptyset \quad (i \neq j, \quad i,j = \overline{1,n}),$$

sau

$$(A_i \cap A) \cup \left(A_j \cap A\right) = \emptyset.$$

Deci

$$P(A) = P(A_1 \cap A) + P(A_2 \cap A) + \cdots + P(A_n \cap A) =$$
$$= \sum_{i=1}^{n} P(A_i \cap A) = \sum_{i=1}^{n} P\left(A_i\right) \cdot P_{A_i}(A).$$

Aplicăm teorema de adunare a probabilităţilor evenimentelor incompatibile și obţinem

$$P(A) = P(A_1 \cap A) + P(A_2 \cap A) + \cdots + P(A_n \cap A),$$

însă pe de altă parte, utilizând proprietăţile probabilităţilor, în cazul evenimentelor dependente, se obţine

$$P(A) = P(A_1) \cdot P_{A_1}(A) + P(A_2) \cdot P_{A_2}(A) + \cdots + P(A_n) \cdot P_{A_n}(A).$$

Exemplul 1. În depozitul unei uzine se află piese de acelaşi tip, provenite de la trei secţii ale uzinei date. Se cunoaşte că prima secţie produce 25% din totalul pieselor, a doua 35% şi a treia 40%, iar rebutul reprezintă: de la prima secţie 2%, de la a doua 3% şi de la a treia 1%. Să se determine probabilitatea că o piesă luată la întâmplare, din depozit, să fie rebut.

Rezolvare. Notăm: evenimentul A_1 - piesa provine de la prima secţie;
evenimentul A_2 – piesa provine de la a doua secţie;
evenimentul A_3 – piesa provine de la a treia secţie;
evenimentul A – piesa este rebut.

Evenimentul *A* poate fi reprezentat ca reuniunea intersecției evenimentelor A_i și *A*

$$A = (A_1 \cap A) \cup (A_2 \cap A) \cup (A_3 \cap A).$$

Cunoaștem probabilitățile evenimentelor A_i

$$P(A_1) = 0{,}25, \quad P(A_2) = 0{,}35$$

și

$$P(A_3) = 0{,}4.$$

De asemenea, din condițiile problemei avem

$$P_{A_1}(A) = 0{,}02, \quad P_{A_2}(A) = 0{,}03$$

și

$$P_{A_3}(A) = 0{,}01.$$

În baza formulei (9), determinăm probabilitatea evenimentului *A*

$$P(A) = 0{,}25 \cdot 0{,}02 + 0{,}35 \cdot 0{,}03 + 0{,}4 \cdot 0{,}01 = 0{,}0195.$$

Răspuns: $P(A) = 0{,}0195$

Exemplul 2. Într-o urnă sunt 26 de cartonașe cu literele alfabetului latin (fără repetare). Se extrag la întâmplare, de 3 ori, câte un cartonaş şi se aşează în ordinea extragerii. Să se determine probabilitatea obţinerii cuvântului ARC.

Rezolvare. Notăm evenimentul A la extragerea a 3 cartonașe, așezate în ordinea succesivă, fără întoarcere, să obținem cuvântul ARC; A_1 – evenimentul ca la prima extragere să obţinem litera A; A_2 – evenimentul ca la a doua extragere să obţinem litera R și A_3 – evenimentul ca la a treia extragere să obţinem litera C.

În aceste condiții, evenimentul A are loc dacă și numai dacă

$$A = A_1 \cap A_2 \cap A_3.$$

Deci

$$P(A) = P(A_1) \cdot P(A_2/A_1) \cdot P(A_3/A_1 \cap A_2) = \frac{1}{26} \cdot \frac{1}{25} \cdot \frac{1}{24} = \frac{1}{15600}.$$

Răspuns: $P(A) = \frac{1}{15600}$

Exerciții și probleme propuse

1. *Într-o urnă sunt nouă bile albe și cinci bile negre, iar în a doua urnă șase bile albe și opt bile negre.*

   a) *Din prima urnă se extrage întâmplător o bilă și se mută în a doua urnă, apoi se extrage o bilă din urna a doua. Să se determine probabilitatea că bila extrasă din urna a doua este neagră.*

   b) *Din prima urnă se mută două bile, extrase întâmplător, în a doua urnă, apoi se extrage o bilă din urna a doua. Să se determine probabilitatea că bila extrasă din urna a doua este albă.*

2. *Un lot cu 1000 de piese este supus unui control de calitate, acest lot va fi respins dacă se va găsi cel puţin un rebut, pe parcursul a cinci verificări consecutive. Să se determine probabilitatea că experţii din comisia de calitate vor respinge lotul cercetat, dacă el conţine 5% rebut.*

3. *Într-un depozit sunt 3 lăzi cu piese: prima ladă are 20 de unităţi dintre care 15 sunt standarte, a doua ladă are 30 de unităţi dintre care 24 sunt standarte, iar a treia ladă are 10 unităţi dintre care 6 sunt standarte. Să se determine probabilitatea că o piesă luată la întâmplare este standartă.*

4. *Fie date 6 urne $U_1, \dots, U_6$, prima și a doua urnă conțin 3 bile albe şi 4 negre; urna a treia, a patra și a cincea conțin 2 bile albe şi 8 negre și a șasea urnă conține 6 bile albe şi 2 negre. Se ia la întâmplare o bilă, să se determine probabilitatea că bila extrasă este albă.*

5. *Dintr-un set de domino se extrage la întâmplare o piesă. Să se determine probabilitatea că a doua piesă extrasă va fi legibilă de pus în continuare.*

§1.7. Formula lui Bayes

În virtutea regulii lui Bayes, pot fi rezolvate probleme de tipul: se consideră un sistem complet de evenimente $A_1, A_2, A_3, \ldots, A_n$ care reprezintă cauzele producerii unui experiment necunoscut, cunoscând probabilităţile $P(A_1), P(A_2), \ldots, P(A_n)$ şi probabilităţile condiţionate

$$P_{A_1}(A), P_{A_2}(A), \ldots, P_{A_n}(A),$$

ce se pot calcula înaintea efectuării unei probe (apriore). Probabilităţile care pot fi calculate după efectuarea probei, se numesc aposterioare.

Fie evenimentul compus $(A_i \cap A)$, unde i este un număr fixat. Conform definiției probabilității condiționate, formulei de înmulțire și respectând formula probabilității totale, rezultă *formula lui Bayes* [1]

$$P(A_i/A) = \frac{P(A_i \cap A)}{P(A)} = \frac{P(A_i) \cdot P(A/A_i)}{\sum_{j \in N} P(A_j)\, P\big(A/A_j\big)}$$

sau

$$P_A(A_i) = \frac{P\big(A_i\big) \cdot P_{A_i}(A)}{P(A_1) \cdot P_{A_1}(A) + P(A_2) \cdot P_{A_2}(A) + \cdots + P(A_n) \cdot P_{A_n}(A)}. \quad (10)$$

Thomas Bayes (1702-1761)

În cazul în care evenimentele A_i $(i = 1, \ldots, n)$ sunt echiprobabile, formula lui Bayes devine

$$P(A_i/A) = \frac{P(A/A_i)}{\sum_{j=1}^{n} P\big(A/A_j\big)}.$$

Formula lui Bayes are numeroase aplicații în jurisprudență, în teoria catastrofelor (cutremure, alunecări de teren etc.), în medicină și în alte domenii legate de cercetările științifice [1].

Exemplul 1. Considerăm condiţiile exemplului 1 din §1.6. *Să se determine probabilitatea că piesa obţinută nestandartă să provină de la prima secţie.*

Rezolvare. În baza formulei (10), determinăm probabilitatea evenimentului $P_A(A_1)$, care poate fi exprimată în fel următor:

$$P_A(A_1) = \frac{P(A_1)P_{A_1}(A)}{P(A_1)P_{A_1}(A) + P(A_2)P_{A_2}(A) + P(A_3)P_{A_3}(A)}.$$

Deci

$$P_A(A_1) = \frac{0{,}25 \cdot 0{,}02}{0{,}25 \cdot 0{,}02 + 0{,}35 \cdot 0{,}03 + 0{,}4 \cdot 0{,}01} = \frac{0{,}005}{0{,}0195} = \frac{10}{39}.$$

Răspuns: $P_A(A_1) = \frac{10}{39}$

Exemplul 2. Un magazin se aprovizionează zilnic de la 3 depozite diferite D_1, D_2, D_3, cu aceleaşi cantităţi globale de marfă, însă în proporţii diferite, în raport cu 2 calităţi ale ei (50%,60%,80%). Un cumpărător procură la întâmplare o unitate de marfă şi se constată că este de calitatea a doua. Să se determine probabilitatea aposteriorii, că marfa dată să provină de la depozitul 3 (D_3).

Rezolvare. Notăm prin evenimentul A_i - marfa provine de la depozitul *i* $(i = \overline{1,3})$ și evenimentul *A* – marfa procurată este de calitatea a doua.

Probabilitățile evenimentelor A_i și anume $P(A_1), P(A_2), P(A_3)$, reprezintă un sistem complet de eveniente, deoarece avem aceleaşi cantităţi de marfă. Aceste evenimente sunt echiprobabile și au probabilitatea

$$P(A_1) = P(A_2) = P(A_3) = \frac{1}{3}.$$

Deoarece

$$A_i \cap A_j = \emptyset \ \ (i \neq j, \quad i,j = \overline{1,n}),$$

atunci

$$(A_i \cap A) \cup \left(A_j \cap A\right) = \emptyset,$$

iar

$$A_1 \cup A_2 \cup A_3 = \Omega.$$

Determinăm probabilitățile condiționate $P_{A_i}(A)$ $(i = 1,2,3)$, obținem

$$P_{A_1}(A) = \frac{5}{10} = \frac{1}{2}, \qquad P_{A_2}(A) = \frac{6}{10} = \frac{3}{5},$$

$$P_{A_3}(A) = \frac{8}{10} = \frac{4}{5}.$$

În baza formulei (10), determinăm probabilitatea evenimentului $P_A(A_3)$

$$P_A(A_3) = \frac{\frac{1}{3} \cdot \frac{4}{5}}{\frac{1}{3} \cdot \frac{1}{2} + \frac{1}{3} \cdot \frac{3}{5} + \frac{1}{3} \cdot \frac{4}{5}} = \frac{\frac{4}{15}}{\frac{1}{6} + \frac{3}{15} + \frac{4}{15}} = \frac{\frac{4}{15}}{\frac{1}{6} + \frac{7}{15}} =$$

$$= \frac{\frac{4}{15}}{\frac{5+14}{30}} = \frac{\frac{4}{15}}{\frac{19}{30}} = \frac{4 \cdot 30}{15 \cdot 19} = \frac{8}{19}.$$

Răspuns: $P_A(A_3) = \frac{8}{19}$

Exerciții și probleme propuse

1. *Ion are 10 monede dintre care una este falsă (pe ambele părţi este stema). Vasile extrage o monedă la întâmplare şi o aruncă de 3 ori. Să se determine probabilitatea:*
   - a) *extragerii unei monede false;*
   - b) *extragerii unei monede bune;*
   - c) *căderii de 3 ori a stemei;*
   - d) *căderii de 3 ori a stemelor de pe moneda falsă.*

2. *Trei uzine furnizează unei întreprinderi aceeași tip de marfă în proporție de 20%, 25% și 55%, iar marfa rebut constituie respectiv 4%, 3% și 2%. În baza garanției, o cantitate de marfă în valoare de 4000 lei, care a fost vândută, este restituită întreprinderii, iar suma respectivă este restituită cumpărătorului. Deoarece nu se cunoaște de la care uzină provine marfa rebut, să se determine sumele ce trebuie imputate uzinelor respective.*

3. *Imprimanta CANON este produs la trei uzine diferite în proporți de 25%, 30% și 45%, iar rebutul constituie respectiv 4%, 3% și 2%. Se procură la întâmplare o imprimantă, să se determine probabilitatea că imprimanta provine de la prima uzină, de la a doua și respectiv de la a treia.*

4. *Piesele fabricate de o secţie a uzinei "Ionel", trec pe la unul din cei doi controlori ai uzinei. Probabilitatea că trece pe la primul este 60%, iar la al doilea 40%. Însă, secţia produce inclusiv şi piese nestandarte, probabilitatea că la primul conrolor trec piese bune este de 0,94, iar la celălalt 0,98. Se cunoaşte că o piesă extrasă la întâmplare s-a dovedit a fi standartă. Să se determine probabilitatea că a fost verificată de primul controlor.*

5. *În clasa a 12-a a liceului "Mihail Sîrghi" sunt 25 de elevi, dintre care 3 elevi au un nivel ridicat de pregătire, 19 elevi au un nivel mediu și 3 elevi au un nivel scăzut. În conformitate cu datele statistice, probabilitățile de promovare a examenului de BAC sunt: 0,95; 0,7 și 0,4. Se cunoaște că un elev a promovat examenul de BAC. Să se determine probabilitatea că:*
   - a) *elevul a fost foarte bine pregătit;*
   - b) *elevul a fost pregătit mediu;*
   - c) *elevul a fost prost pregătit.*

6. *Două mașini fabrică acelaşi tip de piese cu aceeaşi productivitate, și probabilitatea de piese rebut respectiv 0,05 şi 0,02. Pentru realizarea unui control de calitate, se extrage la întâmplare o piesă.*
   - a) *Să se determine probabilitatea ca piesa extrasă să fie rebut.*
   - b) *Să se determine probabilitatea ca piesa extrasă să fie standart.*

c) *Dacă se cunoaște că piesa extrasă este rebut, să se determine probabilitatea că ea provine de la prima maşină.*
d) *Dacă se cunoaște că piesa extrasă este standart, să se determine probabilitatea că ea provine de la a doua maşină.*

§1.8. Scheme clasice de probabilitate

Sub această denumire se întâlnesc câteva modele de experimente, care conduc la calculul rapid al probabilităţii unor evenimente care se produc sau apar în condiţii analoage cu cele ale experimentului model.

Cu alte cuvinte, probabilitatea poat fi calculată în baza unor formule sau scheme de calcul, indiferent de experimentul considerat, evitând procedeele de calcul care aduc la definiţia clasică a probabilităţii.

1. Schema lui Bernoulli cu bila întoarsă (binomială). Schema lui Bernoulli se aplică în cazul repetărilor experimentelor independente şi la fiecare repetare se are în vedere apariţia unui eveniment A cu aceeași probabilitate de apariție p. Se cere de calculat probabilitatea că eveniment A, la n repetări ale experimentului, să se realizeze exact de k ori.

Jakob Bernoulli (1655-1705)

În acest caz, spațiul evenimentelor elementare Ω conține 2^n evenimente.

Modelul probabilistic se realizează cu ajutorul unei urne cu bile de două culori: albe şi negre. Se extrage câte o bilă din urnă, se înregistrează culoarea bilei extrase, apoi bila se întoarce înapoi în urnă. Acest experiment se repetă de n ori.

Se consideră experimentul care cere determinarea probabilității că din n bile extrase, k din ele să fie albe.

Notăm, evenimentul A – din n bile extrase k din ele să fie albe;

evenimentul A_i – la extragerea i se obţine o bilă albă;

evenimentul $\overline{A_i}$ – la extragerea i se obţine o bilă neagră.

Fie că urna examinată conține în total N bile, dintre care a bile albe și b bile negre. Deci $N = a + b$, iar probabilitatea extragerii unei bile albe și respectiv a unei bile negre este

$$P(A_i) = \frac{a}{N} \quad și \quad P(\bar{A_\iota}) = \frac{b}{N}.$$

Probabilitatea unei succesiuni, în care evenimentui A s-a realizat de k ori, iar contrarul său de $n - k$ ori, este

$$P(A_{k,n-k}) = P(A_1 \cap A_2 \cap \dots \cap A_k \cap \bar{A}_{k+1} \cap \bar{A}_{k+2} \cap \dots \cap \bar{A}_n) =$$
$$= \left(\frac{a}{N}\right)^k \cdot \left(\frac{b}{N}\right)^{n-k} = p^k \cdot q^{n-k},$$

unde p şi q sunt probabiităţile succesului şi insuccesului (p – succesul și q – insuccesul)

$$p + q = 1.$$

Numărul tuturor evenimentelor elementare, în care evenimentul A se realizează în k experimente din n efectuate, este egal cu numărul permutărilor cu repetiție de n elemente, din care de k ori să obținem bila albă și evident de $n - k$ ori să obţinem bila neagră, adică

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

În caz general, probabilitatea că din n bile extrase, k din ele să fie albe, este [1]

$$P_n(k) = C_n^k p^k q^{n-k} \quad (k = 0,1,\dots,n). \tag{11}$$

Ultima relație reprezintă ***formula lui Bernoulli*** sau ***formula shemei binomiale***.

Considerăm dezvoltarea binomului lui Newton

$$(p_x + q)^n = p^n x^n + \dots + q^n = \sum_{k=0}^{n} C_n^k p^k q^{n-k} x^k = \sum_{k=0}^{n} P_n(k) x^k.$$

Observăm că probabilitatea $P_n(k)$ coincide cu coeficientul de pe lângă x^k.

Din relația (10), rezultă următoarele cazuri:

1) probabilitatea că evenimentul A să nu apară nici o dată la realizarea a n experimente este
$$P_n(0) = q^n;$$
2) probabilitatea că evenimentul A să apară cel puțin o dată este
$$P_n(\geq 1) = 1 - q^n;$$
3) probabilitatea că evenimentul A să apară cel puțin de k ori este
$$P_n(\geq k) = \sum_{i=k}^{n} P_n(i) = 1 - \sum_{i=0}^{k-1} P_n(i)\,;$$

4) probabilitatea că evenimentul *A* să apară cel mult de k ori este

$$P_n(\leq k) = \sum_{i=0}^{k} P_n(i).$$

Exemplul 1. O unitate hotelieră se consideră normal ocupată, dacă este utilizat cel puţin 80% din capacitatea totală. Dintr-un studiu statistic s-a obţinut, probabilitatea ca hotelul să fie ocupat normal într-o zi este egală cu $\frac{7}{8}$. Să se determine probabilitatea că unitatea hotelieră este normal ocupată în 5 zile din cele 7 ale săptămânii.

Rezolvare. Din condițiile problemei rezultă

$n = 7$ (numărul total de zile cercetate);

$k = 5$ (numărul de zile în care hotelul să fie normal ocupat).

Se cunoaște probabilitatea că hotelul să fie ocupat normal într-o zi, este

$$p = \frac{7}{8}.$$

Determinăm probabilitatea ca hotelul să nu fie ocupat normal într-o zi, adică probabilitatea insuccesului

$$q = 1 - p = 1 - \frac{7}{8} = \frac{1}{8}.$$

În baza formulei (11), calculăm probabilitatea

$$P_n(k) = C_7^5 \left(\frac{7}{8}\right)^5 \left(\frac{1}{8}\right)^2 \approx 0{,}17.$$

Răspuns: $P_n(k) \approx 0{,}17$

2. Schema polinomială. Fie A_i $(i = 1, \dots, r)$ un sistem complet de eveminente independndte cu probabilitățile respective p_i $(i = 1, \dots, r)$, încât

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

Se cere să calculăm probabilitatea că fiecare eveminent A_i, să se realizeze de k_i $(i = 1, \dots, r)$ ori, iar

$$\sum_{i=1}^{r} k_i = n.$$

Cu alte cuvinte, fie dată o urnă cu *N* bile de *r* culori, notăm cu c_i $(i = 1, .., r)$ numărul bilelor de culoarea a_i $(i = 1, .., r)$. Se efectuază *n* extrageri succesive cu întoarceri. Fie evenimentul α_i – extragerea bilei de culoarea a_i, atunci

$$P(a_i) = \frac{c_i}{n}.$$

Observăm că numărul de eveniente elementare este egal cu numărul permutărilor cu repetiție ce se pot face cu cele n elemente, din care α_i sunt egale cu evenimentele considerate. Deci probabilitatea este

$$P_n(\alpha_1, \alpha_2, \dots, \alpha_r) = \frac{n!}{\alpha_1!\ \alpha_2!\ \dots\ \alpha_r!} p_1^{\alpha_1} p_2^{\alpha_2} \dots p_s^{\alpha_r}. \qquad (12)$$

Ultima relație reprezintă ***schema probabilistică polinomială***.

Exemplul 2. La o mașină de spălat rufe apar trei tipuri de defecțiuni: pentru tipul A_1 avem 30%, pentru tipul A_2 avem 60% și pentru tipul A_3 avem 10%. Să se determine probabilitatea că din opt defecțiuni să avem două de tipul A_1, cinci de tipul A_2 și respectiv una de tipul A_3.

Rezolvare. Această problemă corespunde schemei polinomiale, deci

$$\alpha_1 = 2, \qquad \alpha_2 = 5, \qquad \alpha_3 = 1, \qquad n = 8$$

și probabilitățile respective

$$p_1 = 0{,}3; \quad p_2 = 0{,}6; \quad p_3 = 0{,}1.$$

În baza formulei (12), determinăm probabilitatea că din opt defecțiuni să avem două de tipul A_1, cinci de tipul A_2 și respectiv una de tipul A_3

$$P_8(2{,}5{,}1) = \frac{8!}{2!\ 5!\ 1!}(0{,}3)^2 \cdot (0{,}6)^5 \cdot 0{,}1 \approx 0{,}11.$$

Răspuns: $P_8(2{,}5{,}1) \approx 0{,}11$

3. Schema biomială cu bila neîntoarsă (hipergeometrică). Fie dată o urnă ce conţine bile de două culori: albe şi negre. Se extrag bilele din urnă, una câte una. Să se calculeze probabilitatea că din n bile extase, k din ele să fie de culoarea albă şi respectiv $n-k$ de culoare neagră.

Numărul total de cazuri se calculează cu ajutorul combinărilor C_{a+b}^n, unde a este numărul de bile albe, iar b este numărul de bile negre.

Pentru a calcula probabilitatea extragerii bilei albe, de asemenea aplicăm formula combinărilor, și nume C_a^k, respectiv şi pentru calcularea extragerii bilei negre avem C_b^{n-k}. Deci probabilitatea că din n bile extase, k din ele să fie de culoarea albă şi respectiv $n-k$ de culoare neagră este

$$P(n,k) = \frac{C_a^k C_b^{n-k}}{C_{a+b}^n} \quad (a \geq b \geq n, k). \qquad (13)$$

Această formulă poate fi generalizată în felul următor: fie date bile de r culori, notăm a_1 – bila de culoarea 1, a_2 – bila de culoarea 2, ..., a_r – bila de culoarea r. Se extrag n bile fără întoarcere. Să se determine probabilitatea că din n bile k_1 – să fie de culoarea 1, k_2 – să fie de culoarea 2, ..., k_r – să fie de culoarea r. În acest caz vom avea

$$P(n,\ k_1,\ k_2, \dots,\ k_r) = \frac{C_{a_1}^{k_1} C_{a_2}^{k_2} \dots C_{a_r}^{k_r}}{C_{a_1+\ a_2+\dots+\ a_r}^{k_1+\ k_2+\dots+k_r}}. \quad (14)$$

4. Schema lui Poisson. Această schemă se aplică în cazul în care se fac repetări independente ale unui experiment şi la fiecare repetare se are în vedere un anumit eveniment ce apare, în general, cu probabilităţi diferite la fiecare repetare. Se cere de calculat probabilitatea că la n repetări ale experimentului, evenimentul A să apară de k ori.

Simeon Denis Poisson (1781-1840)

Modelul probabilistic poate fi interpretat ca o mulţime din n urne fiecare cu bile de 2 culori, dar în proporţii diferite.

Notăm: p_i - probabilitatea extragerii bilei albe din urna *i*;

q_i - probabilitatea extragerii bilei negre din urna *i*.

Probabilitatea realizării unui număr de k evenimente din cele n, este dată de coeficientul de pe lângă x^k din dezvoltarea binomului lui Newton

$$(p_1 x + q_1)(p_2 x + q_2) \dots (p_n x + q_n). \quad (15)$$

Exemlul 3. Se consideră 4 urne, fiecare cu bile albe și bile negre. În prima urnă avem 50 bile dintre care 10 sunt albe, în a doua urnă 30 bile dintre care 5 sunt albe, în a treia urnă 10 bile dintre care 2 sunt albe, iar în ultima urnă 25 bile dintre care 10 sunt albe. Se se determine probabilitatea că, extrăgând din fiecare urnă câte o bilă, să obținem 3 bile albe.

Rezolvare. Notăm următoarele evenimente independente:

A_1 – bila extrasă din urna U_1 este albă;

A_2 – bila extrasă din urna U_2 este albă;

A_3 – bila extrasă din urna U_3 este albă;

A_4 – bila extrasă din urna U_4 este albă;

A – la extragerea a câte o bilă din cele patru urne să obținem trei bile albe.

Determinăm probabilitățile pentru fiecare eveniment

$$p_1(A_1) = \frac{10}{50} = \frac{1}{5} \Rightarrow q_1(A_1) = \frac{4}{5},$$

$$p_2(A_2) = \frac{5}{30} = \frac{1}{6} \Rightarrow q_2(A_2) = \frac{5}{6},$$
$$p_3(A_3) = \frac{2}{10} = \frac{1}{5} \Rightarrow q_3(A_3) = \frac{4}{5},$$
$$p_4(A_4) = \frac{10}{25} = \frac{2}{5} \Rightarrow q_4(A_4) = \frac{3}{5}.$$

Probabilitatea că din cele 4 bile extrase, 3 să fie albe, este coeficientul de pe lângă x^3 din dezvoltarea binomului lui Newton (15)

$$\left(\frac{1}{5}x + \frac{4}{5}\right)\left(\frac{1}{6}x + \frac{5}{6}\right)\left(\frac{1}{5}x + \frac{4}{5}\right)\left(\frac{2}{5}x + \frac{3}{5}\right) = \frac{1}{25}(x+4)^2 \cdot \frac{x+5}{6} \cdot \frac{2x+3}{5} =$$
$$= \frac{(x^2 + 8x + 16)(2x^2 + 13x + 15)}{750} =$$
$$= \frac{2x^4 + 13x^3 + 15x^2 + 16x^3 + 104x^2 + 120x + 32x^2 + 208x + 240}{750} =$$
$$= \frac{2x^4 + 29x^3 + 151x^2 + 328x + 240}{750}$$

Obţinem, probabilitatea că din cele 4 bile extrase, 3 să fie albe este

$$P(A) = \frac{29}{750}.$$

Răspuns: $P(A) = \frac{29}{750}$

5. Schema lui Pascal. Fie un experiment ce se repetă în aceleaşi condiţii, iar repetările sunt independente. La fiecare repetare se urmăreşte apariţia aceluiaşi eveniment, care apare cu aceeaşi probabilitate. Se cere să determinăm probabilitatea ca până la cea de-a n-a apariţie a evenimentului considerat, să se realizeze contrariul acestui eveniment de k ori.

Modelul probabilistic se realizează printr-o urnă ce conţine bile de două culori, albe şi negre. Se extrag bile din urnă, una câte una, cu întoarcerea bilei extrase în urnă, după ce s-a constatat culoarea ei. Vom spune ca avem succes când se obţine bila de culoare albă, şi respectiv insucces când se obţine bila de culoare neagră. La fiecare extragere succesul apare cu probabilitatea p, iar insuccesul cu probabilitatea $q = 1 - p$.

Este necesar să determinăm probabilitatea că la apariţia celui de-al n-lea succes, să se obţină k insuccese.

Notam evenimentul $B_{n,k}$ - până la apariţia celui de-al n- lea succes să se obţină k insuccese. Deci

$$B_{n,k} = A_{n-1} \cap A_{n+k}$$

unde evenimentul A_{n-1} - la primele $n+k-1$ extrageri să se obţină $n-1$ succese şi k insuccese, iar evenimentul A_{n+k} - la extragerea de rang $n+k$, să se obţină succes.

Evenimentele sunt independente, deci

$$P(B_{n,k}) = P(A_{n-1})P(A_{n+k}).$$

Aplicând schema lui Bernoulli, obține

$$P(A_{n-1}) = C_{n+k-1}^{n-1} p^{n-1} q^k.$$

Deci

$$P(B_{n,k}) = C_{n+k-1}^{n-1} p^n q^k. \tag{16}$$

Ultima egalitatea reprezintă ***schema lui Pascal***.

Blaise Pascal (1623-1662)

Exemplul 4. În urma unui studiu statistic, s-a constatat, probabilitatea ca o persoană care intră într-un magazin să devină cumpărător este $p = 75\%$. *Să se determine probabilitatea că până la primul cumpărător în magazin să fi intrat 5 clienți.*

Rezolvare. În conformitate cu schema lui Pascal (16), avem

$$n = 1, k = 5, p = 0{,}75 \text{ și } q = 0{,}25.$$

Obținem

$$P(1{,}5) = C_5^0 \cdot 0{,}75 \cdot (0{,}25)^5 = \frac{3}{4096}.$$

Răspuns: $P(1{,}5) = \frac{3}{4096}$

6. Teorema locală a lui Laplace. Schema lui Bernoulli, în special formula

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

în cazul când n este destul de mare, duce la calcule foarte mari şi complicate.

O soluţie de calculare a probabilităţii a fost propusă de Laplace.

În cazul când probabilitatea apariţiei evenimetului A este p şi aceeaşi la orice probă, atunci probabilitatea

$$P(n,k)(A) = P_n^k(A),$$

poate fi calculată aproximativ pentru n destul de mare, însă cu exactitate foarte bună.

Teorema locală a lui Laplace. *Dacă probabilitatea evenimentului A este egală cu* p, *este aceiaşi la orice probă şi*

$$P(A) \neq 0 \text{ şi } P(A) \neq 1,$$

atunci $P_n^k(A)$ *se aproximează cu valoarea funcţiei*

$$P_n^k(A) \approx \frac{1}{\sqrt{npq}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} = \frac{1}{\sqrt{npq}} \varphi(x), \qquad (17)$$

unde

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

iar

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

Pierre-Simon de Laplace (1749-1827)

Observația 1. *Valorile funcţiei* $\varphi(x)$ *sunt date tabelar (vezi anexe, Tabelul A1), funcţia dată este pară* $\varphi(x) = \varphi(-x)$, *iar pentru valori mai mari ca 5, valoarea funcţiei este aproximativ nulă.*

S-a constatat [4], cu cât *n* este mai mare, cu atât exactitatea va fi mai bună.

Exemplul 5. Să se determine probabilitatea apariţiei evenimentului A, exact de 80 ori din 400 de probe, dacă se ştie că probabilitatea apariţiei evenimentului A este 0,2.

Rezolvare. Cunoaştem $n = 400, k = 80$ și probabilitatea $p = 0{,}2$, deci probabilitatea insuccesului este

$$q = 1 - 0{,}2 = 0{,}8.$$

În baza formulei (17) și a tabelelor de valori a funcției $\varphi(x)$ (tabelul A1), obținem

$$x = \frac{80 - 400 \cdot 0{,}2}{\sqrt{400 \cdot 0{,}2 \cdot 0{,}8}} = 0;$$

$$\varphi(0) = \frac{1}{\sqrt{2\pi}} e^{-\frac{0^2}{2}} \approx 0{,}3989$$

și

$$P_{400}^{80}(A) \approx \frac{1}{\sqrt{400 \cdot 0{,}2 \cdot 0{,}8}} \cdot \varphi(x) = \frac{1}{8} \cdot \varphi(0) \approx \frac{1}{8} \cdot 0{,}3989 = 0{,}04986$$

De asemenea, pentru calcularea probabilitaţii, putem aplica şi formula lui Bernoulli (11)

$$P(400{,}80)(A) = C_{400}^{80}(0{,}2)^{80}(0{,}8)^{320} = \frac{400!}{80!\,320!}(0{,}2)^{80}(0{,}8)^{320} \approx 0{,}04986$$

Răspuns: $P_{400}^{80}(A) \approx 0{,}04986$

Exemplul 6. Probabilitatea de a nimeri în 100, în jocul Dats, la o aruncare, este de 75%. Să se determine probabilitatea că din 10 aruncări, vom nimeri exact de 8 ori din 100.

Rezolvare. Cunoaștem $n = 10, k = 8$ și probabilitatea $p = 0{,}75$, deci probabilitatea insuccesului este

$$q = 1 - 0{,}75 = 0{,}25.$$

În baza formulei (17) și a tabelelor de valori a funcției $\varphi(x)$ (tabelul A1), obținem

$$x = \frac{8 - 10 \cdot 0{,}75}{\sqrt{10 \cdot 0{,}75 \cdot 0{,}25}} = 0{,}36,$$

$$\varphi(0{,}36) \approx 0{,}3739.$$

Deci

$$P_{10}^{8}(A) \approx \frac{1}{\sqrt{10 \cdot 0{,}75 \cdot 0{,}25}} \cdot \varphi(0{,}36) = 0{,}7301 \cdot \varphi(0{,}36) \approx$$

$$\approx 0{,}7301 \cdot 0{,}3739 = 0{,}273.$$

De asemenea, ca și în exemplul de mai sus, putem aplica şi formula lui Bernoulli (11)

$$P(10{,}8)(A) = C_{10}^{8} \cdot (0{,}75)^8 \cdot (0{,}25)^2 \approx 0{,}282.$$

Răspuns: $P_{10}^{8}(A) \approx 0{,}28$

Analizând aceste exemple, putem concluziona: formula lui Laplace nu poae fi utilizată pentru n mic, dar ne indică exactitatea necesară pentru n destul de mare.

Notăm cu $\lambda = np$ numărul mediu de apariţie a evenimentului A în n experimente. Pentru n destul de mare, iar p – mic, putem aplica formula lui Poisson

$$P_n^k(A) \approx \lim_{n\to\infty} C_n^k \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} = \frac{\lambda^k}{k!} e^{-\lambda}.$$

Ultima egalitate reprezintă ***formula lui Poisson***. Această formulă se utilizează în special la soluţionarea problemelor legate de fluxurile de evenimente elementare, cum ar fi apelurile înregistrate de staţia telefonică centrală, apariţia cumpărătorilor la magazine.

Observația 2. *Valorile repartiției Poisson*

$$P\{X = k\} = \frac{\lambda^k}{k!} e^{-\lambda},$$

sunt date tabelar (vezi tabelul A3).

Dacă notăm cu λ intensitatea fluxului de evenimente, iar p este probabilitatea apariţiei evenimentului A în intervalul de timp t, atunci probabilitatea că în intervalul t va apărea evenimentul A de k ori, se calculează cu ajutorul formulei

$$P_k(A) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}. \tag{18}$$

Exemplul 7. La fabrica "Floarea", în urma unui control s-a determinat: probabilitatea pieselor nestandarte este 0,004. Să se determine probabilitatea că printre 1000 de piese fabricate, 5 vor fi nestandarte.

Rezolvare. Determinăm valoarea parametrului λ, $\lambda = 1000 \cdot 0{,}004 = 4$, și aplicând formula (18), obținem

$$P_{1000}^5(A) = \frac{4^5}{5!} e^{-4} \approx 0{,}168.$$

Răspuns: $P_{1000}^5(A) \approx 0{,}168$

Exerciții și probleme propuse

1. *Se aruncă o monedă de 20 ori. Să se determine probabilitatea că stema apare exact de 10 ori.*
2. *Se aruncă un zar de trei ori. Să se determine probabilitatea obținerii de trei ori a unei fețe cu un număr mai mic ca trei.*
3. *Fie patru telefoane de același tip ce sunt implicate în experimente diferite. Probabilitatea ca fiecare din cele patru telefoane să se defecteze este* $p_1 = 10\%, p_2 = 20\%, p_3 = 30\%,\ p_4 = 40\%.$ *Să se determine probabilitatea că*
   a) *nici un telefon nu se va defecta;*
   b) *se va defecta un telefon;*
   c) *se vor defecta două telefoane;*

d) se vor defecta trei telefoane;

e) se vor defecta patru telefoane.

4. Fie date 3 urne ce conţin: prima urnă conține 3 bile albe şi 2 negre, a doua urnă conține o bilă albă şi 4 negre și a treia urnă conține 2 bile albe şi 3 negre. Se extrage din fiecare urnă câte o bilă, să se determine probabilitatea extragerii unei bile albe și două bile negre.

5. O fabrică a expediat o comandă de 5000 frigidere. Probabilitatea că în timpul deplasării fiecărui frigider i, se produc defecte, este de 0,0002. Să se determine probabilitatea că în urma deplasării vor fi exact 3 frigidere defecte.

6. La staţia telefonică 112, se fece legătură, în mediu, cu 3 apeluri pe minut. Să se determine probabilitatea că în timp de două minute se vor efectua:

a) exact 4 apeluri;

b) nu mai mult de 4 apeluri;

c) nu mai puţin de 4 apeluri, inclusiv.

7. Fie un experiment din 5 probe independente. Să se determine probabilitatea că evenimentul A va apărea nu mai puţin de 2 ori, dacă la fiecare din ele probabilitatea apariţiei este de 0,3.

8. Evenimentul A apare numai atunci când un evenimentul B va apărea nu mai puţin de 2 ori. Să se determine probabilitatea apariţiei evenimentului A, dacă se efectuează 6 probe independente şi la fiecare din ele, probabilitatea apariţiei evenimentului B este 0,4.

9. Se cunoaşte probabilitatea încolțirii semințelor p=0,8. Să se determine probabilitatea că din 1000 de seminţe de floarea-soarelui nu vor încolţi 200.

10. Se cunoaște că probabilitatea de a avea un băiat este de 0,52. Să se determine probabilitatea ca printre 200 de nou-născuți să se nască exact:

a) 100 de băieți;

b) 110 de băieți;

c) 60 de fete.

11. Două personae participă într-un joc format din mai multe partide. Primul participant câştigă o partidă cu probabilitatea egală cu $\frac{1}{3}$. Să se determine probabilitatea că:

a) prima partidă câştigată de primul participant, se produce după 5 partide pierdute;

b) a treia partidă câştigată de primul participant, se produce după un total de 6 partide pierdute.

12. Piesele produse de o uzină sunt puse la 2 maşini control, independente una faţă de alta, probabilitatea ca piesele să treacă prima mașină

este $\frac{2}{3}$ *şi respectiv să treacă la a doua mașină este* $\frac{3}{4}$. *Să se determine probabilitatea că din 5 piese extrase la întâmplare, 2 să treacă ambele teste, una doar primul test, una doar al doilea test şi una să nu treacă nici un test.*

13. *Într-o cutie sunt 12 bile marcate cu cifra 1, 8 bile marcate cu cifra 3 şi 6 bile marcate cu cifra 5. Se extrag la întâmplare 4 bile, să se determine probabilitatea că suma cifrelor marcate pe bile să fie cel mult 13.*

14. *În urma unui studiu statistic s-a constatat probabilitatea ca o persoană care intră într-un magazin să devină cumpărător este* $p = 75\%$. *Să se determine probabilitatea că până la al treilea cumpărător în magazin să fi intrat 7 clienți.*

15. *Două persoane participă într-un joc format din mai multe partide. Primul participant câștigă o partidă cu probabilitatea* $p = \frac{1}{3}$ *și o pierde cu probabilitatea* $q = \frac{2}{3}$. *Să se determine probabilitatea că a treia partidă câștigată de primul participant să se producă după un total de șase partide pierdute.*

CAPITOLUL II. VARIABILE ALEATOARE

§2.1. Definiţia variabilei aleatoare

În practică ne întâlnim zilnic cu mărimi care i-au valori ce se schimbă sub influenţa unor factori întâmplători. De exemplu, numărul de copii care vor răci în prima săptămână de grădiniță, numărul de zile cu ploaie din luna octombrie, perioada de funcționare a unei lămpi sau numărul de puncte care apar la aruncarea unui zar. Aici ne interesează doar acele mărimi care iau un număr finit de valori. Fiecare din mărimile de mai sus, poate lua diferite valori, în diverse exeperienţe, însă, modificarea valorilor de referință au la bază factorii întâmplători. Prin urmare, apare necesitatea introducerii noţiunii de variabilă aleatoare, care de fapt reprezintă o funcţie reală definită pe mulţimea evenimentelor elementare cu valori reale.

Fie Ω mulţimea de evenimente elementare asociată unui experiment. Rezultatele posibile se notează cu ω, este posibil ca acestea să nu fie numere, însă le putem asocia și valori numerice.

Definiția 1 [1]. *Se numeşte variabilă aleatoare, o funcţie definită pe mulţimea de evenimente elementare cu valori numerice*

$$X: \Omega \to \mathbb{R}.$$

Este clar că aceaste valori nu pot fi cunoscute înaintea efectuării experimentului, din cauza factorilor întâmplători, care influențează experimentul considerat.

Exemplul 1. O monedă se aruncă de două ori. Să se determine variabila aleatoare care reprezintă numărul de apariții a banului.

Rezolvare. Notăm cu s și b respectiv apariția stemei și banului. În acest caz, spațiul evenimentelor elementare este

$$\Omega = \{bb, bs, sb, ss\}.$$

Fie X o variabilă aleatoare care reprezintă numărul de apariții a banului, deci X este o funcție care asociază fiecărui element un număr real $X(\omega)$ (tabelul 2)

Tabelul 2. Spațiul evenimentelor elementare în cazul aruncării unei monede de două ori

ω_i	bb	bs	sb	ss
$X(\omega)$	2	1	1	0

Observaţia 1. *Numerele rezultatelor ω_i sunt diferite între ele şi nu depăşesc n.*

Considerăm o variabilă aleatoare X, care înregistrează s valori $\{x_1, x_2, \dots, x_n\}$, în condiţiile în care sunt înregistrate n evenimente elementare.

Fie $\omega_{i1}, \omega_{i2}, \dots, \omega_{ik}$ evenimentele pentru care

$$X(\omega_{ih}) = x_{ih} \;\; i = \overline{1, k}$$

atunci, notând probabilităţile $P_{ik} = X(\omega_{ik})$ primim probabilitatea funcţiei X, când ia valorile x_i

$$P(X = x_i) = P_{i1} + P_{ik}.$$

În dependență de valorile pe care le primesc variabilele aleatoare avem: variabile aleatoare dicrete și continue.

Definiția 2. *O variabilă aleatoare se numeşte discretă, dacă ea are o mulțime finită sau infinită numerabilă de valori.*

Exemple de valori aleatoare discrete: numărul de apeluri telefonice la stația 112 într-o unitate de timp; numărul de piese rebut din 100 de piese; numărul de băieți născuți la maternitatea numărul 1, or Chișinău, în luna august; mașini aflate pe șosea; numărul de aruncări ale unui zar până la obținerea unui fețe cu șase puncte, etc.

Definiția 3. *O variabilă aleatoare se numeşte continuă, dacă mulțimea de valori posibile reprezintă un interval finit sau infinit al dreptei reale.*

Variabilele aleatoare continue se întâlnesc în problemele ce studiază fenomenele unde se efectuează diverse măsurări de timp, masă, lungime, etc.: cantitatea de apa de ploaie; duritatea unui anumit material; viteza vântului; viteza unei molecule de gaz, etc.

§2.2. Repartiția variabilei aleatoare discrete

Pentru a defini o variabilă aleatoare discretă trebuie să cunoaștem toate valorile posibile pe care le poate primi. Însă această condiție nu este suficientă, deoarece mai multe variabile aleatoare discrete pot avea aceleași valori, dar cu probabilităţi diferite. De acceea, este necesar să cunoaștem și probabilitățile fiecăror valori.

Definiția 1. *Se numeşte repartiţia sau distribuția variabilei aleatoare discrete, enumerarea valorilor posibile şi a probabilităţii corespunzătoare.*

Repartiţia variabilei aleatoare discrete poate fi reprezentată sub forma

$$X\colon \begin{pmatrix} x_1 & x_2 & x_3 & \dots & x_n \\ p_1 & p_2 & p_3 & \dots & p_n \end{pmatrix},$$

sau sub formă de tabel (tabelul 3), în care prima linie conține toate valorile posibile, iar a doua linie conține probabilitățile valorilor respective care se scriu sub fiecare valoare corespunzătoare datelor.

În caz general, avem următorul tabel:

Tabelul 3. Repartiţia variabilei aleatoare discrete

X	x_1	x_2	…	x_n
P	p_1	p_2	…	p_n

Având în vedere, că în urma unui experiment variabila aleatoare ia numai una din valorile sale posibile, rezultă că evenimentele respective formează un sistem complet de evenimente, ceea ce înseamnă că suma probailităţilor de jos este egală cu unitatea

$$p_1 + p_2 + p_3 + \cdots + p_n = 1.$$

Exempul 1. Se consideră un joc cu zarul, punctajul care se acordă, celui care aruncă zarul, este:

1 punct – dacă apare una din feţele 1 sau 2;

2 puncte – dacă apare una din feţele 3 sau 4;

3 puncte – dacă apare una din feţele 5 sau 6.

Să se determine repartiția variabilei aleatoare definită de acest joc.

Rezolvare. Notăm cu *X* numărul de puncte obţinute de un jucător la aruncarea zarului. În rezultat obținem o variabilă aleatoare cu repartiția

$$X\colon \begin{pmatrix} 1 & 2 & 3 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix}.$$

Exemplul 2. Să se determine repartiția variabilei aleatoare definită ca suma punctelor ce apar în rezultatul aruncării a două zaruri simetrice.

Rezolvare. Spațiul evenimentelor elementare este format din 36 de elemente (vezi §1.1, exemplul 1)

$$\Omega = \{(1,1), (1,2), (1,3), (1,4), \dots, (6,3), (6,4), (6,5), (6,6)\}.$$

Evident, în rezultatul aruncării a două zaruri, suma punctelor ce apar poate avea valoarea de la 2 până la 12. Cercetăm fiecare din aceste evenimente.

Mulțimea ce definește suma punctelor egală cu 2 este $\{(1,1)\}$, deci probabilitatea acestui eveniment este

$$P(s = 2) = \frac{1}{36}.$$

Mulțimea ce definește suma punctelor egală cu 3 este $\{(1,2), (2,1)\}$, deci probabilitatea acestui eveniment este

$$P(s = 3) = \frac{2}{36}.$$

Mulțimea ce definește suma punctelor egală cu 4 este $\{(1,3), (2,2), (3,1)\}$, deci probabilitatea acestui eveniment este

$$P(s = 4) = \frac{3}{36}.$$

Mulțimea ce definește suma punctelor egală cu 5 este $\{(1,4), (2,3)(3,2), (4,1)\}$, deci probabilitatea acestui eveniment este

$$P(s = 5) = \frac{4}{36}.$$

La fel procedăm până la valoarea sumei punctelor egală cu 12. În rezultat obținem următoarea repartiție a variabilei aleatoare

$$X:\begin{pmatrix} 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 \\ \frac{1}{36} & \frac{2}{36} & \frac{3}{36} & \frac{4}{36} & \frac{5}{36} & \frac{6}{36} & \frac{5}{36} & \frac{4}{36} & \frac{3}{36} & \frac{2}{36} & \frac{1}{36} \end{pmatrix}.$$

Operaţii cu variabile aleatoare discrete. Considerăm variabilele aleatoare X și Y cu repartiţiile

$$X:\begin{pmatrix} x_1 & x_2 & x_3 & \dots & x_n \\ p_1 & p_2 & p_3 & \dots & p_n \end{pmatrix} \quad și \quad Y:\begin{pmatrix} y_1 & y_2 & \dots & y_m \\ q_1 & q_2 & \dots & q_m \end{pmatrix}$$

și un număr real α, atunci în conformitate cu [11-12] avem

Definiția 2. *Suma unei variabile aleatoare X cu un scalar α, este notată $\alpha + X$ cu repartiţia*

$$\alpha + X:\begin{pmatrix} \alpha + x_1 & \alpha + x_2 & \dots & \alpha + x_n \\ p_1 & p_2 & \dots & p_n \end{pmatrix}. \tag{19}$$

Definiția 3. *Suma a două variabile aleatoare X şi Y, este o variabilă aleatoare notată $X + Y$ cu repartiţia*

$$X + Y:\begin{pmatrix} x_i + y_j \\ P_{ij} \end{pmatrix}, \quad P_{ij} = p_i q_j, \quad i = \overline{1,n}, \ j = \overline{1,m}, \tag{20}$$

sau

$$X + Y:\begin{pmatrix} x_1 + y_1 & x_1 + y_2 \dots & x_i + y_j \dots & x_n + y_m \\ P_{11} & P_{12} \ \dots & P_{ij} \ \dots & P_{nm} \end{pmatrix}.$$

Definiția 4. *Produsul unei variabile aleatoare X cu un scalar α, este notat αX cu repartiţia*

$$\alpha X:\begin{pmatrix} \alpha x_1 & \alpha x_2 & \dots & \alpha x_n \\ p_1 & p_2 & \dots & p_n \end{pmatrix}. \tag{21}$$

Definiția 5. *Produsul a două variabile aleatoare X şi Y, este o variabilă aleatoare notată $X \cdot Y$ cu repartiţia*

$$X \cdot Y:\begin{pmatrix} x_i \cdot y_j \\ P_{ij} \end{pmatrix}, \quad P_{ij} = p_i q_j, \quad i = \overline{1,n}, \ j = \overline{1,m}, \tag{22}$$

sau

$$X \cdot Y:\begin{pmatrix} x_1 y_1 & x_1 y_2 & \dots & x_i y_j & \dots & x_n y_m \\ P_{11} & P_{12} & \dots & P_{ij} & \dots & P_{nm} \end{pmatrix}.$$

Definiția 6. *Puterea de ordinul k al variabilei aleatoare discrete X, este variabila aleaoare X^k cu repartiţia*

$$X^k:\begin{pmatrix} x_1^k & x_2^k & \dots & x_n^k \\ p_1 & p_2 & \dots & p_n \end{pmatrix}. \tag{23}$$

Exemplul 3. Fie date două varibile aleatoare cu repartiţiile

$$X:\begin{pmatrix} 1 & 2 & 3 \\ 0{,}2 & 0{,}3 & 0{,}5 \end{pmatrix} \quad și \quad Y:\begin{pmatrix} 1 & 2 \\ 0{,}4 & 0{,}6 \end{pmatrix}.$$

Să se determine repartiţiile variabilelor $4 + X, X + Y, 5 \cdot Y, X \cdot Y, Y^3$.

Rezolvare. În baza formulelor (19)-(23), obținem

$$4 + X:\begin{pmatrix} 5 & 6 & 7 \\ 0{,}2 & 0{,}3 & 0{,}5 \end{pmatrix},$$

$$X + Y:\begin{pmatrix} 2 & 3 & 3 & 4 & 4 & 5 \\ 0{,}08 & 0{,}12 & 0{,}12 & 0{,}18 & 0{,}2 & 0{,}3 \end{pmatrix},$$

$$5 \cdot Y:\begin{pmatrix} 5 & 10 \\ 0{,}4 & 0{,}6 \end{pmatrix},$$

$$X \cdot Y:\begin{pmatrix} 1 & 2 & 2 & 4 & 3 & 6 \\ 0{,}08 & 0{,}12 & 0{,}12 & 0{,}18 & 0{,}2 & 0{,}3 \end{pmatrix},$$

$$Y^3:\begin{pmatrix} 1 & 8 \\ 0{,}4 & 0{,}6 \end{pmatrix}.$$

Dacă variabila aleatoare *X* conţine doar valori pozitive, atunci se poate vorbi despre inversa ei X^{-1} cu repartiţia

$$X^{-1}:\begin{pmatrix} \frac{1}{x_1} & \frac{1}{x_2} & \frac{1}{x_3} & \cdots & \frac{1}{x_n} \\ p_1 & p_2 & p_3 & \cdots & p_n \end{pmatrix}.$$

Fie variabilele aleatoare *X* şi Y, cu condiţia că variabila aleatoare Y admite inversa Y^{-1}, atunci prin definiţie, obținem

$$\frac{X}{Y} = X \cdot Y^{-1},$$

unde variabila aleatoare $X \cdot Y^{-1}$ are repartiţia

$$X \cdot Y^{-1}:\begin{pmatrix} \frac{x_1}{y_1} & \frac{x_2}{y_2} & \frac{x_3}{y_3} & \cdots & \frac{x_n}{y_n} \\ P_{11} & P_{22} & P_{33} & \ldots & P_{nn} \end{pmatrix}, \quad P_{ij} = p_i q_j, \quad i = \overline{1,n}, \quad j = \overline{1,m}. \quad (24)$$

Exerciții și probleme propuse

1. *Se aruncă o monedă până la prima apariție a stemei și fie X numărul de repetări. Să se determine repartiția variabilei aleatoare* X.

2. *Se aruncă două zaruri simentrice Să se determine tabelul de repartiție a variabilei aleatoare definită ca*
   a) *suma punctelor ce apar în rezultatul experimentului să fie pătrat perfect;*
   b) *produsul punctelor ce apar în rezultatul experimentului să fie cel puțin 6.*

3. *Să se determine repartiţiile variabilelor* $10 + X, X + 2Y, 5X, 3XY, X^3, Y^2$, *dacă*

$a)\ X: \begin{pmatrix} 0 & 1 & 2 \\ 0{,}4 & 0{,}2 & 0{,}4 \end{pmatrix}$ și $Y: \begin{pmatrix} 4 & 9 \\ 0{,}55 & 0{,}45 \end{pmatrix}$;

$b)\ X: \begin{pmatrix} 3 & 6 & 9 \\ 0{,}1 & 0{,}4 & 0{,}5 \end{pmatrix}$ și $Y: \begin{pmatrix} 5 & 3 & 2 \\ 0{,}3 & 0{,}4 & 0{,}3 \end{pmatrix}$;

$c)\ X: \begin{pmatrix} 5 & 10 & 15 \\ 0{,}3 & 0{,}3 & 0{,}4 \end{pmatrix}$ și $Y: \begin{pmatrix} 0 & 1 & 2 & 3 \\ 0{,}27 & 0{,}25 & 0{,}25 & 0{,}23 \end{pmatrix}$.

4. *Fie o urnă cu bile de două culori: 6 albe și 4 negre. Se extrag de două ori câte o bilă, după înregistrarea culorii se întoarce înapoi. Fie X numărul de bile albe la prima extragere, iar Y numărul de bile albe la a doua extragere. Să se determine repartiția lui X, repartiția lui Y și repartiția sumei $X + Y$.*

5. *Într-o urnă sunt 9 bile albe și 3 negre. Se i-au la întâmplare două bile și fie că variabila aleatoare X reprezintă numărul de bile albe. Să se determine repartiția variabilei aleatoare X.*

6. *Se aruncă o monedă de trei ori. Să se determine repartiția variabilei aleatoare X, care reprezintă numărul de apariții a stemei.*

§2.3. Funcția de repartiție și densitatea variabilei aleatoare

În paragraful precedent, este indicat faptul că variabila aleatoare discretă este caracterizată de repartiția ei, însă acest mod de definire nu este valabil pentru variabilele aleatoare continui, deoarece nu putem enumera toate valorile posibile ale acestui tip de variabile. Pentru caracterizarea cantitivă a aceastei repartiții, nu vom utiliza probabilitatea evenimentului $\{\omega : X(\omega) = x\}$, dar vom utiliza probabilitatea evenimentului $\{\omega : X(\omega) < x\}$, unde x este o variabilă reală. Evident, ultima probabilitate depinde de x, adică probabilitatea este o funcție de x, care se notează $F(x)$.

Definiția 1. *Se numește funcție de repartiție atașată unei variabile aleatoare X, aplicația definită pe $\mathbb{R}$ cu valori în $\mathbb{R}$ ($F: \mathbb{R} \to \mathbb{R}$)*

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

iar dacă X este discretă, având repartiția

$$X: \begin{pmatrix} x_i \\ p_i \end{pmatrix},$$

atunci funcția de repartiție nu este altceva decât funcția

$$F(x) = \sum_{X<x} p_i.$$

Din punct de vedere geometric, funcție de repartiție $F(x)$, reprezintă probabilitatea ca variabila aleatoare X să i se atribuie valori, care pe o axă numerică se reprezintă la stânga punctului x.

În caz general, o funcție de repartiție se definește după tabelul de valori a variabilei aleatoare în felul următor:

$$F(x) = \begin{cases} 0, & dacă \quad x \leq 0, \\ p_1, & dacă \quad x_1 \leq x < x_2, \\ p_1 + p_2, & dacă \quad x_2 \leq x < x_3, \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ 1, & dacă \quad x \geq x_n. \end{cases}$$

Proprietățile funcției de repartiție

1) Funcția $F(x)$ ia valori din intervalul [0,1], $\forall x \in \mathbb{R}$.
2) În cazul variabilei aleatoare discrete

$$P(\alpha \leq X \leq \beta) = F(\beta) - F(\alpha).$$

3) Dacă $x_1 < x_2 \Rightarrow F(x_1) < F(x_2)$
4) Limita lui $F(x)$, când $n \to \pm\infty$ este

$$\lim_{n \to -\infty} F(x) = 0 \quad și \quad \lim_{n \to \infty} F(x) = 1.$$

5) Funcția de repartiție este continuă la stânga.

În cazul variabilei aleatoare discrete, graficul funcției de repartiție este o funcție în trepte (fig. 11)

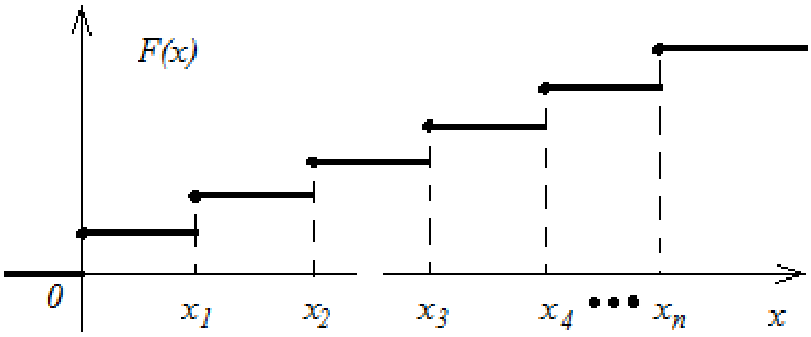


Fig. 11. Graficul de repartiție a variabilei aleatoare discrete

Dacă variabila aletoare este continuă și ia valori dintr-un interval deschis (*a*,*b*), atunci ea poate fi reprezentată grafic în felul următor (fig. 12):

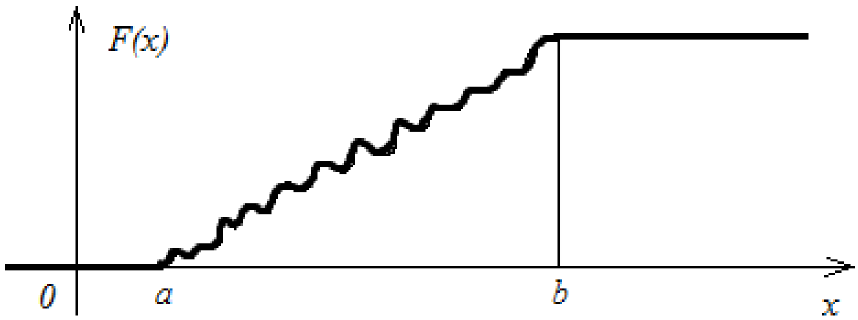


Fig. 12. Graficul de repartiție a variabilei aleatoare continue

O funcție importantă ce caracterizează variabila aleatoare este densitatea și se notează cu $f(x)$.

Definiția 2. *Fie X o variabilă aleatoare discretă ce ia valorile* $x_1, x_2, x_3, \dots, x_n$ *cu probabilitățile* $p_1, p_2, p_3, \dots, p_n$. *Funcția de densitate a probabilității se definește prin*

$$f(x) = \begin{cases} p_i, & dacă\ x = x_i (i = 1,2,3, \dots), \\ 0, & pentru\ celelalte\ cazuri. \end{cases} \tag{25}$$

În cazul variabilei continui, densitatea reprezintă probabilitatea apariției valorilor din intervalul $x + \triangle x$

$$\frac{P(x < X < x + \triangle x)}{\triangle x} = \frac{F(x + \triangle x) - F(x)}{\triangle x}.$$

Trecem la limită când $\triangle x \to 0$

$$\lim_{\triangle x \to 0} \frac{P(x < X < x + \triangle x)}{\triangle x} = \lim_{\triangle x \to 0} \frac{F(x + \triangle x) - F(x)}{\triangle x}.$$

În partea stângă, obținem densitatea

$$f(x) = F'(x).$$

Proprietățile densității

1) Densitatea unei variabile aleatoare ia numai valori pozitive, pentru $\forall x \in \mathbb{R}$.

2) Densitatea este egală cu derivata funcției de repartiție

$$f(x) = F'(x).$$

3) Probabilitatea că variabila aleatoare X, aparține unui oarecare segment $[a, b]$, este

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

4) Funcția de repartiție este egală cu unitatea

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

Exemplul 1. Fie X o variabilă aleatoare discretă care reprezintă numărul de puncte de pe fața obținută, la aruncarea unui zar (tabelul 4).

Tabelul 4. Repartiția variabilei aleatoare la aruncarea unui zar

X	*1*	*2*	*3*	*4*	*5*	*6*
p	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Să se determine funcția de repartiție $F(x)$ *și densitatea acestei funcții* $f(x)$. *Să se ilustreze grafic funcțiile* $f(x)$ *și* $F(x)$.

Rezolvare. În baza tabelului de valori a variabilei aleatoare X, obținem funcția de repartiție $F(x)$

$$F(x) = \begin{cases} 0, & dacă \quad x < 1, \\ \frac{1}{6}, & dacă \quad 1 \le x < 2, \\ \frac{1}{3}, & dacă \quad 2 \le x < 3, \\ \frac{1}{2}, & dacă \quad 3 \le x < 4, \\ \frac{2}{3}, & dacă \quad 4 \le x < 5, \\ \frac{5}{6}, & dacă \quad 5 \le x < 6, \\ 1, & dacă \quad x \ge 6. \end{cases}$$

Cu ajutorul funcției de repartiție $F(x)$, construim graficul acestei funcții (fig. 13)

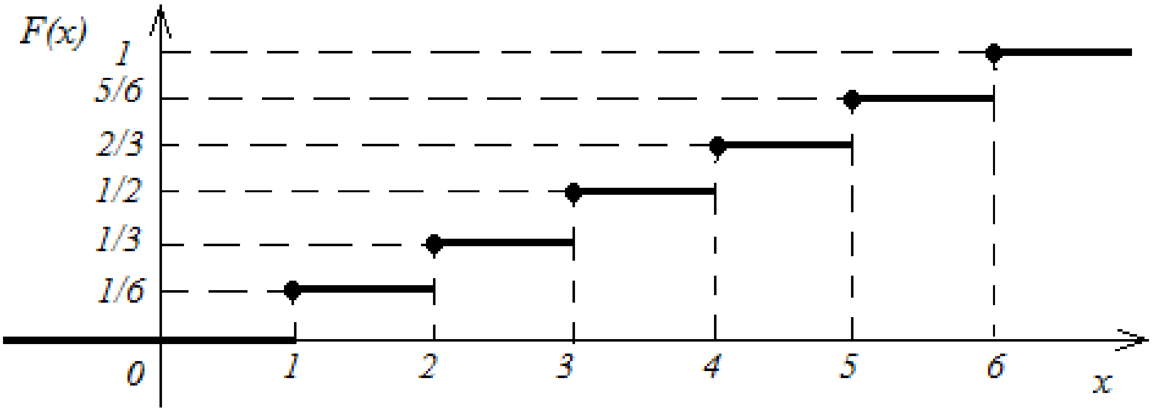


Fig.13. Funcția de repartiție $F(x)$

În baza formulei (25), obținem funcția de densitate a probabilității

$$f(x) = \begin{cases} \frac{1}{6}, & x \in \{1,2,3,4,5,6\}, \\ 0, & pentru\ celelalte\ cazuri. \end{cases}$$

Reprezentăm geometric funcția de densitate $f(x)$ (fig. 14)

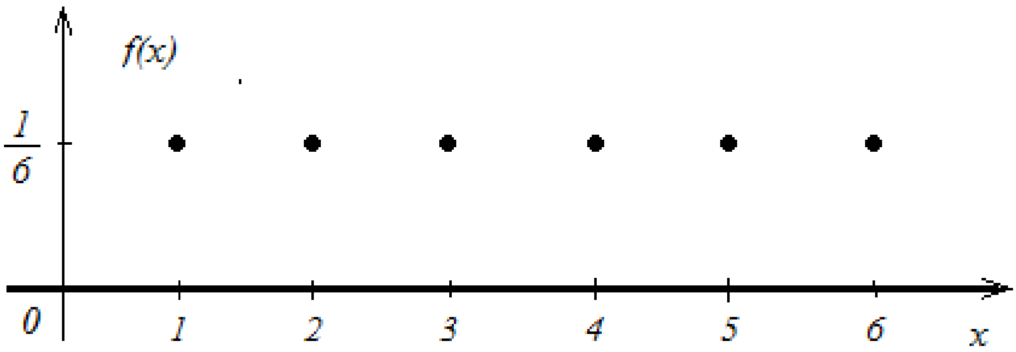


Fig.14. Funcția de densitate $f(x)$

Exerciții și probleme propuse

1. *Se aruncă două zaruri și se notează cu X numărul de puncte ce apar. Să se determine funcția de repartiție și să se ilustreze grafic această funcție.*

2. *Într-o urnă sunt 10 bile de două culori (albe și negre), dintre care 2 negre. Se i-au la întâmplare două bile. Să se determine legea de repartiție a numărului de bile albe dintre cele extrase.*

3. *Fie X o variabilei aleatoare discretă cu repartiția dată de tabelul 5*

Tabelul 5. Repartiția variabilei aleatoare

X	1	2	3	4
p	6/16	5/16	3/16	2/16

Să se determine funcția de repartiție, funcția de densitate a probabilității și să se ilustreze grafic aceaste funcții.

4. *Se aruncă o monedă de patru ori. Să se determine funcția de repartiție și densitatea obținerii stemei. Să se reprezinte geometric funcțiile obținute.*

5. *Înaintea punerii pe piaţă, un produs finit (utilaj) este supus la trei probe succesive de funcţionare. Probabilitatea ca să treacă de oricare din cele trei probe este de 0,8. Dacă se presupune că cele trei încercări sunt independente una de alta, să se determine:*

a) Repartiţia variabilei aleatoare care reprezintă numărul de probe trecute până la prima nereuşită;

b) Să se determine funcţia de repartiţie şi să se construiască graficul funcţiei de repartiţie al variabilei aleatoare de la punctul a).

6. *Fie X o variabilă aleatoare cu densitatea*

$$f(x) = \begin{cases} 2x, & pentru\ \ x \in (0,1), \\ 0, & pentru\ celelalte\ cazuri. \end{cases}$$

Să se determine următoarele probabilități:

a) $P\left(\frac{1}{4} < X < \frac{1}{2}\right)$;

b) $P\left(X > \frac{3}{4} / X > \frac{1}{2}\right)$.

7. *La trei magazine se găsesc articole ce provin de la două uzine: primul magazin se aprovizionează cu* $\frac{1}{3}$ *din articolele de la prima uzină, magazinul al doilea se aprovizionează cu* $\frac{1}{2}$ *din articolele de la prima uzină, iar magazinul al treilea se aprovizionează cu* $\frac{3}{4}$ *din articolele de la prima uzină. O persoană cumpără câte un articol de la fiecare magazin. Fie X numărul articolelor cumpărate de persoana respectivă. Să se determine repartiția variabilei aleatoare X, funcția de repartiție, iar apoi să se reprezinte grafic această funcție.*

§2.4. Valoarea medie şi dispersia variabilei aleatoare

Repartiția variabilei aleatoare, în general, ne descrie comportarea acestei variabile. În cazul când nu este necesară o caracterizare completă a variabilei aleatoare, ne folosim de anumite *caracteristici numerice*, printre care un loc important îl ocupă *valoarea medie, dispersia, mediana, moda* și *momentele de diferite ordine*.

Fie X o variabilă aleatoare cu repartiția

$$X\colon \begin{pmatrix} x_1 & x_2 & x_3 & \dots & x_n \\ p_1 & p_2 & p_3 & \dots & p_n \end{pmatrix}.$$

Se realizează k experimente independente, atunci, este evident, evenimentele x_i $(i = 1, \dots, n)$ apar de mai multe ori.

Fie că x_i apare de k_i $(i = 1, \dots, n)$ ori

x_1 – apare de k_1 ori;

x_2 – apare de k_2 ori;

…

x_n – apare de k_n ori.

Suma tuturor valorilor k_i $(i = 1, \dots, n)$ este

$$k_1 + k_2 + \dots + k_n = \sum_{i=1}^{n} k_i = k.$$

Determinăm *media aritmetică*

$$\bar{x} = \frac{k_1 x_1 + k_2 x_2 + \dots + k_n x_n}{k}$$

sau

$$\bar{x} = \frac{k_1}{k} x_1 + \frac{k_2}{k} x_2 + \dots + \frac{k_n}{k} x_n = \sum_{i=1}^{n} \frac{k_i}{k} x_i.$$

Evident că $\frac{k_i}{k}$ $(i = 1, \dots, n)$ sunt *frecvențele relative* ale evenimentelor x_i $(i = 1, \dots, n)$, deci

$$\bar{x} = f_1 x_1 + f_2 x_2 + \dots + f_n x_n = \sum_{i=1}^{n} f_i x_i.$$

Dacă numărul de experimente este mare, atunci frecvențele relative ale evenimentelor x_i $(i = 1, \dots, n)$ sunt aproximativ egale cu probabilitäile p_i $(i = 1, \dots, n)$.

Definiția 1 *[1]. Valoarea medie a unei variabile aleatoare discrete X, se numește numărul care este egal cu suma produselor dintre valorile ei posibile și probabilitățile corespunzătoare, notată $M(X)$*

$$M(X) = p_1x_1 + p_2x_2 + \cdots + p_nx_n = \sum_{i=1}^{n} p_ix_i. \qquad (26)$$

Proprietăţile valorii medii

1) Valoarea medie de la o constantă este însăşi constanta

$$M(c\ ) = c.$$

2) Dacă c este o constantă reală şi X o variabilă aleatoare, atunci valoarea medie de la suma $c + X$ este

$$M(c + X) = c + M(X).$$

La fel şi valoarea medie a produsului unei constante reale c şi o variabilă aleatoare X este

$$M(c \cdot X) = c \cdot M(X).$$

3) Dacă X şi Y sunt două variabile aleatoare, atunci valoarea medie de la suma/diferenţa lor este

$$M(X \pm Y) = M(X) \pm M(Y).$$

4) Dacă X şi Y sunt două variabile aleatoare independente, atunci valoarea medie de la produsul acestora este egal cu produsul valorilor medii, este

$$M(X \cdot Y) = M(X) \cdot M(Y).$$

5) Dacă pentru oricare două variabile aleatoare avem $X \leq Y$, atunci

$$M(X) \leq M(Y).$$

6) Dacă X şi Y sunt două variabile aleatoare independente, iar a și b sunt două constante reale arbitrare, atunci valoarea medie de la suma produselor constantelor α și respectiv β cu variabilele aleatoare X şi Yeste

$$M(\alpha \cdot X + \beta \cdot Y) = \alpha \cdot M(X) + \beta \cdot M(Y).$$

Definiția 2. *Fie variabila aleatoare X cu funcția de repartiție $F(x)$, atunci variabila aleatoare X se numește continuă, dacă funcția de repartiție este*

$$F(x) = \int_{-\infty}^{\infty} f(x)dx, \qquad (27)$$

unde $f(x)$ este densitatea variabilei aleatoare X.

Definiția 3 *[1]. Dacă X este o variabilă aleatoare continuă, având densitatea de repartiție $f(x)$, atunci numărul*

$$M(X) = \int_{-\infty}^{\infty} xf(x)dx \qquad (28)$$

se numește valoarea medie a variabilei aleatoare X, dacă integrala din membrul drept este absolut convergentă.

Dacă variabila aleatoare continuă X, ia valori numai din intervalul (a, b), atunci $f(x) = 0$ pentru $x \notin (a, b)$ și deci

$$M(X) = \int_a^b xf(x)dx.$$

Exemplul 1. Se aruncă două zaruri simetrice. Să se determine valoarea medie:

a) a sumei punctelor obținute;

b) a produsului punctelor obținute.

Rezolvare. Cunoaştem că în cazul zarurilor, probabilitatea căderii fețelor este echiprobabilă și fie X și Y numerele de puncte obținute în rezultatul experimentului considerat.

a) În caz general, pentru variabilele X şi Y avem repartiția

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \end{pmatrix}.$$

Determinăm valorile medii

$$M(X) = M(Y) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = \frac{7}{2}.$$

$$M(X+Y) = M(X) + M(Y) = \frac{7}{2} + \frac{7}{2} = 7.$$

b) Evindent, variabilele X şi Y sunt independente, deci

$$M(X \cdot Y) = M(X) \cdot M(Y) = \left(\frac{7}{2}\right)^2 = \frac{49}{4} = 12\frac{1}{4}.$$

Răspuns: $M(X+Y) = 7; M(X \cdot Y) = 12\frac{1}{4}$

***Definiția 4** [1]. Mediana unei variabile aleatoare X este valoarea Me a argumentului x, pentru care*

$$F(Me) \leq \frac{1}{2}$$

și

$$F(Me + 0) \geq \frac{1}{2}.$$

Geometric mediana este valoarea variabilei aleatoare cu proprietatea că dreapta $x = Me$ împarte în jumătate aria figurii de sub curba denstății de repartiție.

***Definiția 5.** Valoarea modală a unei variabile aleatoare X, se numește valoarea sa cea mai probabilă (pentru X discret) sau valoarea, care este abscisa punctului maxim al curbei densității de repartiție (pentru variabila aleatoare X continuă), dacă această curbă are un singur vârf determinat.*

Dispersia caracterizează gradul de împrăştire a valorilor variabilei aleatoare faţă de valoarea medie.

Definiția 6. *Se numeşte dispersia variabilei aleatoare X, valoarea medie a variabilei aleatoare* $(X - M(X))^2$.

Dispersia se notează prin *D*

$$D(X) = M\big(X - M(X)\big)^2 =$$
$$= p_1(x_1 - M(X))^2 + p_2(x_2 - M(X))^2 + \cdots + p_n\big(x_n - M(X)\big)^2. \quad (29)$$

Definiția 7. *Variabila aleatoare $X - M(X)$ se numeşte abaterea pătratică faţă de valoarea medie a variabilei aleatoare X*

$$X - M(X) = \begin{pmatrix} x_1 - M(X) & x_2 - M(X) & \dots & x_n - M(X) \\ p_1 & p_2 & \dots & p_n \end{pmatrix}.$$

În baza proprietăților valorii medii, obținem

$$D(X) = M(X^2) - \big(M(X)\big)^2$$

unde

$$M(X^2) = x_1^2 p_1 + x_2^2 p_2 + \cdots + x_n^2 p_n$$

și

$$M(X)^2 = (x_1 p_1 + x_2 p_2 + \cdots + x_n p_n)^2.$$

Această proprietate ușor se arată cu ajutorul proprietăților de mai sus

$$D(X) = M\big(X - M(X)\big)^2 = M\left(X^2 - 2fM(X) + \big(M(X)\big)^2\right) =$$
$$= M(X^2) - 2M(X)M(X) + \big(M(X)\big)^2 = M(X^2) - \big(M(X)\big)^2.$$

Exemplul 2. Fie variabila aleatoare X cu repartiția

$$X\colon \begin{pmatrix} 2 & 3 & 2 \\ 0{,}2 & 0{,}3 & 0{,}5 \end{pmatrix}.$$

Să se determine dispersia variabilei aleatoare X.

Rezolvare. Determinăm valoarea medie a variabilei aleatoare *X*

$$M(X) = 2 \cdot 0{,}2 + 3 \cdot 0{,}3 + 2 \cdot 0{,}5 = 2{,}3.$$

Repartiția variabilei X^2 este

$$X^2\colon \begin{pmatrix} 4 & 9 & 4 \\ 0{,}2 & 0{,}3 & 0{,}5 \end{pmatrix},$$

Deci

$$M(X^2) = 4 \cdot 0{,}2 + 9 \cdot 0{,}3 + 4 \cdot 0{,}5 = 5{,}5.$$

Înlocuim valorile determinate, în formula dispersiei (29), obținem

$$D(X) = 5{,}5 - (2{,}3)^2 = 0{,}21.$$

Răspuns: $D(X) = 0{,}21$

Proprietăţile dispersiei

1) Dispersia de la o constantă este zero

$$D(c) = M(c^2) - \big(M(c)\big)^2 = c^2 - c^2 = 0.$$

2) Dispersia de la produsul unei constantă şi a unei variabile aleatoare

$$D(c \cdot X) = M(c^2 \cdot X^2) - \big(M(c \cdot X)\big)^2 = c^2 \cdot M(X^2) - c^2 \cdot \big(M(X)\big)^2 =$$
$$= c^2(M(X^2) - \big(M(X)\big)^2) = c^2\, D(X).$$

3) Dispersia de la suma unei constante c şi a unei variabile aleatoare X este

$$D(c + X) = 0 + D(X) = D(X).$$

4) Dispersia de la suma/diferenţa a 2 variabile aleatoare X şi Y este

$$D(X \pm Y) = D(X) \pm D(Y).$$

5) Pentru orice numere reale α și β avem

$$D(\alpha X + \beta) = \alpha^2 D(X).$$

6) În cazul când variabilele $x_1, x_2, \dots, x_n$ sunt independente avem

$$D(x_1 + x_2 + \dots + x_n) = D(x_1) + D(x_2) + \dots + D(x_n).$$

Pentru variabilele aleatoare continue, exprimăm dispersia prin valoarea sa medie (28), adică

$$D(X) = M\big(X - M(X)\big)^2 = \int_{-\infty}^{\infty} \big(X - M(X)\big)^2 f(x)dx =$$
$$= \int_{-\infty}^{\infty} (X^2 - 2XM(X) + \big(M(X)\big)^2)\, f(x)dx =$$
$$= \int_{-\infty}^{\infty} X^2\, f(x)dx - 2M(X) \int_{-\infty}^{\infty} X\, f(x)dx + \big(M(X)\big)^2 \int_{-\infty}^{\infty} f(x)dx =$$
$$= M(X^2) - (M(X))^2,$$

deci

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

și

$$\sigma(X) = \sqrt{D(X)}. \tag{30}$$

Observația 1. *Valoarea medie și dispersia, în cazul variabilei aleatoare continue, au aceleași proprietăți ca și în cazul variabilei aleatoare discrete.*

Exemplul 3. Se aruncă o monedă de 3 ori. Fie X variabila aleatoare care reprezintă numerele de apariții a stemei. Să se determine tabloul de repartiție a variabilei aleatoare X, valoarea medie $M(X)$ și dispersia $D(X)$.

Rezolvare. În acest caz, spațiul evenimentelor elementare este format din $2^3 = 8$ elemente

$$\{bbb, bbs, bsb, bss, sbb, sbs, ssb, sss\}.$$

De aici rezultă că într-un singur caz obținem trei steme

$$P_1 = \frac{1}{8},$$

în trei cazuri câte o stemă

$$P_2 = \frac{3}{8},$$

în trei cazuri câte două steme

$$P_3 = \frac{3}{8},$$

și într-un singur caz nici o stemă

$$P_4 = \frac{1}{8}.$$

În baza celor obținute, compunem tabelul de valori (tabelul 6)

Tabelul 6. Tabelul repartiției variabilei aleatoare X

X	0	1	2	3
p	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

Calculăm valoarea medie $M(X)$

$$M(X) = 0 \cdot \frac{1}{8} + 1 \cdot \frac{3}{8} + 2 \cdot \frac{3}{8} + 3 \cdot \frac{1}{8} = \frac{3}{2}.$$

Determinăm $(X - M(X))^2$ pentru fiecare valoare a lui X

Dacă $X = 0$, atunci $(X - M(X))^2 = \left(0 - \frac{3}{2}\right)^2 = \frac{9}{4}$.

Dacă $X = 1$, atunci $(X - M(X))^2 = \left(1 - \frac{3}{2}\right)^2 = \frac{1}{4}$.

Dacă $X = 2$, atunci $(X - M(X))^2 = \left(2 - \frac{3}{2}\right)^2 = \frac{1}{4}$.

Dacă $X = 3$, atunci $(X - M(X))^2 = \left(3 - \frac{3}{2}\right)^2 = \frac{9}{4}$.

Determinăm dispersia (formula (29)), obținem

$$D(X) = M(X - M(X))^2 = \frac{9}{4} \cdot \frac{1}{8} + \frac{1}{4} \cdot \frac{3}{8} + \frac{1}{4} \cdot \frac{3}{8} + \frac{9}{4} \cdot \frac{1}{8} = \frac{3}{4}.$$

Răspuns: $M(X) = \frac{3}{2}; D(X) = \frac{3}{4}$

Observația 1. *Dacă valorile medii, ale n valori aleatoare cu aceleaşi caracteristici, sunt egale*

$$M(x_1) = M(x_2) = \cdots = M(x_n) = m$$

şi de asemenea dispersiile sunt egale

$$D(x_1) = D(x_2) = \cdots = D(x_n) = d,$$

atunci valoarea medie $M(X) = m$, *iar* $D(X) = \frac{d}{n}$.

Dacă $x_1, x_2, \dots, x_n$ sunt independente, atunci

$$\sigma(x_1 + x_2 + \cdots + x_n) = \sqrt{D(x_1) + D(x_2) + \cdots + D(x_n)} =$$
$$= \sqrt{\sigma_1^2(x_1) + \sigma_2^2(x_2) + \cdots + \sigma_n^2(x_n)}.$$

Definiția 8. *Mărimea $\sigma(X)$ este rădăcină de ordinul doi din dispersie $\sigma(X) = \sqrt{D(X)}$ și se numeşte abaterea pătratică (standartă).*

Exerciții și probleme propuse

1. *Să se determine media și dispersia pentru următoarea serie de valori:* 2,2,3,3,3,4.

2. *Să se determine valoarea medie și dispersia variabilei aleatoare care reprezintă numărul de puncte ce apar la aruncarea unui zar.*

3. *Fie două variabile aleatoare independente X și Y cu repartițiile*
$$X\colon\begin{pmatrix} 1 & 2 & 4 \\ 0{,}7 & 0{,}1 & 0{,}2 \end{pmatrix} \quad și \quad Y\colon\begin{pmatrix} 1 & 3 & 5 & 7 \\ 0{,}1 & 0{,}2 & 0{,}3 & 0{,}4 \end{pmatrix}.$$
Să se determine $M(2X + 4Y)$ și $D(2X + 4Y)$.

4. *Fie o monedă care se aruncă de două ori, iar numărul de apariții a banului reprezintă o variabilă aleatoare X. Să se determine valoarea medie și dispersia variabilei aleatoare X.*

5. *Fie două variabile aleatoare cu repartițiile*
$$X\colon\begin{pmatrix} 3 & 5 & 2 \\ 0{,}1 & 0{,}4 & 0{,}5 \end{pmatrix} \quad și \quad Y\colon\begin{pmatrix} 0 & 1 & 2 & 3 & 4 \\ \frac{1}{81} & \frac{8}{81} & \frac{24}{81} & \frac{32}{81} & \frac{16}{81} \end{pmatrix}.$$
Să se determine media și dispersia variabilelor X și Y.

6. *Într-o ladă sunt 8 piese dintre care 6 standarte și 2 nestandarte. Se i-au la întâmplare 2 piese. Fie X o variabilă aleatoare care reprezintă numărul de piese standarte. Să se determine repartiția lui X, valoarea medie $M(X)$ și dispersia $D(X)$.*

7. *Să se determine valoarea medie și dispersia variabilei aleatoare X, care reprezintă numărul de apariții a evenimentului A, din 100 de experimente independente, dacă probabilitatea apariției lui A este 0,85 la toate experimentele.*

8. *Să se determine variabilele aleatoare independente*
$$X\colon\begin{pmatrix} x & x+1 & x+2 & x+3 \\ p & 2p & 3p & 4p \end{pmatrix} \quad și \quad Y\colon\begin{pmatrix} y & 2y & 3y \\ q & q^2 & q^2 \end{pmatrix},$$
dacă $M(X) = 2$ și $M(X) = 7$. Să se determine $M(2X + 3Y)$, $D(X)$, $D(Y)$ și $D(2X + 3X)$.

9. *Se aruncă o monedă de patru ori. Să se determine valoarea medie și dispersia variabilei aleatoare X, care reprezintă numărul de apariții a stemei.*

§2.5. Momente. Coeficientul de asimetrie şi boltire

În teoria probabilităților, o caracteristică a variabilei aleatoare, ce generalizează noțiunile de valoarea medie și dispersie, sunt *momentele*.

Momentele sunt de două tipuri [1]:

- *momente inițiale*;
- *momente centrate*.

Fie X o variabilă aleatoare cu repartiţia

$$X\colon \begin{pmatrix} x_1 & x_2 & x_3 & \dots & x_n \\ p_1 & p_2 & p_3 & \dots & p_n \end{pmatrix}.$$

Definiția 1. *Momentul iniţial de ordinul k, al variabilei aleatoare X, se numeşte valoare medie a variabilei aleatoare X^k şi se notează cu*

$$m_k = M(X^k). \tag{31}$$

Dacă variabila aleatoare X este discretă, atunci

$$m_k = \sum_i x_i^k p_i,$$

iar dacă variabila aleatoare X este continuă, cu densitatea de repartiție $f(x)$, atunci

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

Definiția 2 [2]. *Momentul absolut de ordinul k, al variabilei aleatoare X, se numeşte valoare medie a variabilei aleatoare $|X^k|$, unde*

$$m_k' = M(|X^k|). \tag{32}$$

În particular, utilizând momentele

$$m_1 = M(X) = m \ \text{ și } \ m_2 = M(X^2)$$

și aplicându-le în calculul dispersiei, obținem

$$D(X) = m_2 - m_1^2. \tag{33}$$

Exemplul 1. Să se determine momentul inițial de ordinul k și dispersia variabilei aleatoare X cu repartiția

$$X\colon \begin{pmatrix} 1 & 0 \\ p & q \end{pmatrix}.$$

Rezolvare. În baza formulei (31), determinăm momentul inițial m_k

$$m_k = 1^k \cdot p + 0^k \cdot q = p,$$

iar în baza formulei (33), obținem

$$D(X) = m_2 - m_1^2 = p - p^2 = p(1-p) = pq.$$

Răspuns: $m_k = p; D(X) = pq$

Definiția 3. *Momentul centrat de ordinul k, al variabilei aleatoare X, se numeşte valoare medie a variabilei aleatoare $(X - M(X))^k$ şi se notează cu*

$$\mu_k = M[X - M(X)]^k. \tag{34}$$

Dacă repartiţia variabilei aleatoare este simetrică faţă de valoarea medie, atunci momentele centrate de ordinul k, când k este impar, sunt nule, adică

$$\mu_1 = \mu_2 = \mu_3 = \cdots = 0.$$

Definiția 4. *Momentul centrat absolut de ordinul k, al variabilei aleatoare X, se numeşte numărul*

$$M(|X - M(X)|^k). \tag{35}$$

De asemenea, o importanță deosebită, în studiul teoriei probabilităților, îl ocupă următoarele caracteristici numerice: *coeficientul de asimetri* și *excesul.*

Definiția 5 [1]. *Raportul dintre momentul centrat de ordinul 3 şi cubul abaterii medii centrate (dacă momentele respective există), se numeşte coeficient de asimetrie şi se notează*

$$A_S = \frac{\mu_3}{\sigma^3}. \tag{36}$$

Observația 1. *Dacă variabila aleatoare este simetrică faţă de valoarea medie, atunci coeficientul de simetrie este zero.*

Definiția 6. *Numărul*

$$E_f = \frac{\mu_4}{\sigma^4} - 3, \tag{37}$$

se numeşte exces al variabilei aleatoare X (dacă μ_4 și σ^4 există), iar raportul $\frac{\mu_4}{\sigma^4}$ se numește coeficient de boltire.

Exerciții și probleme propuse

1. *Fie X o varabilă aleatoare cu repartiția dată în tabelul 7.*

Tabelul 7. Tabelul repartiției variabilei aleatoare X

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

Să se determine momentele inițiale și centrate de ordinul doi și trei, coeficientul de asimetrie și excesul.

2. *Să se determine momentul inițial de ordinul k al variabilei aleatoare X cu repartiția*

$$X\colon \begin{pmatrix} a & b \\ p & q \end{pmatrix}.$$

3. *Să se determine momentele m_1 și m_2 pentru variabila aleatoare X cu repartiția*

$$X\colon \begin{pmatrix} 0 & 1 & 2 \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \end{pmatrix}.$$

4. *Fie X o variabilă aleatoare care are densitatea de probabilitate*

$$f(x)=\begin{cases}0, & dacă\ x\notin(0,2),\\ \frac{1}{2}, & dacă\ x\in(0,2).\end{cases}$$

Să se determine momentul inițial de ordinul k.

5. *Fie X o variabilă aleatoare care reprezintă numărul de angajați ai unei firme divizați pe secții (tabelul 8). Să se determine asimetria și boltirea datelor din tabel.*

Tabelul 8. Repartiția variabilei aleatoare X

Secția	*S1*	*S2*	*S3*	*S4*	*S5*	*S6*	*S7*	*S8*	*S9*
Nr. de angajați	4	5	10	10	16	18	18	14	5
p	0,04	0,05	0,1	0,1	0,16	0,18	0,18	0,14	0,05

§2.6. Legi de repartiție discrete clasice

În literatura de specialitate [1-17], întâlnim următoarele legi de repartiție discrete clasice:

1) Repartiția binomială (Bernoulli)

Repartiţia binomială corespunde următorului tip de experiment: Fie A un eveniment care se produce cu probabilitatea p și respectiv evenimentul contrar $\bar{A}$ care se produce cu probabilitatea q. Cele două evenimente formează un sistem de evenimente, producerea unuia excluzând producerea celuilalt. Se repetă experimentul de n ori.

În cele n experimente, evenimentul A s-ar putea să nu se producă nici o dată, s-ar putea să se producă o dată, sau s-ar putea produce de n ori. Ne interesează să determinăm de fiecare dată probabilitatea de realizare a evenimentului A. În acest caz, scrim un tablou de repartiţie de forma

$$X\colon\begin{pmatrix}k\\ P(n,k)\end{pmatrix},$$

unde $k=\overline{0,n}$, iar $P(n,k)$ se calculează conform schemei binomiale (11)

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

Observaţia 1. Repartiţia binomială, în general, este determinată de schema urnei cu 2 bile (albe/negre) şi bila revenită.

Teorema 1. *Dacă variabila aleatoare X, urmează legea de repartiţie binomială, atunci valoarea medie este numărul*

$$M(X)=np, \tag{38}$$

iar dispersia

$$D(X)=npq. \tag{39}$$

Demonstraţie. Media variabilei aleatore X, care ne dă k bile albe din n bile extrase, va fi, prin definiţie

$$M(X) = \sum kC_n^k p^k q^{n-k}.$$

Pentru a calcula această sumă, considerăm următoarea identitate:

$$(pt + q)^n = \sum C_n^k p^k t^k q^{n-k},$$

derivăm ultima egalitate în raport cu t

$$((pt + q)^n)' = \sum C_n^k p^k t^k q^{n-k},$$

$$np(pt + q)^{n-1} = \sum C_n^k p^k k t^k q^{n-k},$$

pentru

$$t = 1 \Longrightarrow \sum C_n^k p^k k q^{n-k}.$$

În rezultat, am obţinut

$$M(X) = np.$$

Folosind aceiaşi identitate, dar derivând de două ori, ușor se arată că dispersia este

$$D(X) = npq.$$

Cunoaşterea mediei şi dispersiei unei variabile aleatoare dă o indicaţie asupra intervalului în care se află valorile variabilei, cu cea mai mare probabilitate. Mai exact, cu cât ne îndepărtam mai mult de valoarea medie, cu atât valorile respective sunt mai puţin probabile ca valori ale variabilei date.

Exemplul 1. Să se determine media și dispersia variabilei aleatoare X, care reprezintă numărul de nimeriri în țintă, din zece trageri, dacă se cunoaște probabilitatea nimeririi la fiecare tragere $p = 0{,}6$.

Rezolvare. Variabila aleatoare X are repartiție binomială, unde

$$n = 10 \quad și \quad p = 0{,}6.$$

Media variabilei aleatoare X, formula (38), este

$$M(X) = 10 \cdot 0{,}6 = 6.$$

Dispersia variabilei aleatoare X, formula (39), este

$$D(X) = npq = np(1 - p) = 10 \cdot 0{,}6 \cdot (1 - 0{,}6) = 2{,}4$$

Răspuns: $M(X) = 6;\ D(X) = 2{,}4$

2) Schema hipergeometrică

Modelul matematic al acestei repartiţii este similar celui binomial, diferenţa constând în faptul că elementul extras pentru control nu se mai întoarce în lot, şi-n consecinţă, la fiecare extragere se modifică condiţiile şi deci şi probabilitatea de extragere a unei piese rebut. Din acest motiv, extragerea se mai numeşte *fără întoarcere***.**

Se consideră un lot de piese, pentru care trebuie verificat coeficientul de rebut p. Cunoscând mărimea lotului de piese n, se poate determina numărul de piese rebut a și respectiv numărul de piese bune b

$$a = npb = n(1-p).$$

Se efectuează m extrageri consecutive, fără a pune piesa extrasă la loc, în cele m extrageri consecutive pot fi obținute 0 piese rebut, 1 piesă rebut, …, sau m piese rebut. În consecinţă putem construi un tablou de repartiţie în care pe prima linie să trecem numărul de piese rebut şi-n a doua linie probabilitatea extragerii fiecăreia. De exemplu

$$X: \begin{pmatrix} 0 & 1 & \dots & n \\ p_0 & p_1 & \dots & p_n \end{pmatrix}.$$

În acest caz, spunem că variabila aleatoare X urmează legea hipergeometrică cu repartiţia

$$X: \begin{pmatrix} k \\ P(n,k) \end{pmatrix},$$

unde $k = \overline{0,n}$, iar $P(n,k)$ se calculează conform formulei (13)

$$P(n,k) = \frac{C_a^k C_b^{n-k}}{C_{a+b}^n}$$

(a – numărul de piese rebut și b – numărul de piese bune).

Teorema 2. *Dacă X este o variabilă aleatoare discretă şi urmează legea hipergeometrică, atunci valoarea medie este*

$$M(X) = np, \tag{40}$$

iar dispersia la pătrat este

$$D^2(X) = npq\frac{N-n}{N-1}, \tag{41}$$

unde $N = a + b$ este numărul total de piese, n este numărul de piese care se extrag, p este probabilitatea că piesele extrase sunt rebut, $p = \frac{a}{a+b}$ și q este probabilitatea că piesele extrase sunt bune $q = \frac{b}{a+b}$.

Exemplul 2. Să se determine media și dispersia variabilei aleatoare X, care reprezintă numărul de de piese rebut, dintr-un lot de 100 de piese, la extragerea a 20 de piese fără întoarcere, dacă se știe că probabilitatea extragerii unei piese rebut este $p = 0{,}02$.

Rezolvare. Variabila aleatoare X are repartiție hipergeometrică, unde

$$n = 20 \quad și \quad p = 0{,}02.$$

Media variabilei aleatoare X, formula (40), este

$$M(X) = 20 \cdot 0{,}02 = 0{,}4.$$

Probabilitatea insuccesului este

$$q = 1 - p = 1 - 0{,}02 = 0{,}98.$$

Dispersia variabilei aleatoare X, formula (41), este

$$D(X) = \sqrt{npq\frac{N-n}{N-1}} = \sqrt{20 \cdot 0{,}02 \cdot 0{,}98 \cdot \frac{100-20}{100-1}} \approx 0{,}56.$$

Răspuns: $M(X) = 0{,}4;\ D(X) \approx 0{,}56$

3) Repartiţia lui Poisson

Notăm cu λ, densitatea de apariţie a unui eveniment A, într-o unitate de timp t, atunci

$$\mu = \lambda t$$

reprezintă media apariţiilor în intervalul cercetat. Posibilitatea apariţiei de k ori a evenimentului A, în acelaşi interval, este

$$P(X = k) = P(k) = \frac{(\lambda t)^k}{k!}e^{-\lambda t} = \frac{\mu^k}{k!}e^{-\mu}.$$

Repartiţia Poisson se aplică în cazul repartiţiei binomiale dacă coeficientul de rebut este foarte mic $p < 0{,}1$ şi mărimea lotului mare $np > 5$. Relaţia care se aplică este

$$P(X = k) = P(k) = \frac{(np)^k}{k}e^{-np}.$$

Teorema 3. *Dacă variabila aleatoare urmează legea lui Piosson, atunci valoarea medie este*

$$M(X) = D^2(X) = \lambda t. \tag{42}$$

Demonstrație.

$$M(X) = \sum_{k=0}^{\infty} k\frac{(\lambda t)^k}{k!}e^{-\lambda t} = e^{-\lambda t}\sum_{k=0}^{\infty} k\frac{(\lambda t)^k}{k!} =$$

$$= \lambda t e^{-\lambda t}\sum_{k=0}^{\infty}\frac{(\lambda t)^{k-1}}{(k-1)!} = \lambda t e^{-\lambda t}e^{\lambda t} = \lambda t.$$

Observația 2. *Legea de repartiție a lui Poisson este cunoscută și sub denumirea de repartiţia evenimentelor rare. Această repartiţie se aplică în cazul avariilor auto sau a accidentelor acestora.*

4) Repartiția geometrică

Variabila aleatoare X urmează legea geometrică, dacă este definită de repartiţia

$$X: \begin{pmatrix} k \\ (1-p)p^k \end{pmatrix}.$$

Teorema 4. *Dacă variabila aleatoare X urmează legea geometrică, atunci valoarea medie este*

$$M(X) = \frac{p}{1-p}, \tag{43}$$

iar dispersia la pătrat

$$D^2(X) = \frac{p}{(1-p)^2}. \tag{44}$$

Exerciții și probleme propuse

1. *Considerăm jocul Dats și X o variabilă aleatoare care reprezintă numărul de nimeriri în 100 din trei aruncări. Să se determine media și dispersia variabilei aleatoare X, dacă se cunoaște că probabilitatea nimeririi în 100 este de 40% la fiecare aruncare.*
2. *Fie X o variabilă aleatoare definită de repartiţie Poisson cu parametrul λ. Pentru această variabilă, să se determine momentul iniţial $m_3(X)$ și momentul central $\mu_3(X)$.*
3. *Să se determine valoarea medie și dispersia variabilei aleatoare X, care reprezintă numărul de apariții a evenimentului A în 100 de experimente independente, dacă se cunoaște probabilitatea apariției evenimentului A, la toate experimentele, $p = 75\%$.*
4. *Să se determine media și dispersia variabilei aleatoare X, care reprezintă numărul de bile negre, dintr-o urnă cu 50 de bile, la extragerea a 10 bile fără întoarcere, dacă se știe că probabilitatea extragerii unei bile negre este $p = 0{,}3$.*

§2.7. Legea numerelor mari

Înaintea efectuării unei experiențe nu putem cunoaște ce valoare va lua o variabilă aleatoare pe care o cercetăm. Întrucât dispunem de puține informații despre fiecare variabilă aleatoare, s-ar părea că determinarea comportării mediei aritmetice a unui număr suficient de mare, de variabile aleatoare, este o problemă dificilă. Însă în realitate, în condiții puțin restrictive, media aritmetică a unui număr mare de variabile aleatoare își pierde caracterul întâmplător, iar în practică, este foarte important să cunoaștem condițiile în care acțiunea combinată a mai multor factori întâmplători conduce la un rezultat care să nu depindă de întâmplare, deci care să ne permită să prevedem mersul fenomenului studiat. Aceste condiții se dau în calculul probabilităților, în teoreme cunoscute sub denumirea comună de *legi ale numerelor mari* [19].

Legea numărului mare a fost menționată pentru prima dată de matematicianul Gerolamo Cardamo, deși fără nici o dovadă riguroasă. Mai târziu, Jacob Bernoulli a reușit să facă o demonstrație completă în lucrarea sa „Ars Conjectandi”, în 1713. În anii 1830, matematicianul Siméon Denis Poisson, a descris în detaliu legea numărului mare, care a ajuns să perfecționeze teoria.

Gerolamo Cardamo (1501-1576)

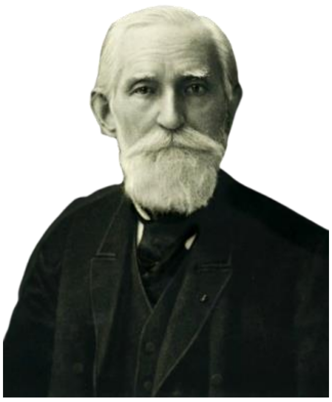
Însă, în 1867 Cebîșev a precizat riguros, din punct de vedere matematic, legea numerelor mari în condiții mai generale.

Teorema 1 (Inegalitatea lui Cebîșev). Fie $\varepsilon > 0$ și X o variabilă aleatoare pentru care există $M(X)$ și $D^2(X)$, atunci are loc egalitatea lui Cebîșev

$$P(|X - M(X)| < \varepsilon) \geq 1 - \frac{D(x)}{\varepsilon^2}, \tag{45}$$

sau

$$P(|X - M(X)| \geq \varepsilon) \leq \frac{D(x)}{\varepsilon^2}. \tag{46}$$



Pafnuti Cebîșev (1821-1894)

Definiția 1. *Vom spune că șirul de variabile aleatoare* (x_n) *converge în probabilitate către o variabilă aleatoare X dacă, fiind date două numere reale pozitive, suficient de mici,* ε *și* η, *există un întreg* N, *încât pentru orice* $n > N$ *avem*

$$P(|x_n - X| \geq \varepsilon) < \eta.$$

Teorema 2. *Dacă* (x_n) *este un șir de variabile aleatoare și* $a \in \mathbb{R}$, *astfel încât*

$$M(x_n) \to a \qquad \text{și} \qquad D^2(x_n) \to 0,$$

atunci $x_n \to a$.

Demonstrație. Se aplică inegalitatea lui Cebîșev

$$P(|x_n - x| \geq \varepsilon) < \frac{M((x_n - a)^2)}{\varepsilon^2} \to 0.$$

Fie șirul de variabile aleatoare (x_n), definit pe câmpul (Ω, K, P) și fie (φ_n) un șir de aplicații $\varphi_n: \mathbb{R}^n \to \mathbb{R}$, simetrice în raport cu argumentele lor. Fie șirul (y_n) dat de $y_n = \varphi(x_1, x_2, \dots, x_n)$.

Definiția 2. *Dacă există un șir de constante* $(c_n)_{n \in N}$, *astfel încât*

$$\lim_{n \to \infty} P(|y_n - c_n| < \varepsilon) = 1,$$

pentru orice $\varepsilon > 0$ *dat, atunci spunem că șirul* (x_n) *urmează legea slabă a numerelor mari.*

Cu alte cuvinte $|y_n - c_n| \to 0$. Una din cele mai frecvente alegeri pentru φ_n este

$$\varphi_n(x_1, x_2, \dots, x_n) = \frac{x_1 + x_2 + \dots + x_n}{n},$$

iar pentru c_n

$$c_n = \frac{M(x_1) + M(x_2) + \dots + M(x_n)}{n}.$$

Teorema 3 (Markov). *Dacă șirul* (x_n) *verifică condiția lui Markov*

$$\lim_{n \to \infty} \frac{1}{n^2} D^2 \left(\sum_{i=1}^{n} x_i \right) = 0,$$

atunci

$$\lim_{n \to \infty} P\left(\left| \frac{1}{n} \sum_{i=1}^{n} x_i - \frac{1}{n} \sum_{i=1}^{n} M(x_i) \right| < \varepsilon \right) = 1, \qquad (47)$$

pentru orice $\varepsilon > 0$ *dat.*

Andrei Marcov (1856-1922)

Teorema 4 (Cebîșev). *Dacă variabilele aleatoare* $x_1, x_2, \dots, x_n$ *sunt independente două câte două și au dispersii finit mărginite (dispersiile nu depășesc o animită constantă), atunci cât de mic nu ar fi* $\varepsilon > 0$, *avem probabilitatea mai mică decât* ε, *și va fi aproape egală de 1, dacă* n *este destul de mare când* $n \to \infty$, *adică*

$$\lim_{n\to\infty} P\left(\left|\frac{x_1 + x_2 + \dots + x_n}{n} - \frac{M(x_1) + M(x_2) + \dots + M(x_n)}{n}\right| < \varepsilon\right) = 1.$$

Observația 1. *Dacă valorile medii sunt egale, adică*

$$M(x_k) = m, k = \overline{1, n},$$

atunci teorema primește următoarea formă:

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

Această observație explică de ce putem să facem afirmații asupra unei populații pe baza unei selecții, având un volum mic comparativ cu cel al întregii populații. Explicația constă în aceea că selecția implică un număr de măsurători suficient.

Teorema 5 (Poisson). *Dacă într-un șir de repetări a unor experimente, probabilitatea apariției evenimentului A, la repetarea* i, *este* P_i, *atunci*

$$\lim_{n\to\infty} P\left(\left|\frac{m}{n} - \frac{1}{n}\sum_{k=1}^{n} P_k\right| < \varepsilon\right) = 1,$$

pentru $\forall\varepsilon > 0$, *unde* m *este numărul de apariții a evenimentului A, la primele* n *repetări.*

Teorema 6 (Bernoulli). *Dacă se fac* n *repetări independente și probabilitatea evenimentului A este constantă la fiecare repetare, atunci*

$$\lim_{n\to\infty} P\left(\left|\frac{m}{n} - P\right| < \varepsilon\right) = 1$$

unde $\varepsilon > 0$, n *reprezintă numărul total de experimente, iar* m *numărul de apariții a evenimentului A în acest experiment.*

Teoremă (Moivre-Laplace). *Dacă variabila aleatoare urmează legea binomială de repartiție, atunci*

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

pentru $a < b$ *avem*

$$P\left(a < \frac{X - np}{\sqrt{npq}} < b\right) \approx \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}}\, dx = \frac{1}{\sqrt{2\pi}} \int_{a}^{b} e^{-\frac{x^2}{2}}\, dx.$$

Utilizând funcția lui Laplace obținem

$$P\left(a < \frac{x-np}{\sqrt{npq}} < b\right) \cong \Phi(b) - \Phi(a), \tag{48}$$

unde $x = 1,2,\dots,n$ *și*

$$P(a < k < b) \cong \Phi\left(\frac{b-np}{\sqrt{npq}}\right) - \Phi\left(\frac{a-np}{\sqrt{npq}}\right). \tag{49}$$

Exerciții și probleme propuse

1. Fie valoarea media $M(X) = 360000kWh$ și abaterea medie pătratică $\sigma(X) = 40000\,kWh$ a consumului de energie electrică în raionul Căușeni, în luna ianuarie. Să se determine probabilitatea că în luna ianuarie anul curent, consumul va depăși 1000000 kWh.

2. Fie variabilele aleatoare $x_1, x_2, \dots, x_n$ care sunt independente două câte două și pot lua valorile $\pm\sqrt{k}$ și zero, cu următoarele probabilități:

$$P\{x_1 = 0\} = 1, \qquad P\{x_k = \sqrt{k}\} = P\{x_k = -\sqrt{k}\} = \frac{1}{k}$$

și

$$P\{x_k = 0\} = 1 - \frac{2}{k}, \quad k = 2,3,\dots$$

Să se demonstreze că șirul x_n se supune legii numerelor mari.

3. Se cunoaște că magazinul X, poate deservi zilnic un număr de 3000 de clienți. Probabilitatea că un client devine cumpărător este p=70%. Să se determine probabilitatea că numărul cumpărătorilor să fie mai mic decât 2150.

4. Fie o variabilă aleatoare X, care reprezintă lungimea unor bastoane, care are media $M(X) = 50$ și dispersia $D(X) = 0{,}1$. Utilizând inegalitatea lui Cebîşev, să se determine probabilitatea ca lungimea variabilei aleatoare X să fie între 49,5 cm şi 50,5 cm.

§2.8. Legi de repartiție continue clasice

1) Repartiția uniformă. *O variabilă aleatoare X este uniform repartizată pe segmentul $[a, b]$, dacă densitatea sa de repartiție este*

$$f(x) = \begin{cases} \dfrac{1}{b-a}, & dacă\ x \in [a,b], \\ 0, & dacă\ x \notin [a,b]. \end{cases}$$

Funcţia de repartiţie corespunzătoare este

$$F(x) = \begin{cases} 0, & dacă \quad x \le a, \\ \dfrac{x-a}{b-a}, & dacă \quad a < x \le b, \\ 1, & dacă \quad x > b. \end{cases}$$

Geometric, funcțiile $f(x)$ și $F(x)$ sunt reprezentate în figura 15.

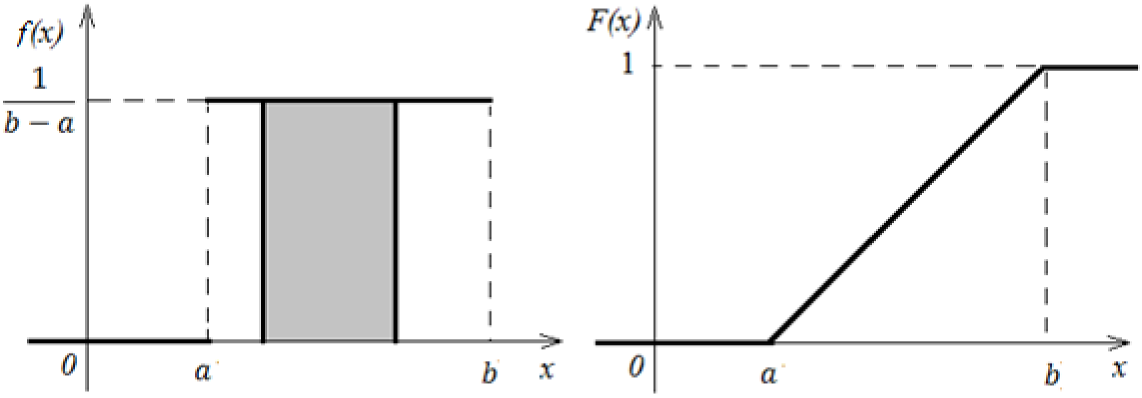


Fig. 15. Graficile funcțiilor $f(x)$ și $F(x)$ pentru repartiția uniformă

Probabilitatea că variabila X să ia valori între α și β $(a \leq \alpha < \beta \leq b)$, este

$$P\{\alpha < X < \beta\} = \int_{\alpha}^{\beta} \frac{dx}{b-a} = \frac{\beta - \alpha}{b-a}. \tag{50}$$

Această probabilitate este proporțională cu $\beta - \alpha$ și nu depinde de α și β, dar de lungimea acestui interval.

Teorema 1. *Dacă X este o variabilă aleatoare uniform repartizată pe intervalul închis $[a, b]$, atunci valoarea medie este*

$$M(X) = \int_{-\infty}^{\infty} xf(x)dx = \int_{a}^{b} \frac{x}{b-a} dx = \frac{a+b}{2}, \tag{51}$$

iar dispersia

$$D(X) = \int_{-\infty}^{\infty} (X - M(X))^2 f(x)dx = \int_{a}^{b} \left(x - \frac{a+b}{2}\right) \frac{1}{b-a} dx = \frac{(b-a)^2}{12}. \tag{52}$$

Observația 1. Abaterea medie pătratică $\sigma(X)$, în cazul variabilei aleatoare continue cu repartiție uniformă, este

$$\sigma(X) = \sqrt{D(X)} = \sqrt{\frac{(a-b)^2}{12}} = \frac{a-b}{2\sqrt{3}} = \frac{\sqrt{3}(a-b)}{6}. \tag{53}$$

2) Repartiția exponențială. Variabila aleatoare X urmează legea exponențială de repartiție cu parametrul $\lambda > 0$, dacă are densitatea de repartiție

$$f(x) = \begin{cases} 0, & dacă \;\; x \leq 0, \\ \lambda e^{-\lambda x}, & dacă \;\; x > 0, \lambda > 0. \end{cases}$$

Funcţia de repartiţie corespunzătoare este

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

deoarece

$$F(x) = \int_{-\infty}^{x} f(t)dt = \int_{-\infty}^{0} 0dt + \int_{0}^{x} \lambda e^{-\lambda t}dt = -e^{-\lambda t}\Big|_0^x = 1 - e^{-\lambda x}.$$

Geometric, funcțiile $f(x)$ și $F(x)$ sunt reprezentate în figura 16.

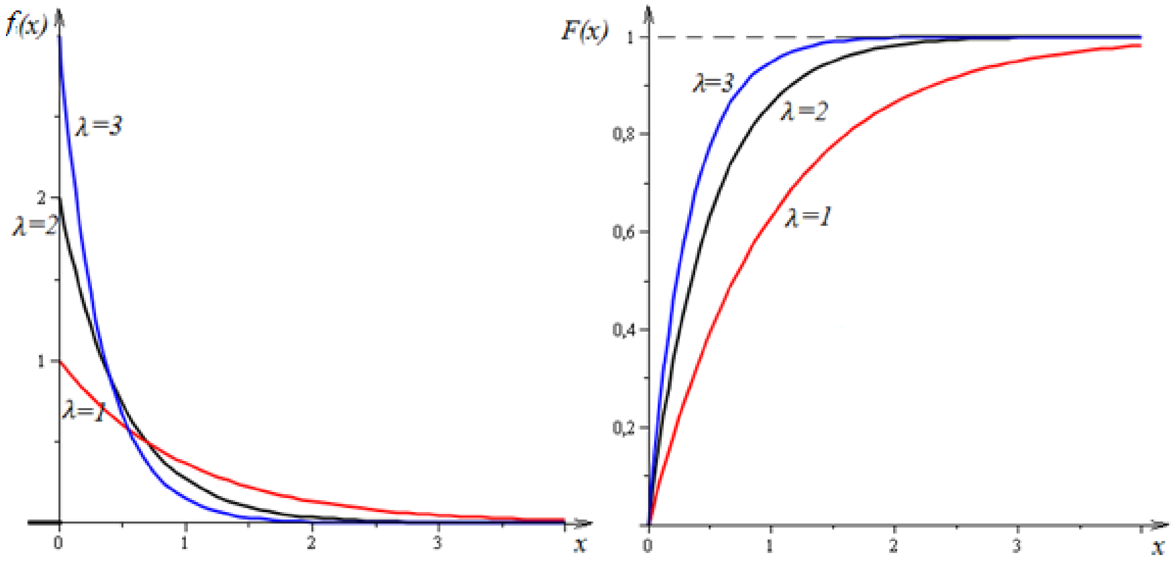


Fig. 16. Graficul densității și funcției de repartiție exponențiale

Dacă variabila aleatoare X urmează repartiția exponențială, atunci

$$P\{X < Y + Z/X \geq Y\} = P\{X < Z\} = 1 - e^{-\lambda x},$$

pentru oricare numere nenegative Y și Z.

Toerema 2. *Dacă X este variabilă aleatoare cu repartiția exponențială cu parametrul* $\lambda > 0$, *atunci valoarea medie este*

$$M(X) = \frac{1}{\lambda}, \tag{54}$$

iar dispersia

$$D(X) = \frac{1}{\lambda^2}. \tag{55}$$

Exemplul 1. Considerăm o variabilă t, *care reprezintă timpul necesar pentru instalarea unui condiționer, ce urmează o repartiție exponențială cu parametrul* $\lambda = 0{,}3$ *(ore). Să se determine probabilitatea că timpul de instalare a unui condiționer să nu întreacă 5 ore.*

Rezolvare. Probabilitatea, în cazul repartiției exponențiale, este

$$P\{t < 5\} = F(5) = 1 - e^{-0{,}3\cdot 5} = 1 - 0{,}223 = 0{,}777.$$

Răspuns: $P\{t < 5\} = 0{,}777$.

3) Repartiția normală (Gauss-Laplace). Este cea mai importantă lege de repartiție fiind cunoscută sub denumirea de legea Gauss-Laplace. *Variabila aleatoare continuă urmează legea de repartiție normală de parametrii* m *și* σ^2, *notată* $N(m, \sigma^2)$, *dacă are densitatea*

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}, m \in \mathbb{R}, \sigma > 0.$$

În mod evident $f(x) > 0$. Efectuând schimbarea de variabilă

$$x - m = \sigma t,$$

obținem

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

Funcţia de repartiţie corespunzătoare este

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

Geometric, funcțiile $f(x)$ și $F(x)$ sunt reprezentate în figura 17.

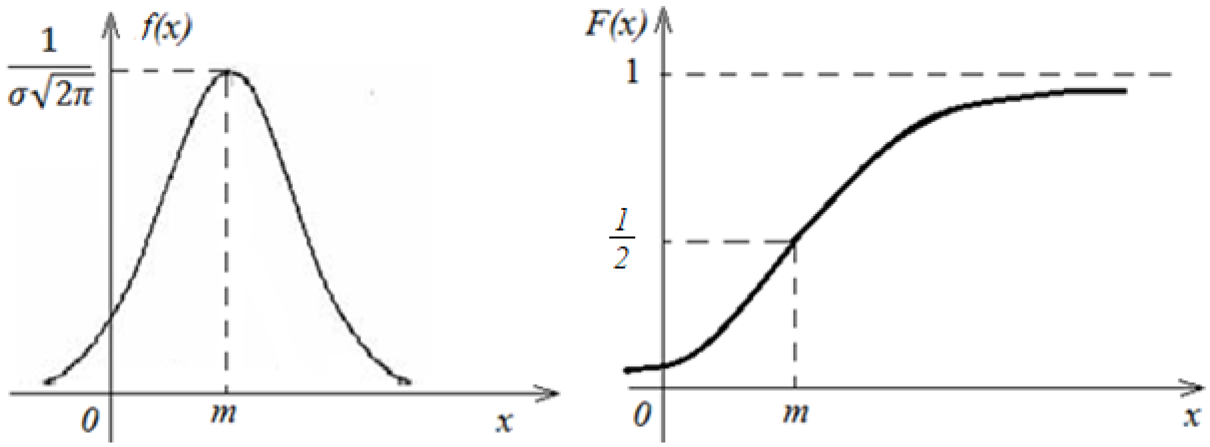


Fig. 17. Graficile funcțiilor $f(x)$ și $F(X)$ pentru repartiția normală

În cazul când $m = 0$ și $\sigma = 1$, atunci legea de repartiție se numește *legea normal normată* (sau *standart*) și se notează cu $N(0,1)$, cu densitatea de reprtiție

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

Observația 1. Valorile funcţiei $\varphi(x)$ sunt date tabelar (vezi tabelul A1).

Teorema 3. *În cazul când variabila aleatoare X urmează legea normală, atunci*

$$P(\alpha \leq X < \beta) = F(\beta) - F(\alpha) = \Phi\left(\frac{\beta - m}{\sigma}\right) - \Phi\left(\frac{\alpha - m}{\sigma}\right), \quad (56)$$

unde

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

Observația 2. Valorile funcţiei $\Phi(x)$ sunt date tabelar (vezi tabelul A2).

Foarte des, este nevoie de calculat probabilitatea abaterii de la $M(X)$ cu condiția ca să nu depășească ε

$$P(|X - M(X)| < \varepsilon).$$

În acest caz avem

$$P(-\varepsilon + m < X < \varepsilon + m) = \Phi\left(\frac{(m+\varepsilon) - m}{\sigma}\right) - \Phi\left(\frac{(m-\varepsilon) - m}{\sigma}\right) =$$

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

Dacă notăm $\frac{\varepsilon}{\sigma} = t$, atunci avem așa numita *regula celor 3 sigma*, iar în cazul când $t = 3$, avem

$$P(|X - m| < 3\sigma) = 3\Phi(3) \approx 0{,}99.$$

În figura 18, sunt ilustrate două curbe normale pentru $m = 0$, când $\sigma = 1$ și $\sigma = 2$.

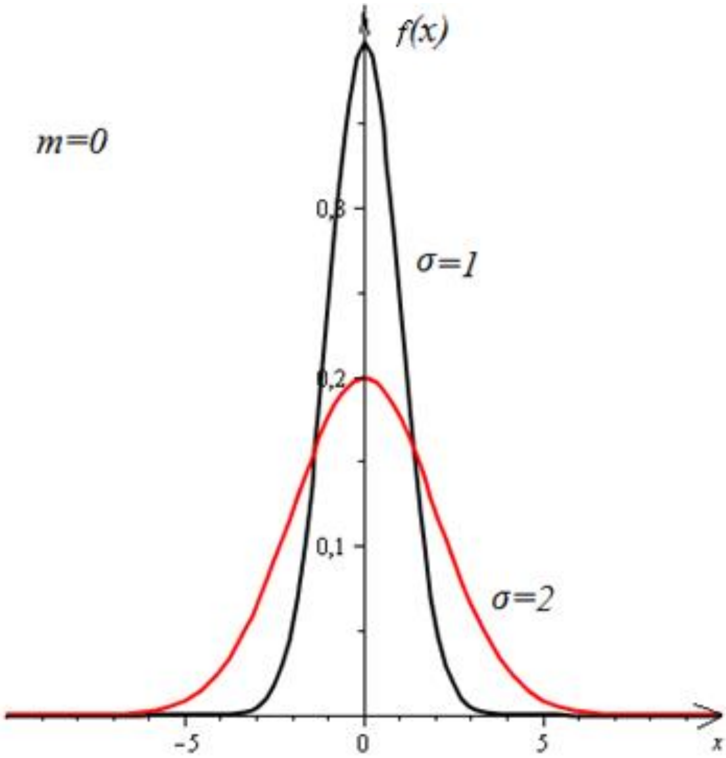


Fig. 18. Graficul curbelor normale pentru $m = 0$, când $\sigma = 1$ și $\sigma = 2$

Teorema 4. *În cazul variabilei aleatoare X cu repartiție normală, avem*

$$M(X) = m \qquad și \qquad D(X) = \sigma^2. \tag{57}$$

4) Repartiția Gamma generalizată. *O variabilă aleatoare are repartiția Gamma generalizată cu parametrii* $\alpha, \lambda > 0$, *dacă densitatea de repartiție este*

$$f(x) = \begin{cases} \dfrac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, & dacă\ x \geq 0, \\ 0, & dacă\ x < 0, \end{cases}$$

unde $\Gamma(\alpha)$ *este funcția Gamma*

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx.$$

Ultima elagitatea reprezintă *funcția Gamma*, care, geometric este reprezentată în figura 19.

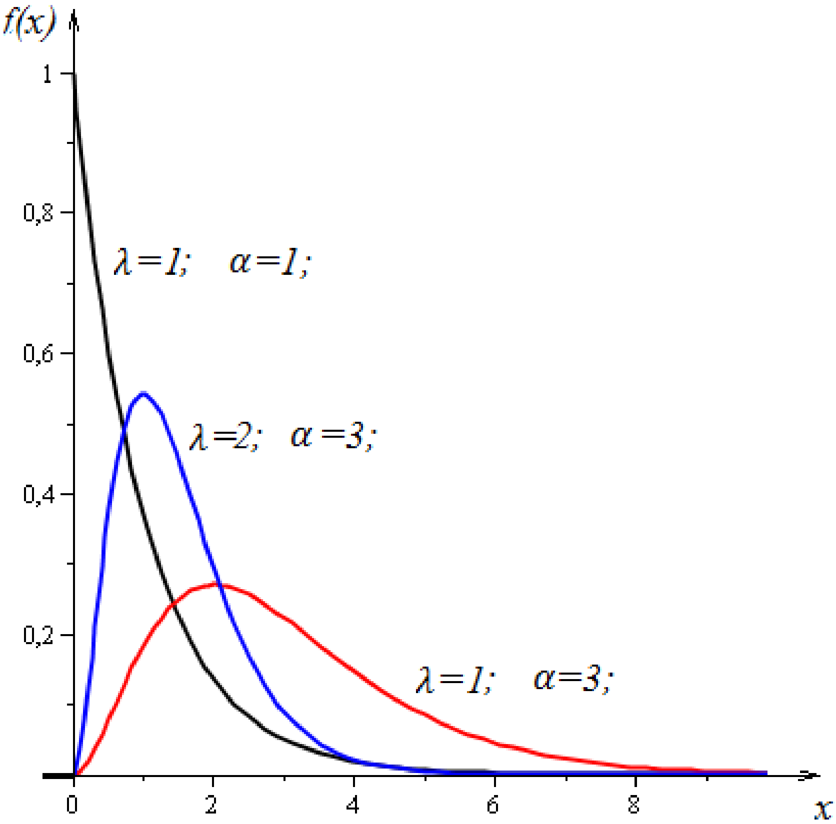


Fig. 19. Repartiția Gamma pentru diverse valori

Integrând $\Gamma(\alpha)$ prin părți, obținem

$$\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1), \alpha > 1.$$

Deoarece

$$\Gamma(1) = \int_0^\infty e^{-x} dx = 1,$$

obținem

$$\Gamma(n) = (n-1)!.$$

Teorema 5. *Dacă variabila aleatoare X urmează legea de repartiție Gamma generalizată, atunci momentele inițiale de ordinul k sunt*

$$m_k(X) = \frac{\alpha(\alpha+1)\dots(\alpha+k-1)}{\lambda^k}, \qquad k \in N^*. \tag{58}$$

Din ultima relaţie deducem media şi dispersia acestei repartiţii

$$M(X) = \frac{\alpha}{\lambda} \qquad \text{și} \qquad D(x) = \frac{\alpha}{\lambda^2}. \tag{59}$$

5) **Repartiția χ^2.** *Repartiția Gamma cu parametrii*

$$a = 1, \lambda = \frac{1}{2} \text{ și } n \in N,$$

se numește repartiția χ^2 (hi pătrat) cu n grade de libertate [21].

Considerăm o variabilă aleatoare X, cu valorile $x_1, x_2, x_3, \dots, x_n$ normal repartizate $N(0,1)$, atunci suma pătratelor variabilelor aleatoare z_i, constituie o nouă variabilă aleatoare și anume χ^2.

Matematic, aceasta reprezintă suma erorilor măsurătorilor până la valoarea 1

$$\chi_i^2 = \sum_{i=0}^{i} z_i^2 = \sum_{i=0}^{i} \frac{(x_i - m)^2}{\sigma}.$$

Densitatea de probabilitate a repartiţiei χ^2, pentru $\nu = n - 1$ grade de libertate, este

$$f(x) = \begin{cases} \dfrac{1}{2^{\nu/2}\Gamma\left(\frac{\nu}{2}\right)} x^{\left(\frac{\nu}{2}-1\right)} e^{-\frac{x}{2}}, & dacă \quad x > 0, \\ 0, & dacă \quad x \le 0. \end{cases}$$

Observația 3. Valorile densității de probabilitate a repartiţiei χ^2 sunt date tabelar (vezi tabelul A4).

În figura 20 sunt reprezentate graficile funcției $f(x)$, pentru diverse valori ale parametrului ν.

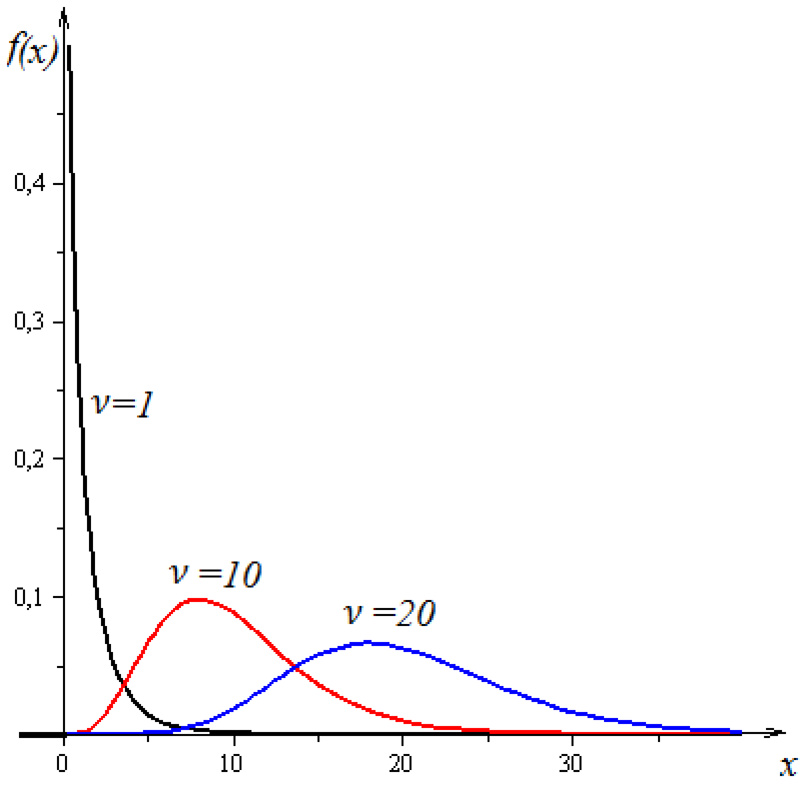


Fig. 20. Repartiția χ^2 pentru $\nu = 1, \nu = 10$ și $\nu = 20$

Teorema 6. *Dacă variabila aleatoare X urmează repartiţia χ^2 cu ν grade de libertate, atunci momentele iniţiale de ordinul k sunt*

$$m_k(X) = \nu(\nu+2)\ldots(\nu+2k-2), \qquad k \in N^*. \qquad (60)$$

Din ultima relaţie, deducem că media şi dispersia variabilei *X* cu repartiţie χ^2, cu ν grade de libertate, sunt

$$M(X) = \nu \quad și \quad D(X) = 2\nu. \qquad (61)$$

6) Repartiția Beta. *Variabila aleatoare X urmează legea beta (X are repartiţie beta) cu parametrii p şi q $(p,q > 0)$, dacă densitatea sa de probabilitate este*

$$f(x) = \begin{cases} \dfrac{1}{B(p,q)} x^{p-1}(1-x)^{q-1}, & dacă \;\; x \in (0,1), \\ 0, & dacă \;\; x \notin (0,1), \end{cases}$$

unde

$$B(p,q) = \int_0^1 x^{p-1}(1-x)^{q-1}dx, \;\; p,q > 0,$$

este integrala lui Euler de primul tip, sau funcţia beta a lui Euler [21].

Teorema 7. *Dacă variabila aleatoare X are repartiţie beta cu parametrii p şi q $(p,q > 0)$, atunci momentele iniţiale de ordinul k sunt*

$$m_k(X) = \frac{p(p+1)\ldots(p+k-1)}{(p+q)(p+q+1)\ldots(p+q+k-1)}, \qquad k \in N^*. \qquad (62)$$

Din ultima relaţie, deducem că media şi dispersia variabilei *X* cu repartiţie beta cu parametrii p şi q sunt

$$M(X) = \frac{p}{p+q} \quad și \quad D(X) = \frac{pq}{(p+q)^2(p+q+1)}. \qquad (63)$$

7) Repartiţia Student. *Variabila aleatoare X urmează legea Student (X are repartiţie Student) cu parametrul* ν (ν ∈ N^*), *dacă densitatea sa de probabilitate (repartiţie) este*

$$f(x) = \frac{1}{\sqrt{\nu\pi}} \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}, \quad x \in \mathbb{R}.$$

Cu alte cuvinte, ultima relație reprezintă *repartiţia Student cu ν grade de libertate*.

Teorema 8. *Dacă variabila aleatoare X are repartiţie Student cu ν grade de libertate, atunci momentele iniţiale de ordinul k, sunt*

$$m_{2k+1}(X) = 0, \qquad 2k+1 < \nu,$$

$$m_{2k}(X) = \frac{\nu^k(2k-1)!}{(\nu-2)(\nu-4)\ldots(\nu-2k)}, \qquad 2k < \nu \in N^*. \qquad (64)$$

Din ultima relație, rezultă că dacă $\nu > 1$, atunci media variabilei X este $M(X) = 0$, iar dacă $\nu > 2$, atunci dispersia variabilei X este

$$D(X) = \frac{\nu}{\nu - 2}. \quad (65)$$

Exerciții și probleme propuse

1. Fie X o variabilă aleatoare cu repartiția normală $N(m, \sigma^2)$ *și densitatea*

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}, m \in \mathbb{R}, \sigma > 0.$$

Să se determine valoarea medie și dispersia variabilei aleatoare X.

2. *Să se determine momentele de ordinul unu și doi a repartiției exponențiale* cu parametrul λ.

3. Să se calculeze probabilitatea $P\{0 < X < 3\}$, *pentru variabila aleatoare X, cu repartiția normală* $N(2, 2^2)$.

4. Fie X o variabilă aleatoare cu repartiție uniformă în intervalul $(-1,1)$. *Se se determine valoarea medie și dispersia variabilei aleatoare X.*

5. *O variabilă aleatoare X are repartiție normală cu parametrii* $M(X) = 30$ *și σ =10. Să se determine probabilitatea că variabila aleatoare X să ia valori în intervalul (10,50).*

CAPITOLUL III. ELEMENTE DE STATISTICĂ MATEMATICĂ

“Într-o zi, gândirea statistică
va fi la fel de necesară oricărui cetățean folositor societății,
ca și scrisul sau cititul”
H. G. Wells

§3.1. Noțiuni generale

În acest capitol este explicată esența statisticii, cum diferă ea de teoria probabilităților, ce o leagă de aceasta, care sunt părțile ei componente și cum începe demersul practic într-o problemă de statistică.

Când omul nu a mai putut intuit, a început să măsoare. Măsurătorile și observațiile au devenit prima treaptă spre întelegerea legilor naturii. Dar, în acest fel, omul nu mai poate să cunoască direct realitatea, el poate numai să o aproximeze succesiv prin modele fizice și apoi prin modele matematice. Dar aceste modele nu descriu exact *realitatea*. Ele o aproximează și apar așa numitele erori. Unele erori sunt previzibile, altele însă sunt întâmplătoare (aleatoare). Aceste ultime erori (aleatoare), au și ele legile lor de manifestare. Apar deci fenomenele aleatoare descrise prin variabilele aleatoare.

Teoria probabilitătilor pleacă de la ipoteza că se cunosc exact aceste variabile aleatoare și anume prin funcțiile de probabilitate, prin funcțiile de repartiție, prin funcțiile caracteristice, etc. Statistica pleacă de la măsuratorile brute și caută să regăsească modelul probabilistic teoretic exact, care se află în spatele acestor măsuratori [4].

Deci, statistica și probabilitatea sunt două domenii strâns legate, dar distincte ale matematicii.

Se spune că "probabilitățile sunt vehiculul statisticii". Aceasta este adevărat, în sensul că dacă nu ar fi legile probabiliste, teoria statistică nu ar fi posibilă. Pentru a ilustra însă diferența dintre probabilități și statistică, să considerăm două urne: una probabilistică și una statistică. În cazul urnei probabilistice se știe că urna conține 5 bile albe, 5 bile negre și 5 bile roșii; problema de probabilitate este: dacă scoatem o bilă, care este șansa ca aceasta să fie albă? În cazul unei urne statistice, nu cunoaștem care este combinația de bile din urnă. Extragem un eșantion și din acest eșantion conjecturăm ce credem că se găsește în urnă. Trebuie reținută deosebirea: probabilitatea pune întrebarea șansei ca ceva (un eveniment) să se întâmple atunci când se cunosc posibilitățile (se cunoaște populația). Statistica ne cere să selectăm un eșantion, să analizăm eșantionul și pe urmă să facem predicție asupra populației pe baza informației găsite la cercetarea eșantionului.

Utilizarea statisticii este nelimitată. Este greu de găsit un domeniu în care statistica nu se folosește. Iată câteva exemple, unde și cum este folosită statistica [1]:

- ✓ în educație - statistica descriptivă este adesea folosită pentru a prezenta rezultatele;
- ✓ în știință - rezultatele experimentale trebuiesc colectate și analizate;
- ✓ guvernele - adună diferite date statistice.

Mulți oameni sunt indiferenți față de descrierea statistică, alții cred că statisticile sunt minciuni. Majoritatea minciunilor statistice sunt inocente și rezultă din folosirea unei statistici neadecvate, sau date obținute dintr-un eșantion nepotrivit. Toate acestea conduc la o înțelegere greșită a informației din partea consumatorului. Folosirea greșită a statisticii duce uneori la încurcături majore.

§3.2. Selecții. Funcția empirică de repartiție

În general, la studierea unei mărimi pe cale experimentală se efectuează diferite măsurări și se înregistrează rezultatele. Este clar că în continuare, datele trebuie prelucrate pentru a putea face câteva concluzii, de aceasta se ocupă statistica.

Definiția 1 [2]. *Statistica este știința colectării, clasificării, prezentării, interpretării datelor numerice și a folosirii acestora pentru a formula concluzii și a lua decizii.*

Există două tipuri de statistică: descriptivă și inferențială.

Definiția 2. *Statistica descriptivă se ocupă cu colectarea, clasificarea și prezentarea datelor numerice.*

Definiția 3. *Statistica inferențială (inferential statistics) se ocupă cu interpretarea datelor oferite de statistica descriptivă și cu folosirea acestora pentru a formula concluzii și lua decizii.*

Fie o problem de cercetare, atunci mulțimea de obiecte care sunt supuse cercetării se numește *populație statistică*, iar fiecare element separat din populația statistică se numește *unitate statistică, membru al populației* sau *individ*. Evident, populația statistică poate fi o mulțime atât finită cât și infinită, dacă este infinită se ia o parte a populației și se cercetează (din populația statistică).

Trăsătura comună a tuturor membrilor unei populații, care ne interesează în cercetare, se numește *caracteristică*. În statistică se studiază două tipuri de caracteristici:

- ➢ calitative;
- ➢ cantitative.

Caracteristici calitative – mărimile care exprimă o anumită calitate, de exemplu culoarea ochilor, culoarea părului, cetățenia, sexul, etc.

Caracteristici cantitative – mărimile care pot fi măsurate, de exemplu înălțimea, greutatea, nota la un examen/evaluare, etc.

Dăcă caracteristica care se cercetează se schimbă cantitativ, atunci ea se numește *variabilă statistică*, iar dacă se schimbă calitativ, se numește *atribut*.

Numărul de elemente din care este formată populația statistică, se numește *volumul populației* sau *efectivul total*.

O populație statistică poate avea mai multe caracteristici.

Dacă se studiază fiecare componentă a populației statistice, atunci se spune că se efectuează o cercetare completă, iar cercetările care se fac asupra unei părți a populației, se numește *cercetare selectivă*, atunci indivizii care au fost selectați, formează selecția sau eșantionul.

Fie că $x_1, x_2, \dots, x_n$ sunt valorile de selecție observate pentru o variabilă X. Valorile caracteristicii se aranjează în ordine crescătoare

$$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$$

și în așa mod se obține seria variațională.

Definiția 4 [1]. *Șirul de perechi ordonate (x_i, n_i), reprezintă repartiția de frecvențe (seria statistică) a variabilei statistice X*

$$X\colon \begin{pmatrix} x_1 & x_2 & x_3 & \dots & x_k \\ n_1 & n_2 & n_3 & \dots & n_k \end{pmatrix},$$

care variază de la o selecție la alta, unde

$$\sum_i n_i = n \quad \text{și} \quad \sum_i \frac{n_i}{n} = 1.$$

Selecția este punctul de plecare al oricărei cercetări statistice asupra unei variabile aleatoare X date. În continuare, notăm funcția de repartiție prin F, x un număr real arbitrar, iar ν_x numărul valorilor de selecție mai mic ca x.

Definiția 5. *Frecvența relativă $\frac{\nu_x}{n}$ a valorilor selecției date mai mici ca x, se numește funcție epirică de repartiție.*

Funcția dată se noteză cu $F_n(x)$, unde n este volumul selecției cercetate, deci

$$F_n(x) = \frac{\nu_x}{n}.$$

Menționăm că funcția empirică $F_n(x)$, posedă aceleași proprietăți ca și funcția de repartiție $F(x)$, a unei variabile aleatoare X (vezi §2.3).

Observația 1. Cu cât este mai mare volumul selecției cu atât, funcția empirică de repartiție ne dă o imagine mai precisă despre funcția teoretică de repartiție.

Dacă avem un număr mare de experiențe, atunci construcția funcției empirice de repartiție $F_n(x)$, este mai complicată. În acest caz este mai comod să utilizăm alte caracteristici ale repartiției empirice analogice densității de reparție $f(x)$.

Definiția 6. *Numărul unităților cu aceeași valoare sau aparținând unuia și aceluiași interval, se numește frecvență absolută.*

Definiția 7. *Frecvența absolută cumulată se numește suma frecvențelor absolute a tuturor variabilelor ce apar până la x_i, frecvența relativă cumulată se numește raportul dintre frecvența absolut cumulată și numărul total de elemente al populației stasistice.*

§3.2. Gruparea datelor. Reprezentarea grafică a datelor statistice

Colectarea datelor referitoare la o caracteristică a populației statistice urmărește scopul stabilirii valorii care poate caracteriza un individ sau formarea unei concluzii observate.

Statistica discriptivă reprezintă forma cea mai simplă de analiză a unei caracteristici a populației statistice și include: colectarea datelor în formă de tabel, întocmirea reprezentării grafice și stabilirea indicatorilor statistici.

În cazul când valorile caracteristice sunt aranjate, atunci primim tabelul primar.

Exemplul 1. În rezultatul unei evaluări, studenții grupei X au primit următoarele rezultate:

9, 4, 8, 7, 6, 8, 7, 6, 10, 8, 5, 6, 7, 5, 8, 9, 6, 7, 5, 8, 5, 9, 7, 6, 8, 5, 8, 6, 7, 9, 8, 7.

Să se grupeze datele într-un tabel.

Rezolvare. Observăm că aceste note (valori), pot fi grupate după valoarea lor.

Tabelul. 9. Gruparea rezultatelor obținute de studenți

Notele	10	9	8	7	6	5	4	3	2	1
Nr. studenților	1	4	8	7	6	5	1	0	0	0

Valorile unei caracteristici reprezentate în tabel, nu întotdeauna sunt simple, de exemplu cercetăm toată facultatea după o anumită caracteristică. În cazul când datele sunt multe, pot fi utilizate și tebele restrânse, adică gruparea pe intervale.

Gruparea simplă, pentru anumite cazuri, reprezintă un tabel în care sunt valorile caracteristicii și numărul de apariții respective (vezi tabelul din exemplu 1). În cazul când avem *m* intervale distincte, gruparea valorilor pe acest interval ne va conduce la un număr de intervale concret, și respectiv, la o frecvență anumită.

În cazul când este nevoie de gruparea pe intervale, lungimea intervalului se calculează în baza formulei

$$l = \frac{x_{max} - x_{min}}{m},$$

unde x_{max} reprezintă valoarea maximă a caracteristicii, x_{min} reprezintă valoarea minimă a caracteristicii, m se calculează după *regula lui Sturges*

$$m = 1 + \frac{10}{3}\lg n,$$

n este numărul total de date, iar când $n > 100$, m se calculează după *regula lui Wald și Mann*

$$m = 4\left(\frac{1}{4}(n-1)\right)^{\frac{1}{5}}.$$

Reprezentarea grafică a datelor statistice, se numește *diagramă* și în general, poate să ne ofere informația privind legile cărora se supun caracteristicile date

În dependență de reprezentarea grafică, diagramele se împart în:

1) diagrame prin butoane;
2) histograme;
3) poligonul frecvențelor absolute;
4) poligonul frecvențelor relative;
5) diagrame structurale.

Diagramele prin butoane se aplică în cazul când caracteristica ia valori discrete care nu sunt grupate în clase și atunci pe axa O*X* se i-au valorile caracteristicii $x_1, x_2, \dots, x_n$ și în fiecare punct se duc drepte (butoane) verticale de lungimi respective (fig. 21).

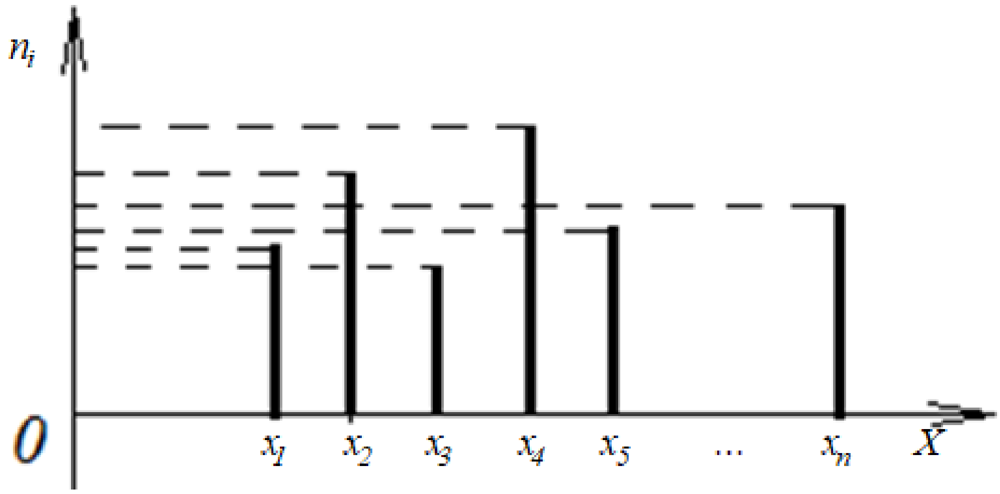


Fig. 21. Diagramă prin butoane

Exemplul 2. Să se reprezinte diagrama prin butoane a datelor din exemplul 1.

Rezolvare.

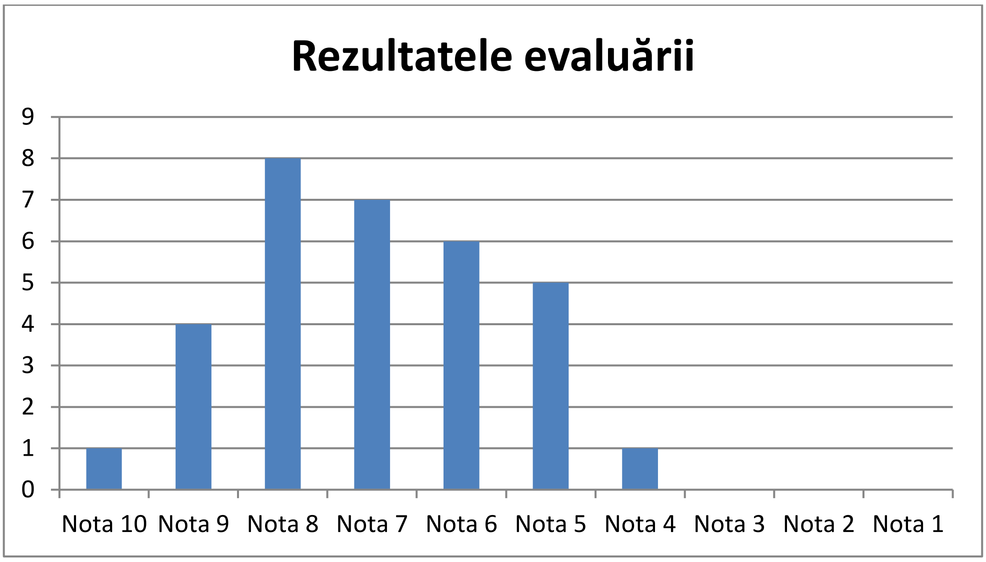


Fig. 22. Reprezentarea rezltatelor evaluării unei grupe de studenți

Histograma este un termen introdus de Karl Pearson (*gr. histos - țesut; grama – scriere, desen*) și se utilizează în cazul în care volumul selecției este mare. În acest caz, datele sunt grupate pe intervale disjuncte, egale între ele [1].

Presupunem că datele de selecție aparțin unui interval $[a, b]$ care se împarte în intervale distincte $[c_{i-1}, c_i)$, $i = 1, \dots, k-1$ și $[c_{i-1}, c_i]$, $i = k,$ cu capetele $c_0 = a$ și $c_k = b$.

Fie n_i numărul valorilor din selecție care aparțin intervalului

$$[c_{i-1}, c_i),\ i = 1, \dots, k-1 \text{ și } [c_{i-1}, c_i],\ i = k.$$

Deci n_i reprezintă frecvența absolută.

Datele obținute se trec într-un tabel de forma

Tabelul 10. Gruparea datelor pe intervale

Intervalul	Frecvența absolută
$[c_0, c_1)$	n_1
$[c_1, c_2)$	n_2
$[c_2, c_3)$	n_3
…	…
$[c_{k-1}, c_k]$	n_k

În baza tabelul 10 se construiește histograma, pe axa OX vor fi reprezentate segmente de lungimile intervalului respectiv și pe aceste intervale se construiesc dreptunghiuri de înălțimile n_i (fig. 23).

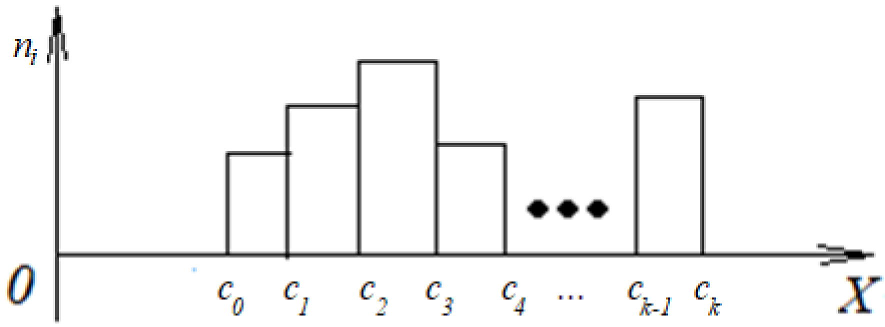


Fig. 23. Histogramă

Exemplul 3. În rezultatul unei evaluări, studenții grupei Y au primit următoarele rezultate: 9,9; 8,7; 5,5; 6,3; 8,0; 7,8; 10; 8,4; 5,0; 7,1; 6,6; 8,2; 7,9; 8,8; 6,4; 5,8; 7,4; 7,3; 8,1; 6,9; 9,3; 5,9; 6,4; 7,5; 8,9; 7,5 9,1; 8,6; 7,9; 8,3; 7,7; 8,5; 7,1; 8,0 și 6,7. Să se reprezinte histograma rezultatelor obținute.

Rezolvare. Observăm că aceste note pot fi grupate pe intervale

Tabelul. 11. Gruparea rezultatelor pe intervale

Intervalul	[5,6)	[6,7)	[7,8)	[8,9)	[9,10]
Frecvența absolută	4	6	10	11	4

În baza tabelului 11, construim histograma corespunzătoare datelor exemplului (fig. 24).

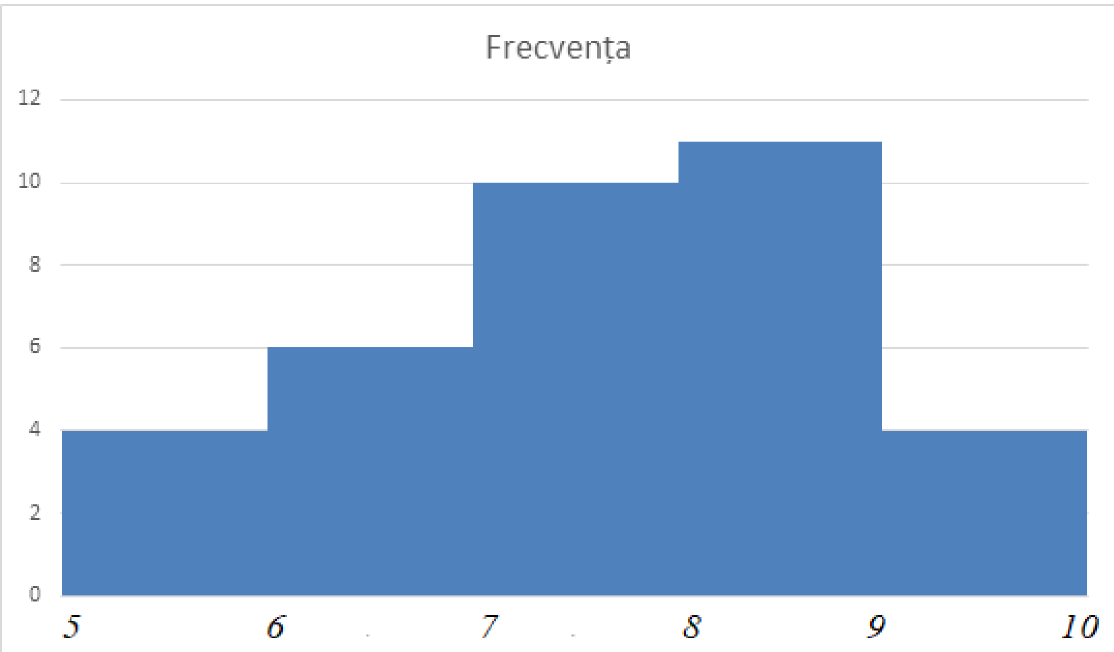


Fig. 24. Reprezentarea datelor problemei 3, prin histogramă

Poligonul frecvențelor absolute este linia poligonală care unește punctele (x_i, n_i), unde x_i reprezintă mijlocul intervalului *i*

$$x_i = \frac{1}{2}(c_{i-1} + c_i), \tag{66}$$

iar n_i este frecvența absolută.

Datele obținute se trec într-un tabel, de forma

Tabelul 12. Gruparea datelor și frecvența absolută corespunzătoare fiecărui interval

Intervalul	Mijlocul intervalului	Frecvența absolută
$[c_0, c_1)$	x_1	n_1
$[c_1, c_2)$	x_2	n_2
$[c_2, c_3)$	x_3	n_3
…	…	…
$[c_{k-1}, c_k]$	x_k	n_k

În baza tabelului 12, se construiește poligonul frecvențelor absolute, pe axa OX fiind reprezentat șirul x_i, adică mijlocurile intervalelor cercetate și respectiv frecvențele absolute corespunzătoare n_i. Unind aceste puncte vom primi poligonul necesar (fig. 25).

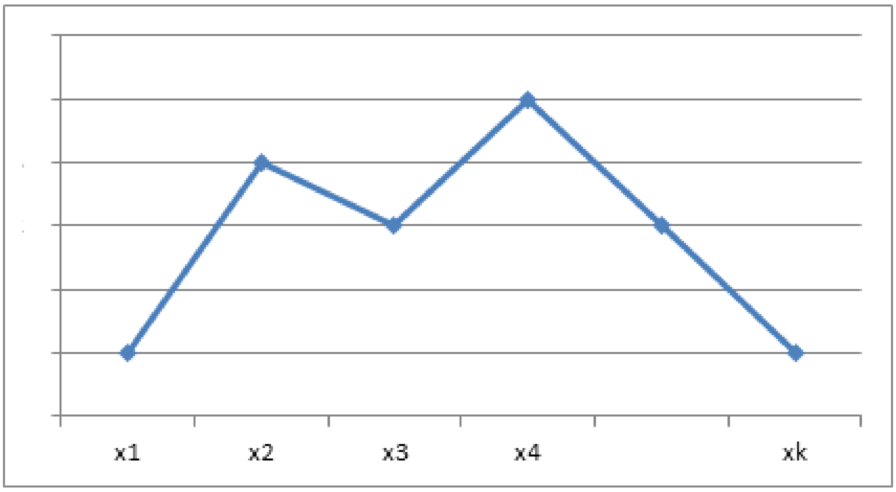


Fig. 25. Poligonul frecvențelor absolute

Exemplul 4. Să se reprezinte poligonul frecvențelor absolute pentru datele din exemplul 3.

Rezolvare. În exemplul 3, datele au fost grupate pe intervale, însă, trebuie să determinăm și mijlocul pentru fiecare interval, în baza formulei (66), obținem

$$x_1 = \frac{1}{2}(5 + 6) = 5{,}5; \qquad x_2 = \frac{1}{2}(6 + 7) = 6{,}5;$$

$$x_3 = \frac{1}{2}(7 + 8) = 7{,}5; \qquad x_4 = \frac{1}{2}(8 + 9) = 8{,}5;$$

$$x_5 = \frac{1}{2}(9 + 10) = 9{,}5.$$

Tabelul. 13. Gruparea datelor pe intervale, mijlocul intervalelor și frecvența absolută corespunzătoare fiecărui interval

Intervalul	Mijlocul intervalului	Frecvența absolută
[5,6)	5,5	4
[6,7)	6,5	6
[7,8)	7,5	10
[8,9)	8,5	11
[9,10]	9,5	4

În baza tabelului 13, construim poligonul frecvențelor absolute corespunzătoare datelor exemplului (fig. 26)

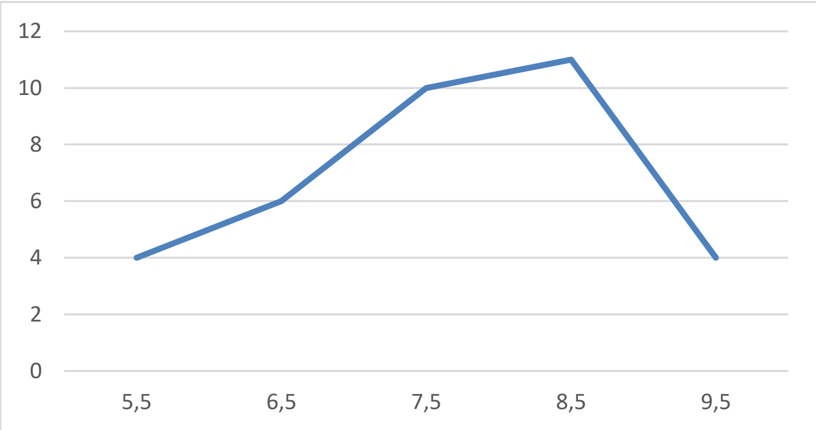


Fig. 26. Poligonul frecvențelor absolute, exemplul 4

Poligonul frecvențelor relative este linia poligonală care unește punctele (x_i, f_i), unde x_i reprezintă mijlocul intervalului *i*, iar

$$f_i = \frac{n_i}{n} \tag{67}$$

frecvența relativă (greprezentarea grafică este asemenea *poligonului frecvențelor absolute*).

Datele obținute se trec într-un tabel de forma

Tabelul 14. Gruparea datelor și frecvența relativă corespunzătoare fiecărui interval

Intervalul	Mijlocul intervalului	Frecvența relativă
$[c_0, c_1)$	x_1	f_1
$[c_1, c_2)$	x_2	f_2
$[c_2, c_3)$	x_3	f_3
…	…	…
$[c_{k-1}, c_k]$	x_k	f_k

În baza tabelului 14, se construiește poligonul frecvențelor relative, pe axa OX fiind reprezentat șirul x_i, adică mijlocurile intervalelor cercetate și respectiv frecvențele relative corespunzătoare f_i. Unind aceste puncte vom primi poligonul necesar (fig. 27).

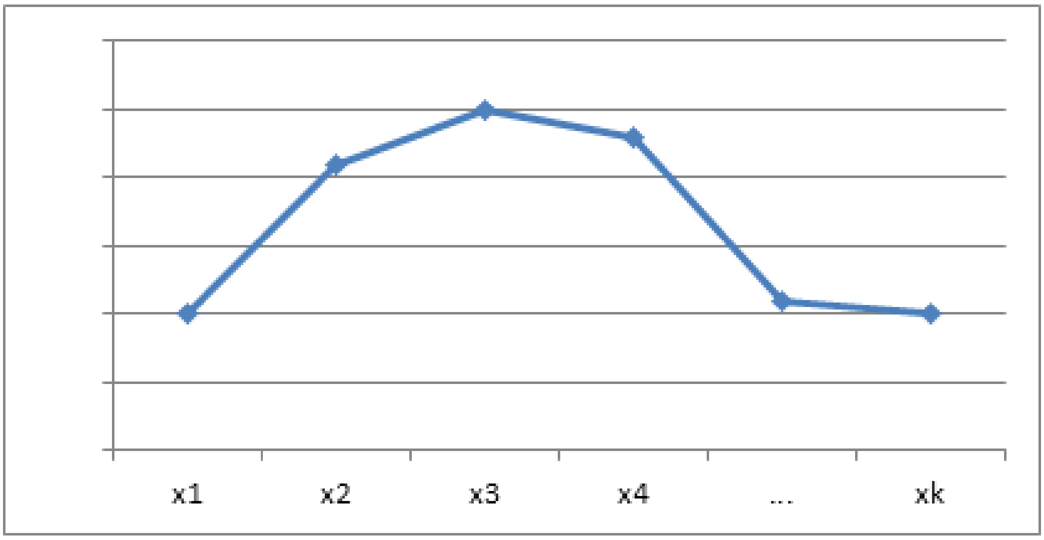


Fig. 27. Poligonul frecvențelor relative

Exemplul 5. Să se reprezinte poligonul frecvențelor relative pentru datele din exemplul 3.

Rezolvare. În exemplul 3, datele au fost grupate pe intervale, iar în exemplul 4, au fost determinate mijlocurile intervalelor, însă, trebuie să determinăm și frecvențele absolute pentru fiecare interval. În baza formulei (67), obținem

$$f_1 = \frac{4}{35} \approx 0{,}11; \qquad f_2 = \frac{6}{35} \approx 0{,}17;$$
$$f_3 = \frac{10}{35} \approx 0{,}29; \qquad f_4 = \frac{11}{35} \approx 0{,}32;$$
$$f_5 = \frac{4}{35} \approx 0{,}11.$$

Tabelul. 15. Gruparea datelor pe intervale, mijlocul intervalelor și frecvența absolută corespunzătoare fiecărui interval

Intervalul	Mijlocul intervalului	Frecvența relativă
[5,6)	5,5	0,11
[6,7)	6,5	0,17
[7,8)	7,5	0,29
[8,9)	8,5	0,32
[9,10]	9,5	0,11

În baza tabelului 15, construim poligonul frecvențelor relative corespunzătoare datelor exemplului (fig. 28)

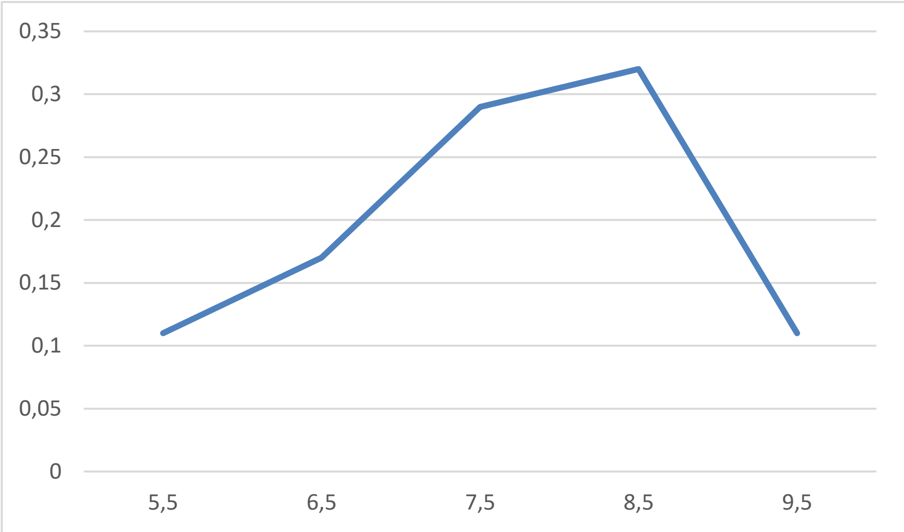


Fig. 28. Poligonul frecvențelor relative, exemplul 5

Diagrame structurale sunt folosite pentru a vizualiza, specifica, construi și documenta, aspectele statice ale sistemului. Acestea sunt organizate în jurul grupelor majore de elemente ce pot fi găsite în modelarea unui sistem. Din aceasta categorie fac parte:

- Diagrame de clase – conțin o mulțime de clase, interfețe și colaborări, precum și relațiile dintre ele. Sunt cele mai folosite diagrame ce ilustrează vederea statică de proiectare a unui sistem.
- Diagrame de obiecte – conțin o mulțime de obiecte și relațiile dintre ele. Sunt folosite pentru a ilustra structurile de date, imagini statice ale instanțelor elementelor din diagramele de clase.
- Diagrame de componente – conțin o mulțime de componente și relațiile dintre ele. Se folosesc pentru a ilustra vederea statică de implementare a unui sistem.
- Diagrame de amplasare – conțin o mulțime de noduri și relațiile dintre ele, ilustrând vederea statică de amplasare a unei arhitecturi.

Exemplul 6. Să se reprezinte datele din exemplul 3, printr-o diagramă structurală.

Rezolvare. În exemplul 3, datele au fost grupate pe intervale și determinată frecvența pe aceste intervale (tabelul 11). În baza tabelului 11, construim o diagramă structurală circulară (fig. 29).

Fig. 29. Diagrama structurală corespunzătoare datelor exemplului 6

Exerciții și probleme propuse

1. Fie dată repartiția elevilor unui liceu, după rezultatele obținute la examenul de limbă română, bacalaureat 2021 (tabelul 16)

Tabelul 16. Repartiția elevilor după rezultatul la limba română, bacalaureat 2021

Nota	*5*	*6*	*7*	*8*	*9*	*10*
Nr. de elevi	*38*	*46*	*86*	*72*	*78*	*51*

Să se reprezinte grafic seria statistică:

a) prin butoane;

b) printr-o diagramă structurală.

2. Un cercetător a studiat timpul de funcționare a 60 de lămpli (în ore) și a obținut următoarele date: 134, 147, 180, 154, 156, 136, 184, 149, 165, 162, 129, 143, 168, 175, 154, 125, 168, 148, 152, 167, 138, 172, 165, 141, 180, 157, 139, 147, 162, 155, 165, 148, 135, 168,159, 145, 167, 144, 158, 152, 146, 171, 158, 164, 133, 169, 155, 168, 142, 160, 157, 166, 148, 137, 152, 169, 134, 182, 190, 125. Să se reprezinte grafic datele obținute. Să se construiască poligonul frecvențelor absolute și relative.

3. Să se cerceteze notele, la obiectul X, a unei grupe (a unui torent) de studenți a facultății. Să se reprezinte grafic datele obținute. Să se construiască poligonul frecvențelor absolute și relative.

§3.3. Indicatorii variabilelor statistice (seriilor statistice)

La cercetarea unei serii satistice este nevoie de câteva caracteristici care ar indica unele rezultate sau concluzii asupra volumului sau populației statistice. Pentru orice caracteristică a populației se observă o tendință de variație cu două aspecte:

- de localizare în jurul unei valori (localizarea poate fi simetrică și asimetrică);
- de împrăștiere.

Dacă dorim să analizăm localizarea, atunci trebuie să cunoaștem că graficile de frecvență ne prezintă doar aspecte calitative, doar o analiză mai profundă care ne permite compararea tendințelor de localizare/împrăștiere poate fi realizată cu ajutorul indicatorilor statistici.

Definiția 1 [24]. *Indicatorul statistic este expresia numerică a unui fenomen, proces sau a unei categorii economico-sociale, definite în timp, spaţiu şi structură organizatorică. Obţinut ca rezultat al procesului cercetării statistice, indicatorul are conţinut real, obiectiv determinat, o formulă de calcul şi o formă specifică de exprimare.*

Indicatorii, ca și aspectele tendinței de variație, se clasifică în:

- indicatori de poziție sau localizare;
- indicatori de variație sau de împrăștiere.

§3.3.1. Indicatorii de poziție sau localizare

Un indicator de poziție este un număr real care face posibilă localizarea valorilor unei serii statistice a variabilei cantitative. Poate fi un indicator de tendință central sau o valoare descentrată, ca maximul sau minimul seriei [1], care în unele cazuri să nu existe printre valorile populației. De fapt, acest indicator ne indică valoarea spre care tinde să se grupeze datele reale.

Indicatorul de bază al tendinței de localizare este media, care se noteză cu $M(X)$.

Clasificarea mărimilor medii se diferenţiază între ele, în funcţie de rolul pe care îl ocupă în analiza statistică şi după modul de obţinere a lor [24]:

1. Clasificrea mediilor după rolul pe care îl ocupă în statistică:

a) mărimi fundamentale (media aritmetică, modul, mediana);

b) mărimi cu aplicaţii speciale (media geometrică, armonică, pătratică).

2. Clasificrea mediilor după modul de obţinere:

a) mărimi medii de calcul (media aritmetică, armonică, geometrică, pătratică);

b) medii de poziţie (modul, mediana).

Media aritmetică se notează $\bar{x}$ și se utilizează când fenomenul supus cercetării înregistrează modificări aproximativ constante în progresia aritmetică.

Media aritmetică este de două tipuri:

- simplă;
- ponderată.

Dacă există o caracteristică numerică x, care ia valorile $x_1, x_2, x_3, \dots, x_n$, atunci

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} \tag{68}$$

reprezintă media aritmetică simplă.

Exemplul 1. Un elev a obținut următoarele note la semestrul I: 8,75; 9,4; 8,2; 9,15; 8,45; 9,5; 7,9; 8,25; 8,85 și 9,1. Să se determine media simestrială a elevului.

Rezolvare. Aplicând formula (68), obținem

$$\bar{x} = \frac{8{,}75 + 9{,}4 + 8{,}2 + 9{,}15 + 8{,}45 + 9{,}5 + 7{,}9 + 8{,}25 + 8{,}85 + 9{,}1}{10} =$$
$$= 8{,}755.$$

Răspuns: $\bar{x} = 8{,}755$

În cazul seriilor de distribuţie, când unele caracteristici se înregistrează de mai multe ori, media aritmetică simplă se înlocuiește cu media aritmetică ponderată. Și anume, dacă caracteristicile numerice $x_1, x_2, x_3, \dots, x_m$ apar cu frecvențele $n_1, n_2, \dots, n_m$, atunci media aritmetică ponderată este

$$\bar{x} = \frac{n_1 x_1 + n_2 x_2 + \cdots + n_m x_m}{\sum_{i=1}^{m} n_i}, \qquad i = 1, .., m. \tag{69}$$

Exemplul 2. Întreprinderea "Ionel" are 1000 de angajați a căror salarii au fost grupate pe intervale (tabelul 17). Să se determine salariul mediu primit de angajații întreprinderii menționate.

Tabelul. 17. Salariul angajaților întreprinderii "Ionel"

Salariul (lei)	Numărul de angajați (n_i)
[3500,4500)	10
[4500,5500)	30
[5500,6500)	100
[6500,7500)	150
[7500,8500)	250
[8500,9500)	210
[9500,10500)	200
[10500,11500]	50

Rezolvare. Se determină mijlocurile intervalelor x_i

$$x_1 = 4000;\ x_2 = 5000; x_3 = 6000;\ x_4 = 7000;$$

$$x_5 = 8000;\ x_6 = 9000; x_7 = 10000;\ x_8 = 11000.$$

Aplicând formula (69), obținem

$$\bar{x} = \frac{1}{1000}(4000 \cdot 10 + 5000 \cdot 30 + 6000 \cdot 100 + 7000 \cdot 150 +$$
$$+8000 \cdot 250 + 9000 \cdot 210 + 10000 \cdot 200 + 11000 \cdot 50) = 8280\ (lei).$$

Răspuns: $\bar{x} = 8280\ (lei)$

Media armonică se notează $\bar{x}_h$ și se calculează din valorile inverse ale termenilor seriei ca medie simplă şi ponderată.

Media armonică simplă se calculează în baza formulei

$$\bar{x}_h = \frac{n}{\sum_{i=1}^{m} \frac{1}{x_i}}, \tag{70}$$

iar media armonică ponderată

$$\bar{x}_h = \frac{\sum_{i=1}^{m} n_i}{\sum_{i=1}^{m} \frac{n_i}{x_i}}. \tag{71}$$

Exemplul 3. Fie datele din exemplul 2. Să se determine salariul mediu al angajaților aplicând media armonică.

Rezolvare. Pentru comoditate vom dezvolta tabelul 17, adăugând coloanele $\frac{1}{x_i}$ și $\frac{1}{x_i} \cdot n_i$, în următorul tabel:

Tabelul. 18. Salariul angajaților întreprinderii "Ionel"

Salariul (lei)	Numărul de angajați (n_i)	x_i	$\frac{1}{x_i}$	$\frac{1}{x_i} \cdot n_i$
[3500,4500)	10	4000	0,00025	0,0025
[4500,5500)	30	5000	0,00020	0,0060
[5500,6500)	100	6000	0,00017	0,0167
[6500,7500)	150	7000	0,00014	0,0214
[7500,8500)	250	8000	0,00012	0,0313
[8500,9500)	210	9000	0,00011	0,0233
[9500,10500)	200	10000	0,00010	0,0200
[10500,11500]	50	11000	0,00009	0,0045

Rezolvare. Aplicând formula (71), obținem

$$\bar{x}_h = \frac{1000}{0{,}1257} \approx 7955{,}45\ (lei).$$

Răspuns: $\bar{x} \approx 7955{,}45\ \ (lei)$

Media pătratică se notează $\bar{x}_p$ și este acea valoare, care înlocuind termenii seriei ridicaţi la pătrat nu modifică suma pătratelor lor. În cazul seriilor simple avem

$$\bar{x}_p = \sqrt{\frac{\sum_{i=1}^{m} x_i^2}{n}}, \tag{72}$$

iar pentru seriile de frecvenţă (ponderate) avem

$$\bar{x}_p = \sqrt{\frac{\sum_{i=1}^{m} x_i^2 \cdot n_i}{\sum_{i=1}^{m} n_i}}. \tag{73}$$

Exemplul 4. Fie datele din exemplul 2. Să se determine salariul mediu al angajaților, aplicând media pătratică.

Rezolvare. Pentru comoditate vom dezvolta tabelul 17 în următorul tabel:

Tabelul. 19. Salariul angajaților întreprinderii "Ionel"

Salariul (lei)	Numărul de angajați (n_i)	x_i	${x_i}^2$	$\sum_{i=1}^{m} x_i^2 \cdot n_i$
[3500,4500)	10	4000	$16\cdot 10^6$	$160\cdot 10^6$
[4500,5500)	30	5000	$25\cdot 10^6$	$750\cdot 10^6$
[5500,6500)	100	6000	$36\cdot 10^6$	$3600\cdot 10^6$
[6500,7500)	150	7000	$49\cdot 10^6$	$7350\cdot 10^6$
[7500,8500)	250	8000	$64\cdot 10^6$	$16000\cdot 10^6$
[8500,9500)	210	9000	$81\cdot 10^6$	$17010\cdot 10^6$
[9500,10500)	200	10000	$100\cdot 10^6$	$20000\cdot 10^6$
[10500,11500]	50	11000	$121\cdot 10^6$	$6050\cdot 10^6$
	1000			$70920\cdot 10^6$

Rezolvare. Aplicând formula (73), obținem

$$\bar{x}_p = \sqrt{\frac{70920 \cdot 10^6}{10^3}} \approx 8421{,}4\ (lei).$$

Răspuns: $\bar{x}_p \approx 8421{,}4\ (lei)$

Media geometrică se notează $\bar{x}_g$ și se bazează pe relaţia de produs a termenilor seriei.

În cazul seriilor simple avem

$$\bar{x}_g = \sqrt[n]{x_1 \cdot x_2 \cdot \ldots \cdot x_m} = \sqrt[n]{\prod_{i=1}^{m} x_i}, \tag{74}$$

iar pentru seriile de frecvenţă (ponderate) avem

$$\bar{x}_g = \sqrt[\sum_{i=1}^{m} n_i]{x_1^{n_1} \cdot x_2^{n_2} \cdot \ldots \cdot x_m^{n_m}} = \sqrt[\sum_{i=1}^{m} n_i]{\prod_{i=1}^{m} x_i^{n_i}}. \quad (75)$$

Prin logaritmare, media geometrică capătă o formă similară celei a mediei aritmetice, cu deosebirea că nu se aplică termenilor seriei ca atare, ci logaritmilor acestora.

Pentru serii simple

$$lg\bar{x}_g = \frac{\sum_{i=1}^{m} lgx_i}{n}, \quad (76)$$

iar pentru seriile de frecvență (ponderate)

$$lg\bar{x}_g = \frac{\sum_{i=1}^{m} n_i \cdot lgx_i}{\sum_{i=1}^{m} n_i}. \quad (77)$$

Exemplul 5. Fie datele din exemplul 2. Să se determine salariul mediu, aplicând media geometrică, primit de angajații întreprinderii "Ionel".

Rezolvare. În cazul dat, putem aplica formula (75), sau formula (77), obținem

$$lg\bar{x}_g = (10 \cdot lg4000 + 30 \cdot lg5000 + 100 \cdot lg6000 +$$
$$+150 \cdot lg7000 + 250 \cdot lg8000 + 210 \cdot lg9000 + 200 \cdot lg10000 +$$
$$+50 \cdot lg11000)/1000 \approx 3{,}9098.$$

Din ultima relație, determinăm media geometrică

$$\bar{x}_g \approx 8124{,}56\ (lei).$$

Răspuns: $\bar{x}_g \approx 8124{,}56\ (lei)$

Relaţia dintre mediile de mai sus, este

$$\bar{x}_h \leq \bar{x}_g \leq \bar{x} \leq \bar{x}_p.$$

Modul sau *dominanta*, se notează M_0, este valoarea caracteristicii cea mai frecventă observată într-o distribuţie și se determină doar în cazul seriilor cu frecvenţe diferite.

I. În cazul variabilei discrete caracterizată de seria $(x_i n_i)$, $i = 1, .., n$, pentru a determina modul se parcurg următorii pași:

a) Se determină frecvența maximă a serei cercetate $n_i = n_{max}$.

Observația 1. *Dacă repartiția are un singur maxim, atunci se numește unimodală, dacă două maxime, atunci bimodală și dacă are mai mult de două maxime se numește polimodală.*

b) Se citește în dreptul frecvenței

$$max = M_0(x_i = M_0).$$

Cu alte cuvinte, aflarea modulului constă în analizarea grafică a seriei și determinarea valorii maxime.

II. În cazul variabilei continue, pentru determinarea modulului avem două moduri:

a) Determinarea modulului M_0, presupune realizarea următoarelor operații:

- aflarea frecvenței maxime $max(n_i, n_{max})$;
- citirea intervalului modal (x_{i-1}, x_i);
- determinarea modulului M_0 în baza formulei

$$M_0 = x_{inf} + l \cdot \frac{\Delta 1}{\Delta 1 + \Delta 2}, \quad (78)$$

unde x_{inf} – reprezintă limita inferioară a intervalului modal;

l – lungimea intervalului modal;

$\Delta 1$ – diferenţa dintre frecvenţa maximă și frecvența intervalului precedent

$$\Delta 1 = n_i - n_{i-1};$$

$\Delta 2$ – diferenţa dintre frecvenţa maximă şi frecvenţa intervalului următor

$$\Delta 2 = (n_i - n_{i+1}).$$

b) Determinarea grafică, care presupune construirea histogramei corespunzătoare seriilor de intervale. Modulul M_0 se află prin metoda diagonalelor ce corespunde cu punctul de intersecţie al diagonalelor ce unește coloana cu înălţimea maximă şi coloanele alăturate.

Mediana se notează Me și se defineşte ca acea valoare a caracteristicii unei serii ordonate crescător sau descrescător, până la care, şi peste care, sunt distribuite în număr egal, unităţile colectivităţii observate, jumătate din unităţi au valori mai mari decât mediana, iar altă jumătate au valori mai mici.

Definirea medianei Me începe de la idea de descompunere a unei distribuţii în 2 părţi: 50% din efectiv, înainte de valoarea medianei și 50% din efectiv, după valoarea medianei.

Ca urmare locul medianei corespunde valorii U^{Me}, denumită *unitate mediană*, care se calculează după relaţia [24]:

$$U^{Me} = \frac{\sum_{i=1}^{n} n_i}{2}, \quad (79)$$

iar în cazul când $n < 100$ unitate mediană se determină în baza formulei

$$U^{Me} = \frac{\sum_{i=1}^{n} n_i + 1}{2}. \quad (80)$$

Determinarea medianei Me se realizează în mod diferit, în funcţie de tipul seriei cercetate:

1. Pentru o serie simplă cu număr impar de termini, aflarea medianei Me constă în găsirea termenului central al seriei:

Exemplul 6. Să se determine mediana și unitatea mediană definită de șirul: 5 10 20 25 35.

Rezolvare. Evident $Me = 20$, iar unitatea mediană (80), este

$$U^{Me} = \frac{5+1}{2} = 3.$$

Răspuns: $Me = 20, U^{Me} = 3$

2. Pentru o serie simplă cu număr par de termini, aflarea medianei Me constă în calculul mediei aritmetice simple a celor doi termeni centrali ai seriei.

Exemplul 7. Să se determine mediana și unitatea mediană definită de șirul: 5 10 20 25 35 42.

Rezolvare. Evident

$$Me = \frac{20+25}{2} = 22{,}5,$$

iar unitatea mediană (80), este

$$U^{Me} = \frac{6+1}{2} = 3{,}5.$$

Răspuns: $Me = 22{,}5;\ U^{Me} = 3{,}5$

3. În cazul caracteristicilor discrete, pentru o serie de frecvenţă, determinarea medianei Me presupune efectuarea următoarelor operaţii:

a) Aflarea frecvenţei cumulate

$$N_i = \sum_{h=1}^{i} n_h.$$

b) Determinarea unității mediane U^{Me} şi aflarea locului ei în şirul frecvenţelor cumulate, respectând condiţia: $N_i \geq U^{Me}$.

c) Se determină nivelul caracteristicii egal cu mediana Me, în dreptul frecvenţelor cumulate egale sau mai mare ca unitatea mediană U^{Me}.

4. În cazul caracteristicii continue, pentru o serie cu frecvenţă, determinarea medianei Me presupune efectuarea următoarelor operaţii:

a) Aflarea şirului frecvenţelor cumulate.

b) Determinarea unităţii mediane şi găsirea locului ei în şirul frecvenţelor cumulate respectând condiţia

$$N_i \geq U^{Me}.$$

c) Determinarea intervalului medianei în dreptul $N_i \geq U^{Me}$.

d) Determinarea medianei Me după formula

$$Me = x_{inf} + l \cdot \frac{U^{Me} - N_{i-1}}{n_{Me}}, \tag{81}$$

unde x_{inf} – reprezintă limita inferioară a intervalului median;

l – lungimea intervalului;

N_{i-1} – frecvenţa cumulată anterior intervalului median;

n_{Me} – frecvenţa intervalului Me.

e) Determinarea medianei Me după modelul grafic.

Exemplul 8. Fie datele din exemplul 2. Să se determine nivelul mediu cu ajutorul medianei.

Rezolvare. Pentru comoditate vom dezvolta tabelul 17, adăugând o coloană cu frecvențele cumulate (tabelul 20).

Tabelul. 20. Salariul angajaților întreprinderii "Ionel" și frecvența cumulată

Salariul (lei)	Numărul de angajați (n_i)	Frecvența cumulată (N_i)
[3500,4500)	10	10
[4500,5500)	30	40
[5500,6500)	100	140
[6500,7500)	150	290
[7500,8500)	250	540
[8500,9500)	210	750
[9500,10500)	200	950
[10500,11500]	50	1000
	1000	

Rezolvare. Aplicând formula (79), obținem

$$U^{Me} = \frac{1000}{2} = 500.$$

Primul interval pentru care

$$U^{Me} < Fc_i$$

este

$$[7500{,}8500).$$

Determinăm mediana (81)

$$Me = 7500 + 1000 \cdot \frac{500 - 290}{250} = 8340\ (lei).$$

Rezultă că jumătate din angajaţi obţin un salariu de până la 8340 lei, iar cealaltă jumătate obţin salarii de peste 8340 lei.

Răspuns: $Me = 8340\ \ (lei)$

Exerciții și probleme propuse

1. Să se determine mediile simestriale personale, pentru anul precedent de studii.

2. La examenul național de bacalaureat 2021, s-au obținut următoarele medii (tabelul 20):

Tabelul. 20. Rezultatele la examenul național de bacalaureat 2021

Mediile liceelor la BAC 2021 (grupate pe intervale)	Numărul licee (n_i)
[5,6)	6
[6,7)	54
[7,8)	219
[8,9]	17

Să se dtermine media aritmetică, geometrică, pătratică, armonică și nivelul mediu cu ajutorul medianei, a rezultatelor obținute.

3. *Să se determine mediana și unitatea mediană definită de șirurile:*

a) 5, 17, 22, 25, 29, 33, 38, 45;

b) 16, 34, 47, 59, 66, 78, 88, 92, 99.

§3.3.2. Indicatorii de variație sau împrăștiere

Indicatorii de variaţie permit separarea acţiunii cauzelor esenţiale de cea a cauzelor întâmplătoare şi oferă posibilitatea identificării modului în care acţionează factorii esenţiali de la o grupă la alta. Dacă avem de caracterizat o serie de repartiţie unidimensională, vom calcula indicatorii de variaţie în raport cu media seriei, sau dacă avem o serie de repartiţie bidimensională, putem calcula indicatorii de variaţie în interiorul fiecărei serii şi indicatorii de variaţie între serii [24].

Indicatorii simpli ai variaţiei servesc pentru a caracteriza gradul de împrăştiere a unităţilor colectivităţii cercetate faţă de medie ($\bar{x}$), sau faţă de o anumită valoare din serie. Acești indicatori se pot exprima în unităţi absolute (kg, lei, metri, etc), aceleaşi ca şi ale caracteristicii studiate, cât şi în mărimi relative (%) calculate în raport cu media.

Indicatorii simpli sunt:

1) *Amplitudinea absolută a variaţiei* se obţine ca diferenţa dintre valoarea maximă (x_{max}) şi valoarea minimă (x_{min}) a seriei

$$A = x_{max} - x_{min}. \qquad (82)$$

În cazul unor serii de distribuţie pe intervale, amplitudinea se determină ca diferenţa dintre limita superioară a ultimului interval şi limita inferioară a primului interval.

2) *Amplitudinea relativă a variaţiei* se calculează ca raportul dintre amplitudinea absolută şi media aritmetică, exprimându-se procentual

$$A_r(\%) \ = \frac{A}{\bar{x}} \times 100. \tag{83}$$

3) *Abaterile individuale absolute* se calculează ca diferenţa dintre fiecare valoare înregistrată şi media aritmetică a seriei

$$d_i \ = \ x_i - \bar{x}\,. \tag{84}$$

4) *Abaterile individuale relative* se calculează ca raportul dintre abaterile individuale absolute şi media aritmetică a caracteristicii studiate, exprimându-se procentual

$$d_i\% = \frac{x_i - \bar{x}}{\bar{x}} \times 100. \tag{85}$$

Acesta abateri pot fi pozitive sau negative în funcţie de mărimea fiecărui termen faţă de media lor. Indicatorii simpli ai variaţiei nu pot exprima şi caracteriza întreaga variaţie a caracteristicii studiate. De aceea, este necesar să se calculeze indicatorii sintetici ai variaţiei.

Indicatorii sintetici ai variaţiei exprimă, în mod sintetic, împrăştierea tuturor nivelurilor individuale ale unei caracteristici faţă de nivelul lor mediu. În calitate de indicatori sintetici, statistica calculează:

1) *Abaterea medie liniară* ($\bar{d}$) care indică variaţia medie în plus şi în minus de la valoarea medie a distribuţiei şi este cu atât mai mică cu cât valorile sunt mai grupate în jurul mediei și se determină

a) pentru seriile simple avem

$$\bar{d} = \frac{\sum_{i=1}^{n}|d_i|}{n}. \tag{86}$$

b) pentru seriile de frecvenţă (ponderate), avem

$$\bar{d} = \frac{\sum_{i=1}^{n}|d_i| \cdot n_i}{\sum_{i=1}^{n} n_i}. \tag{87}$$

2) *Dispersia* (σ^2), pentru seriile simple se calculează după formula

$$\sigma^2 = \frac{\sum_{i=1}^{n} {d_i}^2}{n}, \tag{88}$$

iar pentru seriile de frecvenţă (ponderate)

$$\sigma^2 = \frac{\sum_{i=1}^{n} {d_i}^2 \cdot n_i}{\sum_{i=1}^{n} n_i}. \tag{89}$$

Dispersia este un indicator abstract, nu are formă concretă de exprimare şi arată modul în care valorile caracteristicii gravitează în jurul mediei. Ea măsoară variaţia totală a caracteristicii studiate datorită cauzelor esenţiale şi întâmplătoare.

3) *Abaterea medie pătratică*

$$\overline{\sigma} = \sqrt{\sigma^2}. \qquad (90)$$

4) *Coeficientul de variaţie* (υ), se calculează ca raport dintre abaterea medie pătratică sau abaterea medie liniară şi nivelul mediu şi arată câte unităţi din abaterea medie liniară sau pătratică revin la 100 unităţi de medii

$$\upsilon = \frac{\sigma}{\bar{x}},$$

sau

$$\upsilon = \frac{\bar{d}}{\bar{x}} \cdot 100. \qquad (91)$$

Coeficientul de variaţie ia valori de la $1\% - 100\%$, iar când tinde spre 0 (zero), se consideră o variaţie slabă şi deci este o colectivitate omogenă, în acest caz media are un grad de reprezentativitate ridicat. Cu cât coeficientului de variaţie tinde spre 100%, cu atât variaţia este mai intensă, iar media are un grad de reprezentativitate mai scăzut.

Ca urmare coeficientul de variaţie poate fi folosit ca un test de semnificaţie a reprezentativităţii mediei considerându-se următoarele praguri de semnificaţie [24]:

1. $0 < \upsilon < 17\%$ media este strict reprezentativă, iar colectivitatea este omogenă;

2. $17\% < \upsilon < 35\%$ media este moderat reprezentativă, iar colectivitatea este omogenă;

3. $35\% < \upsilon < 50\%$ media este reprezentativă în sens larg, iar colectivitatea este neomogenă;

4. $\upsilon > 50\%$ media este nereprezentativă.

Exemplul 1. Considerăm datele din exemplul 2 (§3.3.1). Să se determine indicatorii simpli și sintetici ai variației.

Rezolvare. Amplitudinea absolută a variaţiei (82), este

$$A = x_{max} - x_{min} = 11500 - 3500 = 8000.$$

Cunoaștem, din exemplul 2 (§3.3.1), media aritmetică $\bar{x} = 8280\ (lei)$, deci amplitudinea relativă a variaţiei (83), este

$$A_r(\%) = \frac{A}{\bar{x}} \times 100 = \frac{8000}{8280} \times 100 \approx 96{,}62\%.$$

Câmpul de variaţie al salariului muncitorilor este de 8000 lei, aceasta reprezentând 96,62% din salariul mediu al angajaților.

Abaterile individuale absolute (84), sunt

$$d_i = x_i - \bar{x} \rightarrow \begin{cases} x_1 - \bar{x} = 4000 - 8280 = -4280, \\ x_2 - \bar{x} = 5000 - 8280 = -3280, \\ x_3 - \bar{x} = 6000 - 8280 = -2280, \\ x_4 - \bar{x} = 7000 - 8280 = -1280, \\ x_5 - \bar{x} = 8000 - 8280 = -280, \\ x_6 - \bar{x} = 9000 - 8280 = 720, \\ x_7 - \bar{x} = 10000 - 8280 = 1720, \\ x_8 - \bar{x} = 11000 - 8280 = 2720. \end{cases}$$

Abaterile individuale relative (85), sunt

$$d_1 = \frac{-4280}{8280} \times 100 \approx -51{,}69; \quad d_2 = \frac{-3280}{8280} \times 100 \approx -39{,}61;$$

$$d_3 = \frac{-2280}{8280} \times 100 \approx -27{,}54; \quad d_4 = \frac{-1280}{8280} \times 100 \approx -15{,}46;$$

$$d_5 = \frac{-280}{8280} \times 100 \approx -3{,}38; \quad d_6 = \frac{720}{8280} \times 100 \approx 8{,}70;$$

$$d_7 = \frac{1720}{8280} \times 100 \approx 20{,}77; \quad d_8 = \frac{2720}{8280} \times 100 \approx 32{,}85.$$

Pentru a determina indicatorii sintetici ai variației, vom dezvolta tabelul 17 și obținem

Tabelul. 21. Salariul angajaților întreprinderii "Ionel"

Salariul (lei)	Numărul de angajați (n_i)	$\|d_i\|$	$\|d_i\| \cdot n_i$	d_i^2	$d_i^2 \cdot n_i$
[3500,4500)	10	4280	42 800	18318400	183184000
[4500,5500)	30	3280	98 400	10758400	322752000
[5500,6500)	100	2280	228 000	5198400	519840000
[6500,7500)	150	1280	192 000	1638400	245760000
[7500,8500)	250	280	70 000	78400	19600000
[8500,9500)	210	720	151 200	518400	108864000
[9500,10500)	200	1720	344 000	2958400	591680000
[10500,11500]	50	2720	136 000	7398400	369920000

Aplicând formula (87), obținem abaterea medie liniară

$$\bar{d} = \frac{\sum_{i=1}^{n} |d_i| \cdot n_i}{\sum_{i=1}^{n} n_i} = \frac{1\,262\,400}{1\,000} = 1262{,}4.$$

Determinăm intervalul mediu de variație

$$\bar{x} \pm \bar{d} \Rightarrow \begin{cases} \bar{x} + \bar{d} = 8280 + 1262{,}4 = 9542{,}4; \\ \bar{x} - \bar{d} = 8280 - 1262{,}4 = 7018{,}4. \end{cases}$$

Determinăm dispersia (89)

$$\sigma^2 = \frac{2361600000}{1000} = 2361600.$$

Abaterea medie pătratică (90), este

$$\overline{\sigma} = \sqrt{2361600} = 240\sqrt{41}.$$

Coeficientul de variaţie (91),

$$\upsilon = \frac{1262,4}{8280} \cdot 100 \approx 15,25.$$

Deoarece $0 < \upsilon < 17\%,$ rezultă că media este strict reprezentativă, iar colectivitatea este omogenă.

Exerciții și probleme propuse

1. *Să se dtermine indicatorii simpli și sintetici ai variației pentru datele din exemplul 2 (§3.3.1, exerciții și probleme propuse).*
2. *La o probă sportivă de aruncare a unei greutăți în lungime, 50 de elevi au obținut următoarele rezultate (în metri): 15,0; 16,8; 25,6; 35,0; 26,5; 18,5; 17,3; 21,9; 28,0; 23,4; 31,1; 27,9; 25,4; 22,8; 29,3; 31,2; 28,6; 33,5; 19,3; 21,7; 16,8; 28,6; 30,4; 31,7; 28,2; 26,5; 22,2; 31,6; 22,7; 24,8; 26,8; 27,1; 29,2; 16,9; 28,4; 33,0; 29,7; 27,8; 29,6; 33,4; 31,8; 29,6; 31,2; 28,9; 34,0; 26,1; 27,8; 31,4; 26,4 și 28,1. Să se deremine:*
   a) *Amplitudinea absolută și relativă a variaţiei.*
   b) *Abaterile individuale absolute și relative.*
   c) *Abaterea medie liniară.*
   d) *Abaterea medie pătratică.*

§3.3.2. Indicatorii asimetriei și boltirii

În analiza seriilor de distribuţie unidimensională şi unimodale, un interes deosebit prezintă cunoaşterea formei distribuţiei statistice care poate fi apreciată cu ajutorul a 2 categorii de valori tipice:

a) indicatorii de asimetrie, care dau informaţii asupra modului de reprezentare a frecvenţelor de o parte sau alta a valorii centrale a unei serii;

b) indicatorii de boltire care exprimă măsura aglomerării frecvenţelor în zona centrală pe lângă medie.

Asimetria reprezintă o deviaţie de la forma simetrică de distribuţie (observaţiile exprimate prin frecvenţele lor sunt distribuite identic pe de o parte şi de alta a valorii centrale). Ca valori centrale, pentru aprecierea asimetriei sunt folosite: media aritmetică, modul şi mediana.

Indicatorii de asimetrie pot fi calculaţi atât în mărime absolută cât şi în mărime relativă.

Pentru determinarea tipului de asimetrie se utilizează metodele elementare: metoda grafică, momentul centrat de ordinul 3, coeficientul de asimetrie al lui Pearson și al lui Fisher.

➢ ***Metoda grafică***, pentru determinarea gradului de asimetrie se analizează indicatorii tendinţei centrale: media, mediana şi modul. Astfel, în funcţie de raportul dintre aceşti indicatori, avem:

- $x = Me = Mo$ - serie simetrică (fig. 30.a);
- $x < Me < Mo$ - serie cu asimetrie spre stânga (negativă) (fig. 30.b);
- $x > Me > Mo$ - serie cu asimetrie spre dreapta (pozitivă) (fig. 30.c).

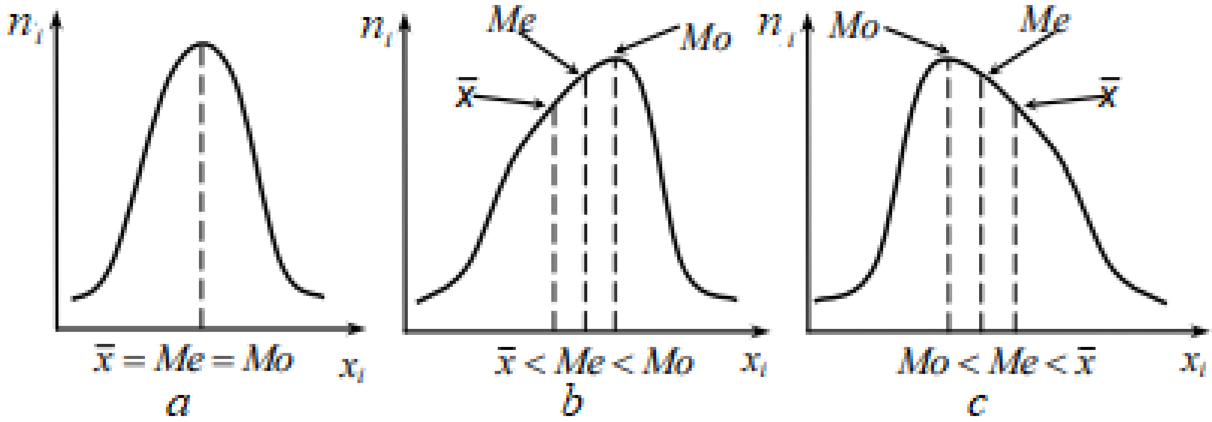


Fig. 30. Serii de repartiţie: a) simetrică; b) cu asimetrie spre stânga (negativă); c) cu asimetrie spre dreapta (pozitivă)

➢ ***Momentul centrat de ordinul 3***:

$$\mu_3 = \frac{\sum(x_i - \bar{x})^3 \cdot n_i}{\sum n_i}. \tag{92}$$

În funcţie de valoarea lui μ_3, avem:

- $\mu_3 = 0$ - serie simetrică;
- $\mu_3 < 0$ - serie cu asimetrie spre stânga (negativă);
- $\mu_3 > 0$ serie cu asimetrie spre dreapta (pozitivă).

Pentru măsurarea statistică a asimetriei se utilizează coeficientul de asimetrie al lui Pearson şi coeficientul lui Fisher.

➢ ***Coeficientul de asimetrie al lui Pearson,*** este cel mai des utilizat și reprezintă un indicator pentru determinarea asimetriei şi se obţine pe baza relaţiei următoare:

$$C_{as} = \frac{\bar{x} - Mo}{\sigma}. \tag{93}$$

Valoarea acestui indicator este abstractă, însă nu este lipsită de semnificaţie și oferă informaţii despre sensul asimetriei, dar și despre intensitatea acesteia. C_{as} ia valori din intervalul $(-1,1)$, dar pentru seriile de repartiţie moderat asimetrice, C_{as} ia valori din intervalul $[-0,3; 0,3]$.

Semnul indicatorului arată sensul asimetriei, și anume:

- $C_{as} = 0$ - serie simetrică;
- $C_{as} < 0$ - serie cu asimetrie spre stânga (negativă);
- $C_{as} > 0$ - serie cu asimetrie spre dreapta (pozitivă).

Dacă seriile sunt uşor asimetrice și se bazează pe un număr mare de cazuri observate, și se verifică relaţia

$$Mo \approx \bar{x} - 3(\bar{x} - Me),$$

se poate folosi un alt coeficient de asimetrie, care se determină în baza formulei

$$C_{as}{}^{*} = \frac{3(\bar{x} - Me)}{\sigma}.$$

Coeficientul $C_{as}{}^{*}$ ia valori din intervalul $(-3,3)$ şi va prezenta un grad mai mare de simetrie cu cât se va apropia mai mult de 0.

Pearson a mai propus un alt coeficient de asimetrie bazat pe momentele centrate de ordinul 2 şi 3:

$$\beta_1 = \frac{(\mu_3)^2}{(\mu_2)^3}. \tag{94}$$

Observăm că $\beta_1 \geq 0$. Acest indicator nu se utilizează pentru aprecierea sensului asimetriei, dar avem următoarele cazuri:

- $\beta_1 = 0$ - serie simetrică;
- $\beta_1 > 0$ - serie cu asimetrie (spre dreapta sau spre stânga).

➢ ***Coeficientul lui Fisher***, se determină în baza formulei

$$\gamma_1 = \sqrt{\beta_1}. \tag{95}$$

Exemplul 1. Să se analizeze asimetria serie din exemplul 2 (§3.3.1).

Rezolvare. Determinăm

$$M_0 = 7500 + 1000 \cdot \frac{100}{140} = 8214\frac{2}{7},$$

iar

$$C_{as} = \frac{8280 - 8214\frac{2}{7}}{240\sqrt{41}} = \frac{65\frac{5}{7}}{240\sqrt{41}} \approx 0{,}043.$$

Deci, avem o serie cu asimetrie spre dreapta (pozitivă)

Boltirea (aplanarea) se defineşte prin raportarea unei distribuţii empirice la distribuţia normală sub aspectul variaţiei variabile x_i şi a frecvenţei relative f_i. Boltirea apare când distribuţia prezintă o variaţie slabă a variaţiei x_i însoţită de o variaţie puternică a frecvenţei relative.

Boltirea se măsoară cu ajutorul coeficienţilor de boltire:

- ***coeficientul de boltire Pearson,*** se determină pe baza momentelor centrate

$$\beta_2 = \frac{\mu_4}{\mu_2^2}, \quad (96)$$

unde μ_2 și μ_4 sunt momentele centrale de ordinal 2 și 4.

Pentru distribuţia normală coeficientul de boltire ia valoarea β_2 =3, iar distribuţia este mezocurtică.

Dacă $\beta_2 > 3$, distribuţia este leptocurtică.

Dacă $\beta_2 < 3$ distribuţia este platicurtică.

- ***coeficientul de boltire Fisher*** măsoară excesul faţă de boltirea unei distribuţii normale Gauss-Laplace, ştiind că pentru o distribuţie normală $\beta_2 = 3$, gradul de exces se determină în baza formulei

$$\gamma_2 = \beta_2 - 3.$$

Dacă $\beta_2 = 3$, atunci $\gamma_2 = 0$ și avem o distribuţie mezocurtică.

Dacă $\gamma_2 > 0$, atunci avem o distribuţie leptocurtică.

Dacă $\gamma_2 < 0$, atunci avem o distribuţie platicurtică.

Exemplul 2. Să se analizeze boltirea serie din exemplul 2 (§3.3.1) în baza coeficientului de boltire Person.

Rezolvare. Determinăm momentul centrat de ordinal 4

$$\mu_4 = 0,$$

deci $\beta_2 = 0 < 3$ distribuţia este platicurtică.

Exerciții și probleme propuse

Situaţia salariului lunar obţinut de angajaţii unei întreprinderi este prezentată în tabelul 22:

Tabelul. 22. Salariul lunar al angajaţilor

Mediile liceelor la BAC 2021 (grupate pe intervale)	Numărul licee (n_i)
[4000,5000)	10
[5000,6000)	25
[6000,7000)	30
[7000,8000)	20
[8000,9000]	15

Să se determine:

a) *Coeficientul de asimetrie.*
b) *Coeficientul de boltire Pearson.*
c) *Coeficientul de boltire Fisher.*

§3.4. Cercetarea prin sondaj

Obiectul statisticii clasice reprezintă obţinerea datelor cu ajutorul observaţiei totale, în timp ce obţinerea informaţiilor pe baza cercetării parţiale, adică prin utilizarea unui eşantion din populaţia statistică, este studiată de către *statistica inferenţială* sau *cercetarea prin sondaj [26].*

Cercetarea prin sondaj vizează cunoaşterea unei populaţii statistice pe baza observaţiilor făcute asupra unuia sau mai multor eşantioane reprezentative extrase din aceasta, şi constă în estimarea unor caracteristici ale repartiţiei statistice, la verificarea unor ipoteze privind legăturile dintre fenomene, în previzionarea evoluţiei viitoare a acestora etc.

Cercetarea parţială are drept scop estimarea, în baza rezultatelor prelucrării datelor obţinute la nivel de eşantion, folosind principiile teoriei probabilităţilor, a parametrilor necunoscuți ai populaţiei statistice studiate.

În acest sens, se culeg şi se prelucrează datele aferente unităţilor statistice incluse în eşantionul studiat, obţinându-se diverşi indicatori sau parametri statistici, care descriu în detaliu respectivul eşantion.

În cazul sondajului, volumul colectivității generale se notează cu N, în cazul unităților simple cu R, iar dacă s-au înregistrat și variabile alternalive, atunci volumul se notează cu M.

Colectivitatea de selecție sau eșantionul reprezintă acea parte a populației generale de la care urmează să fie culese datele în scopul extinderii apoi asupra întregii populații. Volumul eșantionului (selecției) se notează respectiv cu n, r și m.

Un eșantion este considerat reprezentativ, atunci când reproduce în structura sa aceeași structură pe care o reprezintă întreaga populație.

Teoria și practica statistică demonstrează că asigurarea reprezentativității presupune respectarea cu strictețe a următoarelor condiții:

1) Includerea în eșantion a unităților, în mod obiectiv, fără acordarea unor proprietăți. Fiecare unitatea fiind extrasă, trebuie să corespundă unei probabilități diferite de 0.

2) Eșantionul stabilit trebuie să fie suficient de mare ca să permită redarea trăsăturilor esențiale ale populației totale, care v-a permite ulterior obținerea unor indicatori cu grad mai mare de stabilitate.

3) Includerea fiecărei unități trebuie efectuată independent una de alta.

Reprezentativitatea unui eșantion depinde în primul rând de alegerea corectă a procedeelor și tipurilor de selecție.

§3.4.1. Procedee de selecție a unui eșantion

În practică, la formarea eșantionului se utilizează mai multe procedee cunoscute sub numele de:

- *selecții aleatoare,*
- *selecții dirijate,*
- *selecții mixte.*

Folosirea selecției aleatoare exclude orice intervenție subiectivă în alegerea eșantionului, acest obiectiv poate fi realizat numai dacă elementul se selectează în mod aleator, dar cu condiția că fiecare element are șanse egale.

În selecțiile dirijate alegerea unităților se face de către persoanele care „culeg" datele.

Selecția mixtă combină principiile sondajelor aleatoare cu cele dirijate. În acest caz, este necesar mai întâi să se împartă populația în grupe, după o anumită caracteristică, apoi să se extragă în mod aleatoriu câte un reprezentant din fiecare grup.

O altă clasificare a procedeelor este

- *Procedeul tragerii la sorț,* care constă în extragerea dintr-o urnă a unei bile sau alte obiecte identice, reprezentând fiecare o unitate a populației, iar extragerea se face în 2 variante: procedeul selecției repetate și nerepetate. În cazul utilizării bilei întoarse, probabilitatea de includere în eșantion este constanta $\frac{1}{N}$. Tot timpul cât durează operația de eșantionare la sfârșit rămâne $(N-1)$ unități. În cazul bilei neîntoarse șansa fiecărei unități se schimbă în creștere $\frac{1}{N}, \frac{1}{N-1}, \frac{1}{N-n-1}$.

În cazul selecției nerepetate este exclusă probabilitatea extragerii unei unități de mai multe ori, micșorând astfel erorile și mărind astfel precizia. Procedeul tragerii la sorț se utilizează în cazul când populația cuprinde un număr mic de unități pentru care poate fi asigurate obiecte care pot fi incluse în urnă.

- *Procedeul tabelului cu numere aleatoare,* pentru utilizarea acestui procedeu este necesar de numerotat unitățile populației de la 1 la N, apoi se extrag n unități care formează eșantionul. De exemplu, se consideră o populație cu 900 de unități, se intenționează construirea unui eșantion din 10%. Se procedează în felul următor: se alege la întâmplare linia și coloana cu care începe selecția, numărul respectiv și numărul din 3 cifre se vor citi pe coloana respectivă și vor fi notate dacă sunt cuprinse între 1 și 900 și diferă de cele reținute anterior. Se vor enunța dacă sunt peste 900 și continuăm așa până când se vor obține 90 de unități.

Selecția mecanică presupune ordonarea colectivității generale după o caracteristică întâmplătoare prin care să se asigure includerea pe cât de posibil a unităților în bază de sondaj. În acest caz, eșantionul presupune stabilirea pasului de numărare care trebuie calculat ca raportul dintre volumul populației și volumul selecției. Cu ajutorul pasului de numărare se asigură împrăștierea populației în grupuri egale. Pentru a crea un eșeantion, se ia aleator o unitate din primul grup (aplicând procedeul tragerii la sorț) la care se adaugă pasul până la obținerea celor n unități a eșantionului.

Observația 1. Selecțiile dirijate și cele mixte se utilizează în special la sondajele de opinii la cercetările sociologice și în unele cazuri la studiul cererii de consum a populației.

Baza de selecție reprezintă, o listă ce include toate unitățile componente ale colectivității generale, lista este compusă după un criteriu care nu are nimic în comun cu ordinul de mărime al valorilor variabilelor înregistrate. Exemple de baze de sondaj: registrul auto, listele electorale, listele localităților, listele agenților economici, etc.

Cerințele principale pe care trebuie să le îndeplinească o bază de sondaj sunt:

- să fie adecvată scopului urmărit (să cuprindă întreaga populație);
- sa fie actuală;
- să fie exactă;
- sa fie ferită de orice repetiție;
- este convenabil ca să fie disponibilă doar pentru un singur centru de cercetare.

Notățiile folosite uzual pentru indicatorii colectivității generale și cei ai colectivității de selecție sunt date în tabelul 12 [27].

Tabelul 22. Indicatorii pentru medie și dispersie

Indicatorul	Caracteristica nealternativă		Caracteristica alternativă	
	Colectivitatea generală	Eșantionul	Colectivitatea generală	Eșantionul
Media	$\bar{x}_0 = \frac{\sum x_i \cdot N_i}{\sum N_i}$	$\bar{x} = \frac{\sum x_i \cdot n_i}{\sum n_i}$	$p = \frac{M}{N}$	$W = \frac{m}{n}$
Dispersia	$\sigma_0^2 = \frac{\sum (x_i - \bar{x})^2 \cdot N_i}{\sum N_i}$	$\sigma^2 = \frac{\sum (x_i - \bar{x}) \cdot n_i}{\sum n_i}$	$\sigma_p^2 = p(1-p)$	$\sigma_W^2 = W(1-W)$

Menționăm cele mai semnificative avantaje oferite de cercetarea prin sondaj, în cercetarea fenomenelor și proceselor în diverse domenii ale vieții economice și sociale:

- în majoritatea situațiilor, sondajul statistic, este singura alternativă la care se poate recurge și anume, atunci când cercetarea conduce la distrugerea elementelor. De exemplu, estimarea recoltelor înaintea recoltării; determinarea duratei de funcționare a unor piese, a rezistenței diferitor materiale,etc.;

- este mai operativ și mai ieftin, deoarece numărul unităților de la care se culeg datele, este semnificativ mai mic decât colectivitatea generală;
- permite cunoașterea mai completă, în sensul că în cazul unui număr mai mic de unități se poate folosi un program de observare mai amplu comparativ cu cel utilizat în cazul unei înregistrări care epuizează un subiect în întregime;
- erorile de înregistrare sunt mai mici și pot fi depistate mai ușor;
- poate fi utiizat ca mijloc de verificare a rezultatelor unei cercetări totale.

De asemenea se adaugă faptul că sondajul statistic, fiind o cercetare statistică parțială, oferă doar o estimare a parametrilor colectivității generale, deci rezultatele nu sunt determinări exacte.

Concluzionăm, sondajul statistic s-a impus în practica cercetării, din majoritatea domeniilor de activitate, datorită operativității cu care se obțin rezultatele, datorită costului informațiilor și nu în ultimul rând datorită faptului că oferă rezultate suficient de exacte despre colectivitatea studiată.

§3.4.2. Erorile sondajului statistic

La crearea unui sondaj statistic se urmărește, estimarea indicatorilor unei colectivități de mare amploare, pentru determinarea cărora nu este posibilă sau nu se justifică organizarea unei cercetări care epuizează un subiect în întregime, pornind de la indicatorii calculați pe baza datelor eșantionului [27].

Însă, oricât de corectă ar fi făcută eșantionarea, valorile rezultate din prelucrarea datelor corespunzătoare eșantionului se abat de la cele determinate pe baza datelor înregistrate pentru colectivitatea generală.

Erorile de sondaj (de selecție) se consideră diferențele care există între valorile oricărui indicator calculat pe baza datelor eșantionului și valorile aceluiași indicator determinate pe baza datelor corespunzătoare colectivității generale. Sondajele statistice înregistrează două tiputi de erori:

- ✓ *erori de înregistrare*, care sunt comune tuturor tipurilor de observări statistice;
- ✓ *erori de reprezentativitate*, specifice cercetarii prin sondaj.

Erorile de înregistrare ce intervin în sondajele statistice sunt mai mici comparativ cu cele din înregistrările totale, deoarece volumul datelor înregistrate este semnificativ mai mic, iar culegerea datelor se realizează de un personal de specialitate.

Erorile de reprezentativitate sunt cele mai specifice sondajelor statistice și pot fi de două tipuri:

- erori sistematice;

- erori întâmplătoare.

Erorile de reprezentativitate sistematice pot fi evitate dacă se respectă întocmai principiile selecției prin înlăturarea cauzelor de producere. Principalele cauze care pot duce la apariția erorilor sistematice sunt:

- alegerea deliberată a unor unități considerate reprezentative;
- selectarea preferențială a acelor unități care să ducă la rezultatul dorit de cercetător;
- baze de sondaj incomplete;
- volumul redus al eșantionului.

Erorile de reprezentativitate întâmplătoare pot apărea, chiar dacă se respectă toate regulile sondajului statistic, deoarece prin numărul mic de unități care compun eșantionul nu se pot reproduce toate trăsăturile esențiale ale colectivității generale. Deși nu pot fi evitate, ele pot fi determinate anticipat, dacă selecția este probabilistică. Parametrii colectivității generale se estimează cu ajutorul indicatorilor obținuți la prelucrarea datelor eșantionului cu o anumită eroare întâmplătoare de reprezentativitate.

Eroarea de reprezentativitate se determină prin diferența dintre media eșantionului $(\bar{x})$ și media colectivității generale $(\bar{x}_0)$. Eșantion este reprezentativ, dacă eroarea ia valori din intervalul $\pm 5\%$, din care rezultă

$$\frac{\bar{x}-\bar{x}_0}{\bar{x}_0} \cdot 100 \leq |5\%|. \tag{97}$$

În acest caz, se presupune că se cunoaște media colectivității generale. În vederea verificării reprezentativității eșantionului, se recomandă extragerea a două eșantioane de volum diferit și compararea mediilor celor două eșantioane. În cazul când diferența nu este semnificativă, se consideră că oricare din cele două eșantioane poate fi folosite pentru estimarea parametrilor colectivității generale, în caz contrar se recomandă extragerea unui al treilea eșantion (diferit de primele două), volumul căruia este cel puțin egal cu volumul cumulat al primelor două eșantioane. Apoi, se va opta pentru unul din primele două eșantioane la care media se apropie mai mult de media celui de-al treilea eșantion.

Eroarea medie și eroarea limită. După cum am enunțat mai sus, eroarea de reprezentativitate este diferența dintre media eșantionului și media colectivității generale. Însă, dintr-o colectivitate generală de volum N se pot extrage succesiv mai multe eșantioane de același volum n.

În cazul sondajelor repetate, numărul eșantioanelor posibile de format este egal cu N^n, iar în cazul sondajelor nerepetate combinațiile sunt mai puține, deoarece aceeași unitate nu poate participa decât într-un singur eșanton și este egal cu C_N^n. Fiecare eșantion este definit de media $(\bar{x}_i)$ și dispersia (σ_i^2) care se calculează pentru caracteristica sau caracteristicile ce ne interesează. Deoarece mediile de selecție diferă între ele, deci și erorile de reprezentativitate $(\bar{x}_i - \bar{x}_0)$ reprectiv vor fi diferite de la un eșantion la altul.

Dacă se vor lua în considerare toate eșantioanele de un anumit volum n, atunci s-a arătat că mediile de selecție $(\bar{x}_i)$ se distribuie normal față de media care are frecvența cea mai mare de apariție. Însă dacă nu se cunoaște care din eșantioanele posibile s-a extras, nu se cunoaște eroarea de reprezentativitate corespunzătoare.

În acest caz, se aplică *eroarea medie de reprezentativitate* sau *media erorilor de reprezentativitate*.

Eroarea medie de reprezentativitate se calculează ca o abatere medie pătratică a tuturor mediilor de selecție de la media colectivității generale

$$\sigma_{\bar{x}} = \sqrt{\frac{\sum_{i=1}^{k}(\bar{x}_i - \bar{x}_0)^2 \cdot n_i}{\sum_{i=1}^{k} n_i}}, \tag{98}$$

unde k este numărul de eșantioane posibile, iar n_i este frecvența mediilor de selecție posibile.

Eroarea medie de reprezentativitate se poate calcula anticipat, pornind de la relația dintre dispersia colectivității generale (σ_0^2), dispersia mediilor de selecție, de la media colectivității generale $(\sigma_{\bar{x}}^2)$ și volumul eșantionului (n).

Eroarea medie de reprezentativitate se calculează cu ajutorul formulei

$$\sigma_{\bar{x}} = \sqrt{\frac{\sigma_0^2}{n}}, \tag{99}$$

ceea ce înseamnă că mărimea erorii este direct proporțională cu dispersia colectivității generale și invers proporțională, cu volumul eșantionului. Aplicarea ultimei relații presupune că se cunoaște dispersia colectivității generale, dintr-o cercetare totală anterioară, însă așa situații sunt foarte rar întâlnite în practica organizarii unui sondaj statistic.

În cazul când nu se cunoaște dispersia colectivității generale σ_0^2, se accepă ipoteza că dispersia eșantionului extras σ^2 poate caracteriza suficient de corect variația în cadrul colectivității generale, iar eroarea medie de reprezentativitate se determină în baza formulei

$$\sigma_{\bar{x}} = \sqrt{\frac{\sigma^2}{n-1}}. \tag{100}$$

Eroarea medie de reprezentativitate pentru sondajul nerepetat se determină în baza formulei

$$\sigma_{\bar{x}} = \sqrt{\frac{\sigma_0^2}{n}\left(1 - \frac{n}{N}\right)} \qquad (101)$$

și respectiv, dacă nu se cunoaște dispersia colectivității generale, atunci

$$\sigma_{\bar{x}} = \sqrt{\frac{\sigma^2}{n-1}\left(1 - \frac{n}{N}\right)}. \qquad (102)$$

În cazul unei variabile alternative, eroarea medie de reprezentativitate se determină în baza formulei

$$\sigma_p = \sqrt{\frac{p(1-p)}{n}\left(1 - \frac{1}{N}\right)} \qquad (103)$$

și respectiv

$$\sigma_W = \sqrt{\frac{W(1-W)}{n-1}\left(1 - \frac{1}{N}\right)}. \qquad (104)$$

Un eșantion este considerat reprezentativ, dacă abaterea medie de selecție de la media colectivității generale (eroarea de reprezentativitate), este cuprinsă între $\pm 5\%$.

Eroarea limită de reprezentativitate (Δ_x) se determină ca o abatere a mediei de selecție de la media colectivității generale garantată cu suma probabilitaților corespunzatoare limitelor intervalului de variație.

Pentru o variabilă numerică eroarea limită este

$$\Delta_x = z \cdot \sigma_{\bar{x}}, \qquad (105)$$

iar în cazul variabilei alternative

$$\Delta_x = z \cdot \sigma_W. \qquad (106)$$

Argumentul probabilitații z se obține din tabelul întocmit pentru funcția Gauss-Laplace și depinde de probabilitatea cu care se garantează rezultatele sondajului pentru care s-a optat. De exemplu, s-a optat pentru probabilitatea 0.95, atunci $z = 1.96$ și respectiv dacă probabilitatea care se folosește este 0,9973, atunci $z = 3$.

§3.4.3. Tipuri de sondaje folosite frecvent în practica statistică

În practica statistică se pot aplica mai multe tipuri de sondaje în funcție de gradul de omogenitate al colectivității studiate și de forma de organizare a acesteia. Se cunosc următoarele tipuri de sondaje:

- sondajul aleator (întâmplător) simplu;
- sondajul tipic (stratificat);
- sondajul de serii.

Fiecare tip de sondaj menționat poate fi folosit repetat sau nerepetat.

Sondajul aleator simplu se aplică în cazul când colectivitatea generală este omogenă. Pntru a forma un eșantion se extrag aleator unități simple, prin procedeul repetat sau nerepetat. Eroarea de reprezentativitate ($\sigma_{\bar{x}}$) și eroarea limită (Δ_x) se calculează pe baza formulelor din paragraful precedent.

Sondajul tipic (stratificat) este recomandat în situațiile când colectivitatea este neomogenă. În cazul dat, se separă unitațile simple pe straturi (grupe) mai omogene după o variabilă calitativă sau cantitativă. De exemplu, dacă colectivitatea generală este formată din totalitatea agenților economici, în vederea separării pe straturi (grupe, tipuri) se utilizează următoarele caracteristici: domeniul de activitate, numărul angajaților, cifra de afaceri etc.

Eșantionul este format prin extragerea din fiecare strat a unui număr de unități simple (subeșantioane, n_i), fapt ce conduce la o mai mare reprezentativitate, și ca atare la erori mai mici.

Eroarea de reprezentativitate și eroarea limita se determină tinând cont de variația la nivelul fiecarui strat (grup), care presupune determinarea dispersiei fiecarui grup sau strat (σ_1^2).

În acest caz, variația din toate grupele (straturile) se sintetizează prin media dispersiilor de grupă

$$\left(\bar{\sigma}^2 = \frac{\sum_{i=1}^{k} \sigma_i{}^2 \cdot n_i}{\sum_{i=1}^{k} n_i}\right). \tag{107}$$

Deoarece $\bar{\sigma}^2 < \sigma^2$, sondajul tipic înregistrează erori mai mici comparativ cu sondajul aleator simplu.

Formulele de calcul a erorii de reprezentativitate și a erorii limită de reprezentativitate, în cazul sondajului tipic repetat sunt

$$\sigma_{\bar{x}} = \sqrt{\frac{\sigma_0^2}{n}} \quad sau \quad \sigma_{\bar{x}} = \sqrt{\frac{\bar{\sigma}^2}{n-1}} \tag{108}$$

și

$$\Delta_x = z\sqrt{\frac{\sigma_0^2}{n}} \quad sau \quad \Delta_{\bar{x}} = z\sqrt{\frac{\bar{\sigma}^2}{n-1}}, \tag{109}$$

iar în cazul sondajului tipic nerepetat

$$\sigma_{\bar{x}} = \sqrt{\frac{\bar{\sigma}_0^2}{n}\left(1-\frac{n}{N}\right)} \quad sau \quad \sigma_{\bar{x}} = \sqrt{\frac{\bar{\sigma}^2}{n-1}\left(1-\frac{n}{N}\right)} \tag{110}$$

și

$$\Delta_x = z\sqrt{\frac{\bar{\sigma}_0^2}{n}\left(1-\frac{n}{N}\right)} \quad sau \quad \Delta_{\bar{x}} = z\sqrt{\frac{\bar{\sigma}^2}{n-1}\left(1-\frac{n}{N}\right)}. \qquad (111)$$

Selecția tipică este de două tipuri: proporțională și optimă.

Selecția tipică proporțională este acea selecție la care volumul subeșantioanelor diferă în raport cu ponderea pe care o are fiecare grupă în colectivitatea generală și se respectă proporția de selecție.

Selecția tipică optimă este selecția în care, la formarea eșantionului se ea în considerare ponderea grupelor în colectivitatea generală și mărimea variației din interiorul grupelor măsurate prin abaterea mediei pătratice.

Sondajul de serii se aplica în cazul când colectivitatea care este supusă studiului este formată din unități complexe (echipe de muncitori, gospodari, grupe de studiu), denumite serii. Pentru formarea eșantionului se extrage un anumit număr de unități complexe (serii), prin unul din procedeele menționate mai sus. Pentru fiecare serie se determină media, iar cu ajutorul acestora se determina media colectivității generale ($\bar{x}_0$) sau media eșantionului.

Deoarece nu se cunosc valorile pentru fiecare unitate simplă care compune seria, ci doar media seriei, la determinarea indicatorilor sondajului se folosește dispersia dintre grupe sau dintre medii

$$\delta^2 = \frac{\sum_{i=1}^{k}(\bar{x}_i - \bar{x})^2}{n}. \qquad (112)$$

Într-o colectivitate generală, numărul seriilor existente se notează de regula cu R, iar numărul seriilor care compun eșantionul, cu r.

Eroarea medie de reprezentativitate și eroarea limită pentru sondajul de serie repetate sunt

$$\sigma_{\bar{x}} = \sqrt{\frac{\delta^2}{r}} \quad și \quad \Delta_x = z\sqrt{\frac{\delta^2}{r}}, \qquad (113)$$

iar pentru sondajul de serie nerepetate sunt

$$\sigma_{\bar{x}} = \sqrt{\frac{\delta^2}{r}\cdot\frac{R-r}{R-1}} \quad și \quad \Delta_x = z\sqrt{\frac{\sigma^2}{r}\cdot\frac{R-r}{R-1}}. \qquad (114)$$

§3.4.4. Determinarea volumului unui eșantion

Realizarea unui sondaj statistic în vederea estimării indicatorilor colectivității generale, presupune să se decidă asupra mărimii eșantionului. Criteriile, în funcție de care se decide, privesc exactitatea cu care se estimează indicatorii colectivității generale, costurile realizării sondajului ș.a.

Volumul eșantionului se deduce în cazul fiecărui tip de sondaj, din formula erorii limită. Prin ridicarea la pătrat a formulei erorii limită (Δ_x) se deduce volumul eșantionului.

În cazul sondajului aleator simplu repetat, avem [27]

$$\Delta_x = z\sqrt{\frac{\sigma^2}{n}} => {\Delta_x}^2 = z^2\frac{\sigma^2}{n} \qquad (115)$$

deci

$$n = \frac{z^2\sigma^2}{{\Delta_x}^2},$$

în cazul sondajului aleator simplu nerepetat se începe de la relația

$$\Delta_x = z\sqrt{\frac{\sigma^2}{n-1}\left(1-\frac{n}{N}\right)}, \qquad (116)$$

iar dacă se renunță la corecția cu 1 din numitorul membrului drept se obține

$$n = \frac{z^2\sigma^2}{{\Delta_x}^2 + \frac{z^2\sigma^2}{N}}. \qquad (117)$$

Similar se deduc relațiile privind volumul eșantionului pentru celelalte tipuri de sondaje.

§3.4.5. Estimarea parametrilor colectivității generale

La organizarea unui sondaj statistic se urmarește, cel mai adesea, estimarea parametrilor colectivității generale. În acest scop se utilizează cel mai frecvent procedeul extinderii directe. La aplicarea acestui procedeu se estimează intervalul de încredere pentru media colectivității generale și limitele între care se va încadra nivelul totalizat al caracteristicii pe întreaga colectivitate ($\sum x_i$).

Estimarea parametrilor colectivității generale se bazează pe media eșantionului $\bar{x}$ și pe eroarea limită.

Media colectivității generale se estimează în baza formulei

$$\bar{x} - \Delta_x < \bar{x}_0 < \bar{x} + \Delta_x, \qquad (118)$$

iar limitele între care variază nivelul totalizat al caracteristicii în colectivitatea generală se estimează după formula

$$N(\bar{x} - \Delta_x) < \sum x_i < (\bar{x} + \Delta_x)N. \qquad (119)$$

CAPITOLUL III. IMPLEMENTAREA TEHNOLOGIILOR INFORMAȚIONALE LA REZOLVAREA UNOR PROBLEME DIN PROBABILITATE

Instruirea asistată de calculator reprzintă un instrument util şi eficient aplicat nu numai în învățământ, dar și toate domniile existente. În învățământ există divrse programe destinate învăţării printre care se numără aplicația Maple și mediul de programare Delphi, în care pot fi realizate programe, astfel încât să prezinte elevului/studentului, inclusiv și cadrelor didactice, o cantitate de informaţie, iar apoi, alternativ, să testeze modul de înţelegere şi însuşirea informaţiei respective.

Tradițional, cursul teoria probabilităților și statistica matematică studiază mai mult datele teoretice decât datele experimentale, deoarece ultimele necesită utilaje costisitoare, instrumente, timp, spațiu, etc. De aici și apare confuzia: de ce diferă probabilitatea teoretică și cea experimentală, cu excepția așa numitor *cazuri perfecte* (când coincide probabilitatea teoretică și cea experimentală)?

Este bine venit, ca la cursul teoria probabilităților și statistica matematică, și nu numai la el, să aplicăm unele simulări la calculator care să ne ofere posibilitatea realizării experimentelor, dar și realizarea rapidă a unor calcule necesare.

În aceast capitol vor fi prezentate câteva simulări a unor probleme din probabilitate în aplicația Maple, dar și în mediul de programare Delphi. Însă codul programul va fi prezentat succint (doar secvențele cele mai importante).

Problema 1.

Se aruncă o monedă de n ori. Să se determine probabilitatea căderii stemei.

Rezolvare. Cunoaștem că probabilitatea teoretică este

$$P(A) = \frac{1}{2},$$

însă prezintă interes probabilitatea experimentală. În acest caz, s-a creat o procedură în aplicația Maple care realizează aruncarea unei monede de *n* ori. Aceasta implică un ciclu repetitiv, de exemplu, instrucțiunea *for* care la rândul său va coține o condiție de verificare, pentru a determina dacă valoarea atribuită aleator variabilei a, este sau nu caz favorabil. În acest scop, s-a utilizat instrucțiunea if, care va stoca într-o variabilă contor c, numărul de cazuri favorabile pentru $a = 1$ (apariția stemei).

În rezultat, s-a obținut următoarea procedură:

```
> moneda := proc(n);
  local p, i , c, a;
  c := 0;
  for i from 1 to n do
  a := RandomTools[Generate](integer(range = 1 ..2));
  if (a = 1) then c := c + 1; fi;
  end do;
  p := evalf(c/n);
  return(p);
  end proc;
moneda := proc(n)
    local p, i, c, a;
    c := 0;
    for i to n do
        a := RandomTools[Generate](integer(range = 1 ..2)); if a = 1 then c := c + 1
        end if
    end do;
    p := evalf(c / n);
    return p
end proc
> exp1 := moneda(100)
                              exp1 := 0.5400000000
> exp1 := moneda(1000)
                              exp1 := 0.4850000000
```

Problema 2.

Se aruncă un zar de n ori. Să se determine probabilitățile căderii fiecărei fețe a zarului.

Rezolvare. Cunoaștem că probabilitatea teoretică este

$$P(A) = \frac{1}{6}.$$

Pentru determinarea probabilității experimentale s-a utilizat, ca și în problema precedentă, o procedură Maple, care conține un ciclu repetitiv *for*, în care, de n ori, i se atribuie unei variabile b, un număr aleator de 1 la 6 și în dependență de valoarea variabilei obținute, cu ajutorul instrucțiunii *if*, se verifică pentru care față este caz favorabil și se adună la variabila corespunzătoare c_i, dar poate fi utilizată și instrucțiunea *case*.

În rezultat, la realizrea a n experimente, calculatorul va returna probabilitatea căderii fiecărei fețe a zarului.

```
> zar := proc(nn);
  local p1, p2, p3, p4, p5, p6, j , b, c1, c2, c3, c4, c5, c6;
  c1 := 0; c2 := 0; c3 := 0; c4 := 0; c5 := 0; c6 := 0;
  for j from 1 to nn do
  b := RandomTools[Generate](integer(range = 1 ..6));
  if (b = 1) then c1 := c1 + 1; fi;
  if (b = 2) then c2 := c2 + 1; fi;
  if (b = 3) then c3 := c3 + 1; fi;
  if (b = 4) then c4 := c4 + 1; fi;
  if (b = 5) then c5 := c5 + 1; fi;
  if (b = 6) then c6 := c6 + 1; fi;
  end do;
```

$p1 := evalf\left(\frac{c1}{nn}\right); p2 := evalf\left(\frac{c2}{nn}\right); p3 := evalf\left(\frac{c3}{nn}\right);$

$p4 := evalf\left(\frac{c4}{nn}\right); p5 := evalf\left(\frac{c5}{nn}\right); \mathrm{p6} := evalf\left(\frac{c6}{nn}\right);$

```
  return(p1, p2, p3, p4, p5, p6);
  end proc;
zar := proc(nn)
    local p1, p2, p3, p4, p5, p6, j, b, c1, c2, c3, c4, c5, c6;
    c1 := 0;
    c2 := 0;
    c3 := 0;
    c4 := 0;
    c5 := 0;
    c6 := 0;
    for j to nn do
        b := RandomTools[Generate](integer(range = 1 ..6));
        if b = 1 then c1 := c1 + 1 end if;
        if b = 2 then c2 := c2 + 1 end if;
        if b = 3 then c3 := c3 + 1 end if;
        if b = 4 then c4 := c4 + 1 end if;
        if b = 5 then c5 := c5 + 1 end if;
        if b = 6 then c6 := c6 + 1 end if
    end do;
    p1 := evalf(c1/nn);
    p2 := evalf(c2/nn);
    p3 := evalf(c3/nn);
    p4 := evalf(c4/nn);
    p5 := evalf(c5/nn);
    p6 := evalf(c6/nn);
    return p1, p2, p3, p4, p5, p6
end proc
```

```
> exp2 := zar(100)
exp2 := 0.2100000000, 0.2000000000, 0.1500000000, 0.1500000000, 0.1700000000, 0.1200000000
> exp2 := zar(1000)
exp2 := 0.1710000000, 0.1680000000, 0.1800000000, 0.1390000000, 0.1670000000, 0.1750000000
> exp2 := zar(100000)
exp2 := 0.1673400000, 0.1667700000, 0.1659100000, 0.1685000000, 0.1661700000, 0.1653100000
```

Problema 3

Într-o urnă sunt 100 de bile numerotate de la 1 la 100, fără repetare. Se extrage, aleator o bilă. Să se determine probabilitatea că bila extrasă are imprimat, pe ea, un număr prim (exerciții propuse, exemplul 8(a), §1.3) [30].

Rezolvare. Este cunoscut că de la 1 la 100 sunt 25 de numere prime. Deci, în baza formulei (3), obținem

$$P(X) = \frac{25}{100} = 0{,}25.$$

Prezenta condiție a fost simulată în mediul de programare Delphi. Pe forma aplicației, a fost plasată o imagine *Image1*, care va încarca o imagine ce conține condiția problemei, dintr-un directoriu curent img

image1.Picture.LoadFromFile('img\1.jpg');

De asemenea, pe formă a fost plasat un *Button 1* (START) în care s-a programat experimentul problemei. La realizarea unui click pe butonul START, se va realiza *n* experimente, unde *n* este numărul introdus în *Edit1* (fig. 31)

n:=strtoint(edit1.Text);

Se selectează aleator de *n* ori câte un număr de la 1 la 100 și se verifică dacă numărul selectat este prim, atunci într-o variabilă contor *c*, îl numără ca un caz favorabil.

Pașii enumerați mai sus, se realizează prin următoarea secvență:

```
c:=0;
for j:=1 to n do
begin
i:=random(100);
if (i=2) or (i=3) or (i=5) or (i=7) or (i=11) or (i=13) or (i=17) or
(i=19) or (i=23) or (i=29) or (i=31) or (i=37) or (i=41) or (i=43) or
(i=47) or (i=53) or (i=59) or (i=61) or (i=67) or (i=71) or (i=73) or
(i=79) or (i=83) or (i=89) or (i=97)
then
c:=c+1;
end;
```

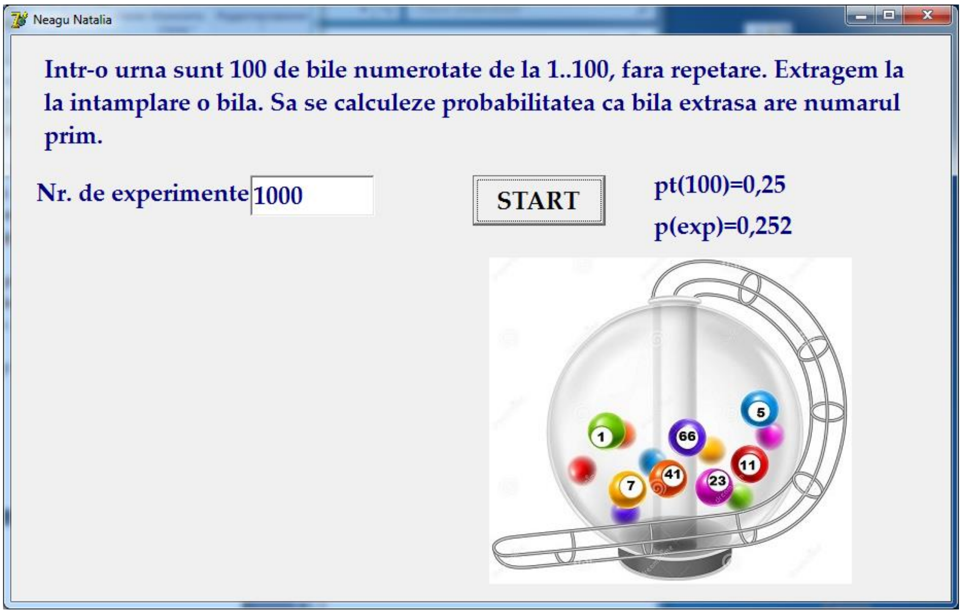


Fig. 31. Simularea experimentului din problema 3

Problema 4

Fie două ceruri concentrice C_1 și C_2 de raze $r_1 = 200$ și $r_2 = 100$ cu condiția $r_1 > r_2$. Să se determine probabilitatea că un punct luat la întâmplare din cercul C_1, să nu aparțină cercului C_2 (exemplul 3, §1.4) [28].

Rezolvare. Din punct de vedere teoretic, această problemă a fost rezolvată în exemplul 3, §1.4. Însă, în continuare vom prezenta succint simularea condiției problemei în mediul de programare Delphi (fig. 32).

La activarea soft-ului, se pornește un *Timer1*, care extrage numărul de experimente N, de pe formă, dintr-un *Edit1*

N:=strtoint(edit1.Text);

În această simulare este posibil de selectat valorile razelor cercurilor: manual și aleatoriu. După ce sunt selctate razele cercurilor concentrice C_1 și C_2 ($r_1 > r_2$), se construiesc două cercuri corespunzătoare datelor pe un *Paintbox1*

paintbox1.Canvas.Ellipse(x0-r1,y0-r1,x0+r1,y0+r1);

paintbox1.Canvas.Ellipse(x0-r2,y0-r2,x0+r2,y0+r2);

unde $(x0, y0)$ este centrul cercurilor, iar $r1$ și $r2$ sunt razele respective ale cercurilor. Apoi se activează un *Timer2* ce realizează *N* probe, adică aleatoriu se construiesc *N* puncte în interiorul cercului C_1

x:=random(paintbox1.width);

y:=random(paintbox1.Height);

d:=sqrt(sqr(x-x0)+sqr(y-y0));//distanța de la punctul luat aleatoriu pînă la centru

if d<r1 then

begin

```
if d>=r2 then
begin
sc1:=sc1+1;
paintbox1.Canvas.Brush.Color:=clred;
paintbox1.Canvas.Ellipse(x-2,y-2,x+2,y+2);
end;
if d<r2 then
begin
sc2:=sc2+1;
paintbox1.Canvas.Brush.Color:=clwhite;
paintbox1.Canvas.Ellipse(x-2,y-2,x+2,y+2);
end;
n:=n+1;
end;
```

În acelaşi timp se numără punctele din interiorul cercului C_2 şi din discul format de aceste cercuri, în dependență de distanța dintre punctul luat aleator şi centrul cercurilor. Dacă această distanță este $r_1 \geq d \geq r_2$, atunci punctul dat va fi numărat ca şi un caz favorabil

sc1 – variabilă contor ce numără punctele din disc;

sc2 – variabilă contor ce numără punctele din cercul C_2.

După ce s-au construit *N* puncte *Timer2* se opreşte, iar la un experiment nou se activează şi respectiv la finisare se dezactivează.

Într-un Label, pe formă, se afişează probabilitatea teoretică calculată în baza formulei (3), iar în altul probabilitatea experimentală (fig. 32)

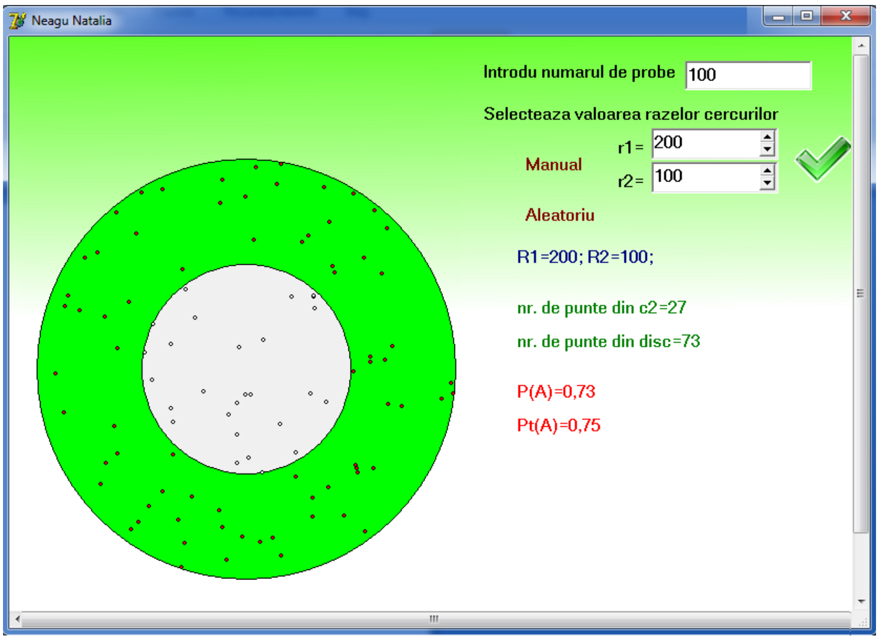


Fig. 32. Realizarea unui experiment din 100 probe

Realizând un alt experiment, dintr-un număr de probe mai mare (fig. 33), observăm că probabilitatea experimentală se apropie de probabilitatea teoretică.

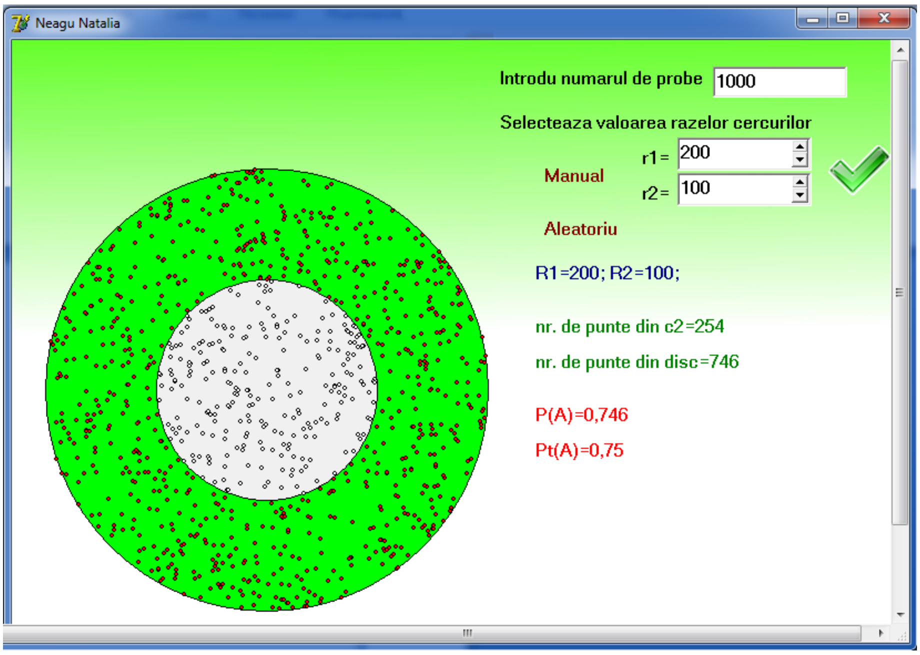


Fig. 33. Realizarea unui experiment din 1000 probe

Problema 5

Andrei și Ion au decis să vină între orele 18^{00} și 19^{00} la Librăria Făt-Frumos, și să se aștepte 15 min. Determinați probabilitatea întâlnirii celor 2 prieteni (exemplul 4, §1.4) [29].

Rezolvare. Din punct de vedere teoretic, această problemă a fost rezolvată în exemplul 4, §1.4. Însă condiția acestei probleme, a fost simulată pe calculator în mediul de programare Delphi (fig. 34).

În această simulare are posibilitatea de a modifica numărul de minute de așteptare și viteza de realizare a experimentului (*Interval*-ul componentei *Timer1*), care se determină din poziția componentei *Scroolbar1*

timer1.Interval:=round(1000 div scrollbar1.Position);

În partea stângă a formei se construiește un pătrat, care va reprezenta cazurile totale, iar în interiorul lui, în conformitate cu minutele incluse pe *Form1*, se construiește figura cazurilor favorabile – fâșia marcată.

De asemenea, pe acest pătrat este realizată o rețea cu marcajul minutelor, care este creată de următorul ciclu:

for i:=0 to 12 do // pe orizontală

begin

*if i=0 then paintbox1.Canvas.TextOut(x0-250+i*40, y0+240, '18.00') else*

*if i=1 then paintbox1.Canvas.TextOut(x0-250+i*40, y0+240, '18.05') else* *if i=12 then paintbox1.Canvas.TextOut(x0-250+i*40, y0+240, '19.00') else*

*paintbox1.Canvas.TextOut(x0-250+i*40, y0+240, '18.'+inttostr(i*5));*

end;

Această aplicație, funcționează cu ajutorul a două *Timer*-e. Unul este activ la accesarea soft-ului, *Timer2*, care extrage numărul de probe (*N*), de pe formă, din *Edit1*

n:=strtoint(edit1.Text);

și realizează construcția geometrică cu datele respective de pe *Form1*.

Timer1 se activează la începerea experimentului și se finisează în momentul în carea a realizat *n* probe utilizând o repartiție întâmplătoare a cazurilor de sosire a celor 2 prieteni

```
x:=x0-239+random(478);
y:=y0-239+random(478);
d1:=(8*min)/sqrt(2);
d2:=Abs(x+y-x0-y0)/sqrt(2);
if d2<=d1 then
begin
c:=c+1;
paintbox1.Canvas.Pen.Color:=clred;
paintbox1.Canvas.Ellipse(x-1,y-1,x+1,y+1);
n:=n+1;
end else
begin
cn:=cn+1;
paintbox1.Canvas.pen.Color:=clblue;
paintbox1.Canvas.Ellipse(x-1,y-1,x+1,y+1);
n:=n+1;
end;
```

În același timp, în *Timer2* se numără cazurile din fâșia respectivă și se afișează într-un *Label*

label4.Caption:='nr. de cazuri fav='+inttostr(c);

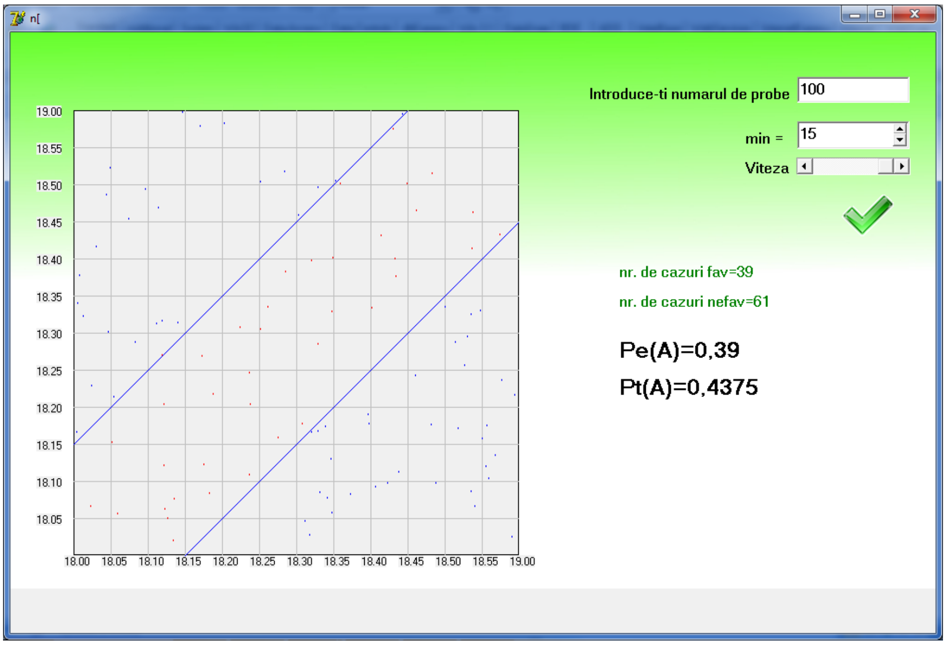


Fig. 34. Realizarea unui experiment din 100 probe

În figura 35 este prezentat un alt rezultat al experimentului ce corespunde problemei.

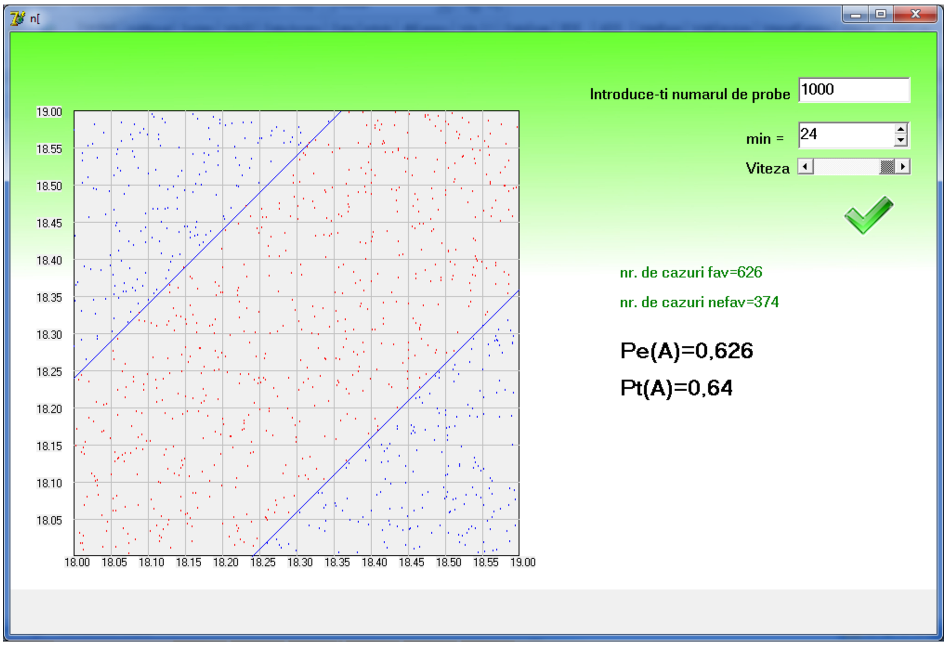


Fig. 35. Realizarea unui experiment din 1000 de probe

Problema 6

Într-o urnă sunt 26 de bile, pe care sunt indicate câte o literă a alfabetului latin (fără repetare). Se extrag, aleator, fără întoarcere, N bile (N - numărul de litere din cuvântul țintă). Să se determine probabitatea că literele extrase (în ordinea respectivă) formează cuvântul ARC (exemplul 2, §1.6) [30].

Rezolvare. Pentru realizarea experimentului, a fost creată o simulare în mediul de programare Delphi care ne realizează exprimentul problemei, dar și ne calculează atât probabilitatea teoretică cât și cea experientală pentru oricare cuvânt ce satisface condiția menționată mai sus.

În acest caz, pentru realizarea constructivistă a programului, am creat mai multe funcții și proceduri separate care sunt apelate în componenta Button2 (START).

```
procedure TForm1.Button2Click(Sender: TObject);
var p,pt:extended;
t:integer;
s:string;
begin
randomize;
  label1.Caption:='';
  label8.Caption:='';
  label9.Caption:='';
 s:=edit1.Text;
 if (unique_str(s)) then
   begin
    experiment(strtoint(edit2.Text),s,p);
 label1.Caption:='Pexp('+edit1.Text+')='+floattostr(p);
    nl:=length(edit1.Text);
    pt:=1;
    for  t:=0 to nl-1 do
    begin
    pt:=pt*(1)/(26-t);
    end;      label8.Caption:='Pt('+edit1.Text+')='+floattostr(pt);
   end
 else label9.Caption:='Atentie! Literele nu trebuie sa se repete!';
end;
```

De asemenea, în acest buton, după realizarea experimentului și calcularea probabilității experimentale, se calculează și probabilitatea teoretică, după cum rezultă și din codul de mai sus a evenimentului Button2Click.

Funcția *unique_str(s)*, este inclusă în condiția instrucțiunii *if*, dacă aceasta este adevărată, atunci se realizează experimentul problemei și are codul

```
function unique_str(str:string):boolean;
var i,j:integer;
begin
 for i := 1 to length(str) do
```

```
for j := 1 to length(str) do
  if(i<>j) then
    if str[i]=str[j] then
      begin
        Result:=false; Exit;
      end;
 Result:=true;
end;
```

Cu ajutorul acestei funcții se verifică dacă literere introduse în componenta *Edit2* sunt unice (nu se repetă în cuvântul țintă), în caz contrar nu se realizează experimentul, dar se afișează un mesaj de avertizare, care ne atenționează despre repetarea acestor litere în cuvânt, adică încălcăm condiția problemei 6 (fig. 36).

Fig. 36. Cazul nerespectării condiției problemei 6

În *procedura experiment* se realizează experimentul problemei menționate, apelându-se funcțiile ajutătoare *IndexOfArray*, *In_Array* și *Compare_Arr*.

```
procedure experiment(n:integer;str:string;var prob:extended);
var a,b:array of integer;
 z,i,r,c,j:integer;
s:string;
begin
  z:=length(str);
  c:=0;
  s:='';
  nl:=0;
```

```
setlength(a,z);
setlength(b,z);
 for j := 1 to n do
  begin
   for I := 0 to z-1 do
     begin
       a[i]:=IndexOfArray(str[i+1],arr);
       repeat
       r:=random(26);
       until not In_Array(r,b);
       b[i]:=r;
       s:=s+arr[r];
       end;
   if (Compare_Arr(a,b)) then c:=c+1;
   s:=s+AnsiString(#13#10);
   end;
  form1.Memo1.Text:=s;
 prob:=c/n;
end;
```

Rezultatul experimentului se afișează în componenta *Memo1*, de pe formă, unde pot fi vizualizate toate rezultatele extragerii, dar și probabilitatea experimentală și cea teoretică, care sunt afișate *Label1* și respectiv *Label8* (fig. 37).

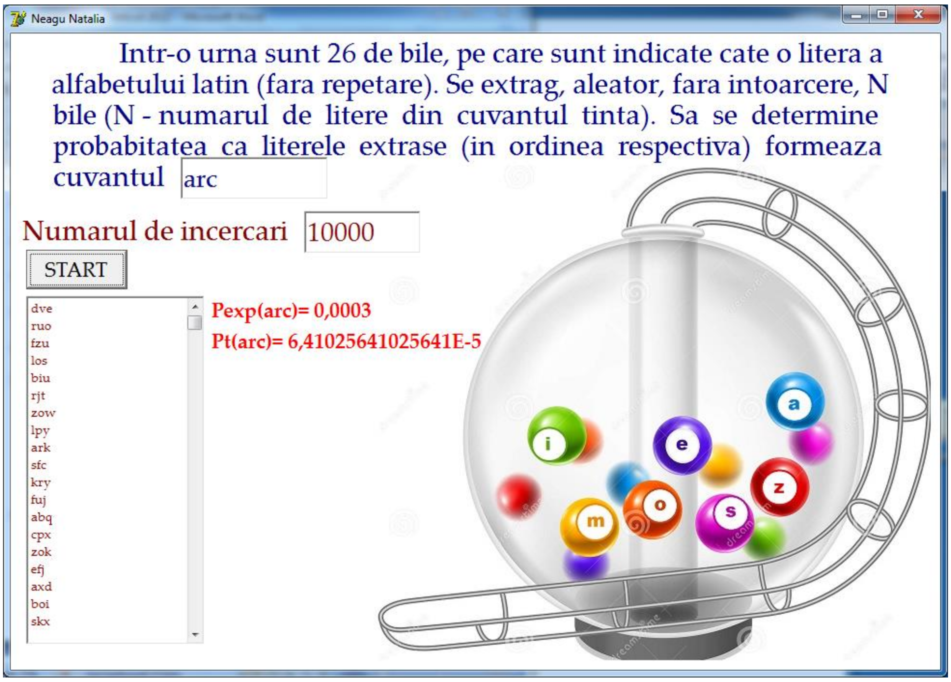


Fig. 37. Simularea experimentului din problema 6

De asemenea avem posibilitatea să modificăm cuvântul țintă. Pentru aceasta, pe formă am introdus o componentă *Edit2*, în care poate fi introdus oricare cuvânt ce satisfce condiției problemei 6.

Fie cuvântul țintă MARC, atunci în rezultatul experimentului, vom obține următoarea probabilitatea teoretică și cea experimentală (fig. 38):

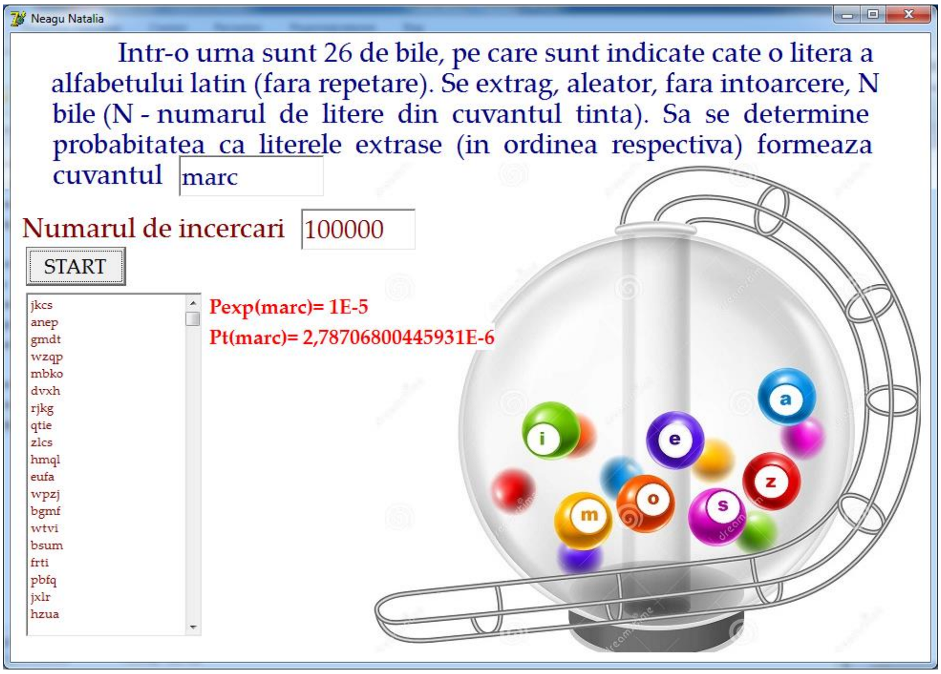


Fig. 38. Rezultatul extragerii cuvântului MARC

Este cunoscut, cu cât lungimea cuvântului țintă este mai mare cu atât probabilitatea extragerii cuvântului dat, va tinde la zero. Din acestă cauză am selectat cuvintele țintă ARC și MARC, însă programul funcționează pentru oricare lungime posibilă de cuvânt (lungimea nu trebie să fie mare, deoarece pentru realizarea calculelor va fi necesar un interval mare de timp).

Remarcă. Este important să cunoaștem că numărul de încercări contează foarte mult, deoarece cu cât acest număr este mai mare cu atât eroarea de calcul, a probabilității experimentale, este mai mică și valoarea sa este mai aproape de cea teoretică [1].

Aceste simulări sunt bine venite în cadrul liceelor, dar și în cadrul instituțiilor superioare, unde evelii sau studenții studiază capitolul Probabilitate și statistică.

În Republica Moldova, astfel de simulări, nu se realizează, iar ceea ce am realizat este o simulare unică.

În concluzie, menționez, cu cât numărul de experimente este mai mic cu atât probabilitatea experimentală poate varia mai mult față de cea teoretică. O dată nu creșterea numărului de experiențe, eroare de calcul se micșorează, iar rezultatele experimentului devin mai exacte. Simulările unor probleme din probabilitate sunt binevenite în învățăânt și pot înlocui unele experimente care necesită utilaje costisitoare, instrumente, timp, etc.

BIBLIOGRAFIE

1. Ciuac P., ș.a. *Teoria probabilităților & elemente de statistică matematică*, Editura "Tehnica" UTM, Chișinău, 2003, 278 p. ISBN 9975-9704-7-8
2. Mihoc I., Fătu C.I., *Calculul probabilităţilor şi statistică matematică*, Casa de editură-Transilvania Press, Cluj – Napoca, 2003, 577 p. ISBN: 973-95635-8-9
3. Petrehuș V., Popescu S.-A., *Probabilități și statistică*, Universitatea Tehnică de Construcții, București, 2005, 225 p.
4. Petrehuș V., Popescu S.-A., *Probabilități și statistică (teorie, exemple, probleme)*, Universitatea Tehnică de Construcții, București, 1997, 213 p.
5. Trandafir R., *Introducere în teoria probabilităţilor*, Editura Albatros, 1979, 253 p.
6. Corlat A., *Elemente de teorie probabilităților*, Biotehdesign, Chişinău, 2017, p 44. ISBN 978-9975-108-30-0
7. Chirilov P., ș.a. *Teoria probabilităţilor şi statistica matematică*, UCCM, Chișinău, 2013, 308 p.
8. Chirilov P., Vulpe I., *Curriculum la disciplina "Teoria probabilităţilor şi statistica matematică": Pentru studenţii tuturor specialităţilor, studii cu frecvenţă la zi şi cu frecvenţă redusă*, UCCM, Chişinău, 2011, 18 p. ISBN 978-9975-4198-6-4
9. Poştaru A., *Teoria probabilităţilor*, CEP USM, Chişinău, 2008, 367 p. ISBN: 978-9975-70-474-8
10. Poştaru A., Benderschi O., *Teoria Probabilităților: Exemple și probleme*, CEP USM, Chişinău, 2015, 258 p. ISBN: 978-9975-71-614-7
11. Poştaru A., Leahu A., *Probabilitate - procese aleatoare şi aplicaţii*, Ştiinţa, Chişinău, 1991,150 p. ISBN: 5-376-01104-6
12. Marcov B, ș.a., *Culegere de probleme la teoria probabilităţilor*, Universitatea de Stat din Tiraspol, Chișinău, 1998, 82 p.
13. Căbulea L., Aldea M., *Elemente de teoria probabilităţilor şi statistică matematică*, Editura Didactica, Alba Iulia, 2004.
14. Tudor C., *Teoria Probabilităţilor*. Editura Universităţii din Bucureşti, 2004.
15. Reischer C., Sâmboan A., *Culegere de probleme de teoria probabilităţilor şi statistică matematică*, Editura Didactică şi Pedagogică, Bucureşti, 1972, 228 p.
16. Blaga P., *Statistică matematică*. Ediţia a II-a, Universitatea „BabeşBolyai", Cluj-Napoca, 2001.

17. Șterbeți C., *Calculul probabilităților*, 163 p. (accesat 04.11.2021) https://www.ucv.ro/pdf/departamente_academice/dma/suporturi_curs/Matematica-curs-STERBETI_CATALIN.pdf
18. *Variabile aleatoare discrete*, 58 p. (accesat 22.02.2022) http://math.etc.tuiasi.ro:81/rosu/didactic/MS%20II_Slides_Variabile%20aleatoare%20discrete.pdf
19. Trîmbițaș R., *Legea numerelor mari și legi limită*, 2012, 28p. (accesat 11.03.2022) http://math.ubbcluj.ro/~tradu/geologi/lnmtlc.pdf
20. *Variabile aleatoare continue*, 10 p. (accesat 10.01.2021) http://math.etc.tuiasi.ro/rstrugariu/cursuri/SPD2015/c7.pdf
21. *Probabilităţi şi statistică*, 302 p. (accesat 19.03.2022) http://dep2.mathem.pub.ro/pdf/didactice/Probabilitati%20si%20statistica.pdf
22. Balmuș I. *Teoria Probabilităţilor şi a Informaţiei în Sistemul de programe Mathematica*, Editura „Tehnica-UTM", Chișinău, 2017, 132 p. (accesat 09.08.2022) http://calc.fcim.utm.md/biblioteca/arhiva/Anul%20I/Semestru%20II/Matematici%20Speciale/Teoria%20Probabilitatilor.pdf
23. Popescu, A., *Teoria probabilităţilor şi statistică matematică*. *Culegere de probleme*, Editura Universitas Petroşani, 2015, 195 p. (accesat 01.07.2022) https://www.researchgate.net/publication/313887330_Teoria_probabilitatilor_si_statistica_matematica_Culegere_de_probleme
24. Popovici L., *Statistica. Suport de curs*, Chișinău, 2019, 58 p. (accesat 11.08.2022) https://irek.ase.md/xmlui/bitstream/handle/1234567890/736/Suport_Statistica_CON.pdf?sequence=1&isAllowed=y
25. *Indicator de poziție* (accesat 15.07.2022) https://ro.frwiki.wiki/wiki/Indicateur_de_position
26. *Cercetarea prin sondaj*, Universitatea de Vest din Timișoara, 5 p. (accesat 21.09.2022) https://www.studocu.com/ro/document/universitatea-de-vest-din-timisoara/statistica-aplicata/cercetarea-prin-sondaj/13291929
27. *Elemente de sondaj statistic* (accesat 02.09.2022) https://www.stiucum.com/economie/statistica/Elemente-de-sondaj-statistic33195.php
28. Neagu N., *Calculul probabilității geometrice asistat de calculator*, Chișinău, 2016, 3 p. (accesat 22.10.2022)

http://dir.upsc.md:8080/xmlui/bitstream/handle/123456789/982/NEAGU%2CN._CALCULUL_PROBALITATII_GEOMETRICE_ASISTAT_DE_CALCULATOR.pdf?sequence=1&isAllowed=y

29. Neagu N., *Simularea aplicaţiei probabilităţii geometrice*. În: Abstracts, Conferința "Probleme ale ştiinţelor socioumanistice şi modernizării învăţământului", Seria 19, Vol. 2, Chișinău, 2017, p. 170-175 https://ibn.idsi.md/sites/default/files/imag_file/170-175_2.pdf
30. Neagu N., *Implementarea tehnologiilor informaționale la rezolvarea unor probleme din probabilitate*. În *Materialele conferinţei ştiinţifice naționale cu participare internaţională "Probleme ale științelor socioumanistice și ale modernizării învățământului"*. Chișinău, UPSC, 25 martie, 2022, seria XXIV, vol. II, pp. 293–298. ISBN 978-9975-46-651-6 http://dir.upsc.md:8080/xmlui/handle/123456789/3647

ANEXE

Tabelul A1. Densitatea de probabilitate a repartiției normale normate $N(0,1)$

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

x	0	1	2	3	4	5	6	7	8	9
0,0	0,3989	3989	3989	3988	3986	3984	3982	3980	3977	3973
0,1	3970	3965	3961	3956	3951	6945	3939	3932	3925	3918
0,2	3910	3902	3894	3885	3876	3867	3857	3847	3836	3825
0,3	3814	3802	3790	3778	3765	3752	3739	3726	3712	3697
0,4	3683	3668	3653	3637	3621	3605	3589	3572	3555	3538
0,5	3521	3503	3485	3467	3448	3429	3410	3391	3372	3352
0,6	3332	3312	3292	3271	3251	3230	3209	3187	3166	3144
0,7	3123	3101	3079	3056	3034	3011	2989	2966	2943	2920
0,8	2897	2874	2850	2827	2803	2780	2756	2732	2709	2685
0,9	2661	2637	3613	2589	2565	2541	2513	2492	2468	2444
1,0	0,2420	2396	2371	2347	2323	2299	2275	2251	2227	2203
1,1	2179	2155	2131	2107	2083	2059	2036	2012	1989	1965
1,2	1942	1919	1895	1872	1849	1826	1804	1781	1758	1736
1,3	1714	1691	1669	1647	1626	1604	1582	1561	1539	1518
1,4	1497	1476	1456	1435	1415	1394	1374	1354	1334	1315
1,5	1295	1276	1257	1238	1219	1200	1182	1163	1145	1127
1,6	1109	1092	1074	1057	1040	1023	1006	0989	0973	0957
1,7	0940	0925	0909	0893	0878	0863	0848	0833	0818	0804
1,8	0790	0775	0761	0748	0734	0721	0707	0694	0681	0669
1,9	0656	0644	0632	0620	0608	0596	0584	0573	0562	0551
2,0	0,0540	0529	0519	0508	0498	0488	0478	0468	0459	0449
2,1	0440	0431	0422	0413	0404	0396	0387	0379	0371	0363
2,2	0355	0347	0339	0332	0325	0317	0310	0303	0297	0290
2,3	0283	0277	0270	0264	0258	0252	0246	0241	0235	0229
2,4	0224	0219	0213	0208	0203	0198	0191	0189	0184	0180
2,5	0175	0171	0167	0163	0158	0154	0151	0147	0143	0139
2,6	0136	0132	0129	0126	0122	0119	0116	0113	0110	0107
2,7	0104	0101	0099	0096	0093	0091	0088	0086	0084	0081
2,8	0079	0077	0075	0073	0071	0069	0067	0065	0063	0061
2,9	0060	0058	0056	0055	0053	0051	0050	0048	0047	0046
3,0	0, 0044	0043	0042	0040	0039	0038	0037	0036	0035	0034
3,1	0033	0032	0031	0030	0029	0028	0027	0026	0025	0025
3,2	0024	0023	0022	0022	0021	0020	0020	0019	0018	0018
3,3	0017	0017	0016	0016	0015	0015	0014	0014	0013	0013
3,4	0012	0012	0012	0011	0011	0010	0010	0010	0009	0009
3,5	0009	0008	0008	0008	0008	0007	0007	0007	0007	0006
3,6	0006	0006	0006	0005	0005	0005	0005	0005	0005	0004
3,7	0004	0004	0004	0004	0004	0004	0003	0003	0003	0003
3,8	0003	0003	0003	0003	0003	0002	0002	0002	0002	0002
3,9	0002	0002	0002	0002	0002	0002	0002	0002	0001	0001

Tabelul A2. Funcția de repartiție normală normată (Laplace) $N(0,1)$

$$\Phi(-x) = 1 - \Phi(x), \ \ \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt$$

x	0,0000	0,0100	0,0200	0,0300	0,0400	0,0500	0,0600	0,0700	0,0800	0,0900
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7290	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9779	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9983	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986

Cuantilele repartiție $N(0,1)$

p	0,90	0,95	0,975	0,99	0,999	0,9995
u_p	1,282	1,645	1,960	2,326	3,090	3,291

Tabelul A3. Repartiția Poisson $P\{X = k\} = \frac{\lambda^k}{k!}e^{-\lambda}$

k \ λ	0,1	0,2	0,3	0,4	0,5
0	0,90484	0,81873	0,74082	0,67032	0,60653
1	0,09048	0,16375	0,22225	0,26813	0,30327
2	0,00452	0,01638	0,03334	0,05363	0,07583
3	0,00015	0,00109	0,00333	0,00715	0,01204
4		0,00006	0,00025	0,00072	0,00158
5			0,00002	0,00006	0,00016
6					0,00001

k \ λ	0,6	0,7	0,8	0,9
0	0,54881	0,49659	0,41933	0,40657
1	0,32920	0,34761	0,35946	0,36591
2	0,09879	0,12166	0,14379	0,16166
3	0,01976	0,02839	0,03831	0,01919
4	0,00296	0,00497	0,00767	0,01112
5	0,00036	0,00070	0,00123	0,00200
6	0,00004	0,00008	0,00016	0,00030
7		0,00001	0,00002	0,00004

k \ λ	1,0	2,0	3,0	4,0	5,0
0	0,36788	0,13534	0,04979	0,01832	0,00674
1	0,36788	0,27067	0,14936	0,07326	0,03369
2	0,18394	0,27067	0,22404	0,14653	0,08422
3	0,06131	0,18045	0,22404	0,19537	0,14037
4	0,01533	0,09022	0,16803	0,19537	0,17547
5	0,00307	0,03609	0,10082	0,15629	0,17547
6	0,00051	0,01203	0,05041	0,10419	0,14632
7	0,00007	0,00344	0,02160	0,05954	0,10445
8	0,00001	0,00080	0,00810	0,02977	0,06528
9		0,00019	0,00270	0,01323	0,03627
10		0,00004	0,00081	0,00529	0,01813
11		0,00001	0,00022	0,00193	0,00824
12			0,00006	0,00064	0,00343
13			0,00001	0,00020	0,00132
14				0,00006	0,00047
15				0,00002	0,00016
16					0,00005
17					0,00001

Tabelul A4. Cuantilele repartiție χ^2 (hi pătrat)

$$f_{\chi^2(\nu)}(x) = \frac{1}{2^{\frac{\nu}{2}}\Gamma\left(\frac{\nu}{2}\right)} x^{\left(\frac{\nu}{2}-1\right)} e^{-\frac{x}{2}}, \qquad x > 0, \nu > 0$$

k\p	**0,005**	**0,010**	**0,025**	**0,05**	**0,10**	**0,20**	**0,30**	**0,70**	**0,80**	**0,90**	**0,95**	**0,975**	**0,990**	**0,995**	**0,999**
1	$0,0^4393$	$0,0^3157$	$0,0^3982$	$0,0^239$	0,0158	0,0642	0,148	1,07	1,64	2,71	3,84	5,02	6,63	7,88	10,8
2	0,0100	0,0201	0,0506	0,103	0,211	0,446	0,713	2,41	3,22	4,61	5,99	7,38	9,21	10,6	13,8
3	0,0717	0,115	0,216	0,352	0,584	1,005	1,424	3,66	4,64	6,25	7,81	9,35	11,3	12,8	16,3
4	0,207	0,297	0,484	0,711	1,060	1,650	2,190	4,88	5,99	7,78	9,49	11,10	13,3	14,9	18,5
5	0,412	0,554	0,831	1,150	1,610	2,340	3,000	6,06	7,29	9,24	11,10	12,80	15,1	16,7	20,5
6	0,676	0,872	1,240	1,640	2,200	3,070	3,830	7,23	8,56	10,60	12,60	14,40	16,8	18,5	22,5
7	0,989	1,240	1,690	2,170	2,830	3,820	4,670	8,38	9,80	12,00	14,10	16,00	18,5	20,3	24,3
8	1,340	1,650	2,180	2,730	3,490	4,590	5,530	9,52	11,00	13,40	15,50	17,50	20,1	22,0	26,1
9	1,730	2,090	2,700	3,330	4,170	5,380	6,390	10,70	12,20	14,70	16,90	19,00	21,7	23,6	27,9
10	2,160	2,560	3,250	3,940	4,870	6,180	7,270	11,80	13,40	16,00	18,30	20,50	23,2	25,2	29,6
11	2,600	3,050	3,820	4,570	5,580	6,990	8,150	12,90	14,60	17,30	19,70	21,90	24,7	26,8	31,3
12	3,070	3,570	4,400	5,230	6,300	7,810	9,030	14,00	15,80	18,50	21,00	23,30	26,2	28,3	32,9
13	3,570	4,110	5,010	5,890	7,040	8,630	9,930	15,10	17,00	19,80	22,40	24,70	27,7	29,8	34,5
14	4,070	4,660	5,630	6,570	7,790	9,470	10,80	16,20	18,20	21,10	23,70	26,10	29,1	31,3	36,1
15	4,600	5,230	6,260	7,260	8,550	10,30	11,70	17,30	19,30	22,30	25,00	27,50	30,6	32,8	37,7
16	5,140	5,810	6,910	7,960	9,310	11,20	12,60	18,40	20,50	23,50	26,30	28,80	32,0	34,3	39,3
17	5,700	6,410	7,560	8,670	10,10	12,00	13,50	19,50	21,60	24,80	27,60	30,20	33,4	35,7	40,8
18	6,260	7,010	8,230	9,390	10,90	12,90	14,40	20,60	22,80	26,00	28,90	31,50	34,8	37,2	42,3
19	6,840	7,630	8,910	10,10	11,70	13,70	15,40	21,70	23,90	27,20	30,10	32,90	36,2	38,6	43,8
20	7,430	8,260	9,590	10,90	12,40	14,60	16,30	22,80	25,00	28,40	31,40	34,20	37,6	40,0	45,3
21	8,030	8,900	10,30	11,60	13,20	15,40	17,20	23,90	26,90	29,60	32,70	35,50	38,9	41,4	46,8
22	8,640	9,540	11,00	12,30	14,00	16,30	18,10	24,90	27,30	30,80	33,90	36,80	40,3	42,8	48,3
23	9,260	10,20	11,70	13,10	14,80	17,20	19,10	26,00	28,40	32,00	35,20	38,10	41,6	44,2	49,7
24	9,890	10,90	12,40	13,80	15,70	18,10	19,90	27,10	29,60	33,20	36,40	39,40	43,0	45,6	51,2
25	10,50	11,50	13,10	14,60	16,50	18,90	20,90	28,20	30,70	34,40	37,70	40,60	44,3	46,9	52,6
26	11,20	12,20	13,80	15,40	17,30	19,80	21,80	29,20	31,80	35,60	38,90	41,90	45,6	48,3	54,1
27	11,80	12,90	14,60	16,20	18,10	20,70	22,70	30,30	32,90	36,70	40,10	43,20	47,0	49,6	55,5
28	12,50	13,60	15,30	16,90	18,90	21,60	23,60	31,40	34,00	37,90	41,30	44,50	48,3	51,0	56,9
29	13,10	14,30	16,00	17,70	19,80	22,50	24,60	32,50	35,10	39,10	42,60	45,70	49,6	52,3	58,3
30	13,80	15,00	16,80	18,50	20,60	23,40	25,50	33,50	36,30	40,30	43,80	47,00	50,9	53,7	59,7
35	17,20	18,50	20,60	22,50	24,80	27,30	30,20	38,90	41,80	46,10	49,80	53,20	57,3	60,3	66,6
40	20,70	22,20	24,40	26,50	29,10	32,30	34,90	44,20	47,30	51,80	55,80	59,30	63,7	66,8	73,1
45	24,30	25,90	28,40	30,60	33,40	36,90	39,90	49,50	52,70	57,50	61,70	65,40	70,0	73,2	80,1
50	28,00	29,70	32,40	34,80	37,70	41,40	44,30	54,70	58,20	63,20	67,50	71,40	76,2	79,5	86,7
75	47,20	49,50	52,90	56,10	59,80	64,50	68,10	80,90	85,10	91,10	96,20	100,3	106,4	110,3	118,6
100	67,30	70,10	74,20	77,90	82,40	87,90	92,10	106,9	111,7	118,5	124,3	129,6	135,6	140,2	143,4