Universitatea din Bucureşti
Facultatea de Matematică şi Informatică

METODE MODERNE DE CALCUL ŞI SIMULARE

Cursurile nr. 1-4 – Anul III – Tehnologia Informaţiei

Asist. dr. Bianca Mogoş

Bucureşti, Februarie 2014

Cuprins

Capitolul 1

Noţiuni de probabilităţi şi statistică- Partea I

1.1 Experiment aleator. Spaţiu de selecţie.

Definiţia 1.1.1. *(Experiment aleator)*

Un experiment aleator este definit ca fiind o acţiune al cărei rezultat nu poate fi prezis cu certitudine şi care se poate modifica în urma repetării experimentului.

Observaţie 1.1.2. Variabilitatea rezultatelor ieşite în urma unui experiment aleator poate apărea ca urmare a unor erori de măsurare, a alegerii unor obiecte diferite pentru testare, etc.

Exemplul 1.1.3.

- Observarea pe un interval T de timp a funcţionării unui calculator.
- Înregistrarea consumului de energie electrică a unui combinat.

Definiţia 1.1.4. *(Spaţiu de selecţie)*

Spaţiul de selecţie al unui experiment aleator, notat prin Ω, reprezintă mulţimea tuturor rezultatelor posibile.

Exemplul 1.1.5.

- Prin aruncarea banului se pot obţine două rezultate. Astfel, spaţiul de selecţie este $\{0,1\}$.
- Prin rostogolirea unui zar cu şase feţe şi numărarea punctelor de pe o faţă se pot obţine şase rezultate posibile. Astfel, spaţiul de selecţie este $\{1,2,3,4,5,6\}$.

1.2 Evenimente

Definiţia 1.2.1. *(Eveniment)*

Un eveniment este orice submulţime de rezultate conţinute în spaţiul de selecţie. Un eveniment este elementar dacă el constă dintr-un singur rezultat şi compus dacă constă din mai multe rezultate.

Definiţia 1.2.2. *(Evenimentul sigur şi evenimentul imposibil)*

Evenimentul sigur este evenimentul care se realizează întotdeauna ca rezultat al experimentului; va fi notat cu Ω (se asociază mulţimii totale de rezultate).

Evenimentul imposibil nu se poate realiza ca rezultat al unui experiment; va fi notat cu $\emptyset$ (corespunde mulţimii vide).

Definiţia 1.2.3. *(Reuniunea şi intersecţia evenimentelor)*

Numim reuniunea evenimentelor $A_1, A_2, \ldots, A_k$, notată prin $\bigcup\limits_{j=1}^{k} A_j$, evenimentul care se realizează când cel puţin unul dintre evenimentele $A_1, A_2, \ldots, A_k$ se realizează.

Numim intersecţia evenimentelor $A_1, A_2, \ldots, A_k$, notată prin $\bigcap\limits_{j=1}^{k} A_j$, evenimentul care se realizează când se realizează toate evenimentele $A_1, A_2, \ldots, A_k$.

1.3 Axiomele probabilităţii. Câmp de probabilitate

Probabilitatea unui eveniment este o măsură prin care gestionăm/modelăm incertitudinea producerii fenomenelor şi a apariţiei datelor din lumea reală. Aceasta ne permite să apreciem gradul de încredere şi să cuantificăm lipsa de siguranţă inerentă în procesul care generează datele analizate.

Definiţia 1.3.1. *(Definiţia clasică a probabilităţii)*

Dacă într-un experiment cu "n" rezultate, "k"dintre ele favorizează realizarea evenimentului A, definim probabilitatea P(A) a evenimentului A prin

$$P(A) = \frac{k}{n} \tag{1.3.1}$$

Exerciţiul 1.3.2. *(Aruncarea cu zarul)*

1. Calculaţi probabilitatea ca în urma aruncării unui zar să iasă un număr par.
2. Care este probabilitatea ca în urma aruncării unui zar să iasă un număr impar, strict mai mare ca 3?

Teoria modernă a probabilităţilor a fost fondată la începutul secolului XX. O contribuţie esenţială a avut-o matematicianul rus A. Kolmogorov.

Fie Ω mulţimea evenimentelor elementare (rezultatelor posibile) ale unui experiment aleator.

Definiţia 1.3.3. *(σ – algebră)*

$\mathcal{B} \subset \mathcal{P}(\Omega)$ este o σ – algebră dacă

1. $\emptyset, \Omega \in \mathcal{B}$

2. Dacă $A \in \mathcal{B}$ atunci $A^C \in \mathcal{B}$

3. Dacă $(A_n)_{n \in N^*}, A_n \in \mathcal{B}$ atunci $\bigcup_{n=1}^{\infty} A_n \in \mathcal{B}$

Perechea $(\Omega, \mathcal{B})$ se numeşte câmp de evenimente.

Definiţia 1.3.4. *(Probabilitate)*

Se numeşte probabilitate pe $\mathcal{B}$ o funcţie nenegativă $P : \mathcal{B} \to [0, 1]$ cu proprietăţile:

1. $P(\Omega) = 1$

2. Dacă $A, B \in \mathcal{B}$ şi $A \cap B = \emptyset$ atunci $P(A \cup B) = P(A) + P(B)$.

Proprietăţi 1.3.5.

1. $P(A^C) = 1 - P(A)$, deoarece $A \cup A^C = \Omega, A \cap A^C = \emptyset$.

2. $P(\emptyset) = P(\Omega^C) = 1 - P(\Omega) = 0$.

3. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$, dacă $A \cap B \neq \emptyset$.
Relaţia 3) se demonstrează astfel:

$A = A \cap (B \cup B^C) = (A \cap B) \cup (A \cap B^C)$ cu $A \cap B \cap B^C = \emptyset$.

$B = B \cap (A \cup A^C) = (B \cap A) \cup (B \cap A^C)$ cu $B \cap A \cap A^C = \emptyset$.
Dar, $A \cup B = (A \cap B) \cup (A \cup B^C) \cup (B \cap A^C)$.
Aplicăm axioma 2) din definiţia probabilităţii şi obţinem

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

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

$P(A \cup B) = P(A \cap B) + P(A \cap B^C) + P(B \cap A^C)$
Din cele trei relaţii deduse mai sus rezultă proprietatea 3.

Definiţia 1.3.6. *(Câmp de probabilitate)*

Tripletul $(\Omega, \mathcal{B}, P)$ se numeşte câmp de probabilitate.

Exerciţiul 1.3.7. *(Problema controlului alimentelor)*
Se ştie că într-o cutie sunt 550 de mere. Se verifică aleator 25 dintre ele dacă sunt stricate. Dacă 2% dintre merele din cutie sunt stricate, care este probabilitatea ca printre cele 25 testate săgăsim 2 mere stricate?

Exerciţiul 1.3.8. *(Problema zilei de naştere)*
Care este probabilitatea ca cel puţin 2 persoane dintr-un grup de n indivizi să aibă aceeaşi zi de naştere?

1.4 Probabilităţi condiţionate

Definiţia 1.4.1. *Evenimente independente*

Două evenimente A şi B din $\mathcal{B}$ sunt independente dacă nu se influenţează, adică realizarea evenimentului A nu depinde de realizarea lui B şi reciproc.

Definiţia 1.4.2. *(Probabilitate condiţionată)*

Se numeşte probabilitate condiţionată a evenimentului A de către evenimentul B şi se notează $P(A|B)$ sau $P_B A$ probabilitatea de realizare a evenimentului A calculată în condiţia că evenimentul B s-a realizat $(P(B) > 0)$ şi se defineşte ca:

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

Exemplul 1.4.3.

Un tratament s-a aplicat la 700 de pacienţi care sufereau de piatră la rinichi. S-a constatat că doar 562 s-au vindecat. De asemenea, se ştie că 357 au pietre mici şi 315 dintre aceştia s-au vindecat, iar 343 au pietre mari şi 247 s-au vindecat. Să se calculeze probabilitatea ca un pacient să se vindece ştiind că acesta are piatră mică. Dar, dacă are piatră mare?

Soluţie:

Fie

$\Omega = \{1, 2, \ldots, 700\}$ mulţimea tuturor pacienţilor,

$S = \{1, 2, \ldots, 562\}$ mulţimea pacienţilor vindecaţi şi

$E = \{563, \ldots, 700\}$ mulţimea pacienţilor nevindecaţi.

Obţinem următoarele probabilităţi:

$P(\omega) = \frac{1}{700}, \forall$ evenimentul elementar $\omega \in \Omega$

$P(S) = \frac{562}{700} \approx 0,8$

$P(E) = \frac{138}{700} =\approx 0,2$

Notăm cu

$\Omega_1 =$ pacienţii care au pietre mici

$\Omega_2 =$ pacienţii care au pietre mari

Rezultă $P(\Omega_1) = \frac{357}{700} = 0,51$ şi $P(\Omega_2) = \frac{343}{700} = 0,49$. Astfel

$$P(S|\Omega_1) = \frac{P(S \cap \Omega_1)}{P(\Omega_1)} = \frac{315}{700} \cdot \frac{1}{0,51} = 0,88$$

Analog, se arată că

$P(S|\Omega_2) = 0,72.$

Exerciţiul 1.4.4.

Presupunem că 1% din populaţie este bolnavă. Un test de analiză a sângelui dă următoarele rezultate: în 95% dintre cazuri testul iese pozitiv pentru persoanele bolnave şi în 2% dintre cazuri iese pozitiv pentru persoanele sănătoase. Care este probabilitatea ca o persoană căreia i-a ieşit testul pozitiv să fie bolnavă?

Definiţia 1.4.5. *(Evenimente independente în totalitate)*

Spunem că evenimentele $A_1, A_2, \ldots, A_k$ sunt independente în totalitate dacă sunt independente două câte două, sunt independente trei câte trei,..., sunt independete toate "k''.

Formal, avem relaţiile:

$$P(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_p}) = P(A_{i_1})P(A_{i_2}) \ldots P(A_{i_p}),$$
$$\forall i_1, \ldots, i_p \in \overline{1,k} \text{ distincte }, p = \overline{1,k}$$

Exemplul 1.4.6. *(Evenimentele nu sunt independente în totalitate)*

Fie spaţiul de selecţie $\Omega = \{1, 2, 3, 4\}$ şi evenimentele.

$A = \{1, 2\}$
$B = \{1, 3\}$
$C = \{1, 4\}$

Cum $A \cap B = \{1\}$ rezultă $P(A \cap B) = \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} = P(A)P(B)$. Rezultă că evenimentele A şi B sunt independente.

Iar $P(A \cap B \cap C) = \frac{1}{4} \neq P(A)P(B)P(C) = \frac{1}{8}$. Deducem că evenimentele A, B şi C nu sunt independente.

1.5 Formula lui Bayes

Fie $(A_i)_{i=\overline{1,k}}$ o partiţie a mulţimii Ω, ceea ce înseamnă că

$$A_i \cap A_j = \emptyset, \forall i, j \in \overline{1,k}, i \neq j, P(A_j) > 0, \forall j = \overline{1,k} \text{ şi } \bigcup_{j=1}^{k} A_j = \Omega.$$

Fie $X \in \mathcal{B}$ un eveniment oarecare. Atunci

$$P(X) = P(X \cap \Omega) = P(\bigcup_{j=1}^{k} X \cap A_j) = \sum_{j=1}^{k} P(X \cap A_j) = \sum_{j=1}^{k} P(A_j) \cdot P(X|A_j)$$

Obţinem relaţia

$$P(X) = \sum_{j=1}^{k} P(A_j) \cdot P(X|A_j)$$

Atunci probabilitatea condiţionată a evenimentului A_j de către evenimentul X se poate calcula cu formula

$$P(A_j|X) = \frac{P(A_j) \cdot P(X|A_j)}{\sum_{j=1}^{k} P(A_j) \cdot P(X|A_j)}. \tag{1.5.1}$$

Relaţia obţinută în ecuaţia (1.5.1) se numeşte formula lui Bayes.

Exemplul 1.5.1. *(Exemplu bazat pe formula lui Bayes)*

Presupunem că trei maşini A, B şi C lucrează cu următoarele performanţe: probabilitatea ca o piesă să fie defectă este 0,01; 0,02 şi 0,04 pentru maşinile A, B şi respectiv C. În aceeaşi zi maşina A a produs 3900 de piese, maşina B 4200 de piese, iar maşina C 3600 de piese. Se alege la întâmplare o piesă din producţia acelei zile şi se constată că este defectă. Care este probabilitatea să fi fost produsă de maşina C?

Soluţie: Notăm cu

Ω = mulţimea tuturor pieselor produse într-o zi

A = evenimentul: piesa este produsă de maşina A

B = evenimentul: piesa este produsă de maşina B

C = evenimentul: piesa este produsă de maşina C

Atunci

$P(A) = \dfrac{3900}{11700} = 0,33$

$P(B) = \dfrac{4200}{11700} = 0,36$

$P(C) = \dfrac{3600}{11700} = 0,31$

De asemenea, notăm cu D = evenimentul: piesa este defectă. Avem din enunţ probabilităţile condiţionate: $P(D|A) = 0,01; P(D|B) = 0,02$ şi $P(D|C) = 0,04$

Atunci aplicând formula Bayes obţinem

$$P(C|D) = \frac{P(C) \cdot P(D|C)}{\sum_{A \in \{A,B,C\}} P(A) \cdot P(D|A)} = \frac{0,31 \cdot 0,04}{0,0229} = 0,54.$$

Capitolul 2

Noţiuni de probabilităţi şi statistică - Partea a II-a

2.1 Variabile aleatoare discrete

2.1.1 Variabile aleatoare. Funcţie de repartiţie

Fie $(\Omega, \mathcal{B})$ un câmp de evenimente.

Definiţia 2.1.1. *(Variabilă aleatoare)*

Numim variabilă aleatoare o funcţie $X : \Omega \to \mathbb{R}$ cu proprietatea că pentru fiecare $x \in \mathbb{R}$ evenimentul $X^{-1}(x) = \{\omega \in \Omega | X(\omega) = x\}$ aparţine mulţimii $\mathcal{B}$.

Fie evenimentul $A = \{\omega \in \Omega | X(\omega) \leq x\} \in \mathcal{B}$.

Definiţia 2.1.2. *(Funcţie de repartiţie)*

Funcţia $F_X : \mathbb{R} \to [0, 1]$ definită prin

$$F_X(x) = P(\{\omega \in \Omega | X(\omega) \leq x\}) \tag{2.1.1}$$

se numeşte funcţia de repartiţie a variabilei aleatoare X.

Proprietăţi 2.1.3.

1. $F_X(x) \in [0, 1]\,, \quad \lim\limits_{x\to\infty} F_X(x) = 1, \lim\limits_{x\to-\infty} F_X(x) = 0.$
2. $F_X(x + h) \geq F_X(x)$, pentru $h > 0$, deoarece
   $F_X(x + h) = F_X(x) + P(\{\omega \in \Omega | x < X(\omega) < x + h\})$.
3. $P(\{\omega \in \Omega | a < X(\omega) \leq b\}) = F_X(b) - F_X(a)$.

Definiţia 2.1.4. *(Variabilă aleatoare discretă)*

Dacă valorile luate de variabila aleatoare X sunt în număr finit sau numărabil, spunem că X este o variabilă aleatoare discretă.
Ea este perfect determinată dacă se cunosc valorile luate de variabila aleatoare,

precum şi probabilităţile cu care apar aceste valori.
Putem scrie

$$X : \begin{pmatrix} x_1 & x_2 & \dots & x_n & \dots \\ p_1 & p_2 & \dots & p_n & \dots \end{pmatrix} \tag{2.1.2}$$

unde $p_n = P(\{\omega \in \Omega | X(\omega) = x_n, \}), n = 1, 2, \dots$

Definiţia 2.1.5. *(Repartiţia unei variabile aleatoare discrete)*

Tabloul din ecuaţia (2.1.2) se numeşte repartiţia variabilei aleatoare discrete X.

Observaţie 2.1.6.

1. $p_1 + p_2 + \ldots + p_n + \ldots = P(\{\omega \in \Omega | X(\omega) = x_1, \}) + P(\{\omega \in \Omega | X(\omega) = x_2, \}) +$

   $+ \ldots + P(\{\omega \in \Omega | X(\omega) = x_n, \}) + \ldots = P(\Omega) = 1$

   deoarece evenimentele $A_i = \{\omega \in \Omega | X(\omega) = x_i, \}$ şi $\{\omega \in \Omega | X(\omega) = x_j, \}$ cu $i \neq j$ sunt disjuncte (adică, $A_i \cap A_j = \emptyset$; altfel avem evenimentul $e \in A_i \cap A_j$, de unde rezultă $X(e) = x_i = x_j$, ceea ce este imposibil pentru $i \neq j$) şi $\bigcup\limits_i A_i = \Omega$

2. O variabilă aleatoare discretă realizează o partiţie a câmpului de evenimente.

2.1.2 Operaţii cu variabile aleatoare discrete

Definiţia 2.1.7. *(Operaţii cu variabile aleatoare discrete)* Dacă X şi Y sunt definite pe acelaşi câmp de probabilitate $(\Omega, \mathcal{B}, P)$ atunci:

$$\begin{aligned} &(X+Y)(\omega) = X(\omega) + Y(\omega) \\ &(aX)(\omega) = aX(\omega), \text{ pentru } a \in \mathbb{R} \\ &(XY)(\omega) = X(\omega)Y(\omega) \\ &\frac{X}{Y}(\omega) = \frac{X(\omega)}{Y(\omega)}, \ \textit{dacă } Y(\omega) \neq 0 \text{ pentru orice } \omega \in \Omega \end{aligned}$$

Dacă

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

atunci

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

unde $p_{ij} = P(\{\omega \in \Omega | X(\omega) = x_i\} \cap \{\omega \in \Omega | Y(\omega) = y_j\})$.

În cazul în care evenimentele

$$\{\omega \in \Omega | X(\omega) = x_i\} \text{ şi } \{\omega \in \Omega | Y(\omega) = y_j\}$$

sunt independente pentru orice $1 \le i \le n, 1 \le j \le m$ spunem că v.a. X şi Y sunt independente. În acest caz, avem $p_{ij} = p_i \cdot q_j$.

Deducem analog relaţiile:

$$aX : \begin{pmatrix} ax_1 & ax_2 \dots & ax_n & \\ p_1 & p_2 & \dots & p_n \end{pmatrix}$$

$$X \cdot Y : \begin{pmatrix} x_1y_1 & x_1y_2 & \dots & x_iy_j & \dots & x_ny_m \\ p_{11} & p_{12} & \dots & p_{ij} & \dots & p_{nm} \end{pmatrix}$$

şi

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

$y_j \neq 0$ pentru orice $j = 1, 2, \dots, m$.

2.1.3 Momente. Definiţii

Fie $(\Omega, \mathcal{B}, P)$ spaţiul de probabilitate pe care este definită variabila aleatoare discretă X.

Definiţia 2.1.8. *(Valoare medie a unei variabile aleatoare)*

Dacă X este o variabilă aleatoare, $X : \Omega \to \mathbb{R}$ atunci numim valoare medie a variabilei X,

$$E[X] = \sum_{i=1}^{\infty} x_i p_i \tag{2.1.3}$$

pentru cazul în care X ia un număr numărabil de valori şi seria este convergentă. Dacă X ia un număr finit de valori, valoarea medie se defineşte printr-o sumă finită.

Definiţia 2.1.9. *(Momentul de ordinul r)*

Momentul de ordinul r al v.a. X este $E_r[X] = E[X^r]$. Se mai foloseşte notaţia $\mu_r(X) = E[X^r]$.

Definiţia 2.1.10. *(Momentul centrat de ordin r)*

Momentul centrat de ordin r al v.a. X este numărul

$$\mu_r(X) = E[(X - E[X])^r] \tag{2.1.4}$$

Pentru $r = 2$

$$\mu_2(X) = E[(X - E[X])^2] \text{ se numeşte dispersia variabilei } X. \tag{2.1.5}$$

O notăm cu $D(X)$ şi dă o măsură a împrăştierii valorilor variabilei X în jurul valorii ei medii.

Definiţia 2.1.11. *(Coeficientul de variaţie)*

Cantitatea $\sqrt{D(X)}/E[X]$ se numeşte coeficientul de variaţie pentru variabila X.

Definiţia 2.1.12. *(Coeficientul de asimetrie)*

Cantitatea $E[(X - E[X])^3]/D^{3/2}(X)$ se numeşte coeficientul de asimetrie pentru variabila X.

Definiţia 2.1.13. *(Coeficientul de aplatizare sau de exces)*

Cantitatea $E[(X - E[X])^4]/D^2(X)$ se numeşte coeficientul de aplatizare sau de exces pentru variabila X.

2.1.4 Proprietăţi ale valorii medii

Proprietăţi 2.1.14.

$E[aX + b] = aE[X] + b, a, b \in \mathbb{R}$

Proprietăţi 2.1.15.

$E[X + Y] = E[X] + E[Y]$

Demonstraţie

$$\begin{aligned} E[X+Y] &= \sum_{i=1}^{n}\sum_{j=1}^{m}(x_i + y_j)p_{ij} = \sum_{i=1}^{n} x_i \sum_{j=1}^{m} p_{ij} + \sum_{j=1}^{m} y_j \sum_{i=1}^{n} p_{ij} \\ &= \sum_{i=1}^{n} x_i P(X = x_i) + \sum_{j=1}^{m} y_j P(Y = y_j), \end{aligned}$$

deoarece

$$\sum_{j=1}^{m} p_{ij} = P\left((X = x_i) \cap \Omega\right) = P(X = x_i)$$

şi

$$\sum_{i=1}^{n} p_{ij} = P\left(\Omega \cap (Y = y_j)\right) = P(Y = y_j).$$

Proprietăţi 2.1.16. Dacă X şi Y sunt v.a. independente, atunci

$$E[X \cdot Y] = E[X] \cdot E[Y]$$

Demonstraţie

Avem $E[X \cdot Y] = \sum_{i=1}^{n} \sum_{j=1}^{m} x_i y_j p_{ij}$. Cum X şi Y sunt independente

$$p_{ij} = P(X = x_i) \cdot P(Y = y_j).$$

Astfel,

$$E[X \cdot Y] = \sum_{i=1}^{n} P(X = x_i) \cdot \sum_{j=1}^{m} P(Y = y_j).$$

2.2 Variabile aleatoare continue

2.2.1 Funcţia de repartiţie. Densitatea de probabilitate

Definiţia 2.2.1. *(Variabilă aleatoare continuă)*

Spunem că $X : \Omega \to \mathbb{R}$ este v.a. continuă dacă

$$P(\{\omega \in \Omega | X(\omega) = x\}) = 0, \tag{2.2.1}$$

pentru orice $x \in \mathbb{R}$.

Observaţie 2.2.2.

Deşi probabilitatea evenimentului din definiţie este zero nu înseamnă că evenimentul nu se poate realiza.

Proprietăţi 2.2.3.

Dacă X este o variabilă aleatoare continuă atunci funcţia sa de repartiţie este continuă.

Demonstraţie

Trebuie să demonstrăm că

$$\lim_{\substack{x \to x_0 \\ x < x_0}} F_X(x) = \lim_{\substack{x \to x_0 \\ x > x_0}} F_X(x) = F_X(x_0).$$

Fie $x > x_0$. Avem relaţia

$$F_X(x) - F_X(x_0) = P(\{\omega \in \Omega | x < X(\omega) \le x_0\}).$$

Dar,

$$\lim_{\substack{x \to x_0 \\ x > x_0}} P(\{\omega \in \Omega | x < X(\omega) \le x_0\}) = 0$$

Atunci

$$\lim_{\substack{x \to x_0 \\ x > x_0}} F_X(x) = F_X(x_0).$$

Analog, se arată că

$$\lim_{\substack{x \to x_0 \\ x < x_0}} F_X(x) = F_X(x_0).$$

Definiţia 2.2.4. *(Densitate de probabilitate)*

Dacă există

$$\lim_{h\to 0} \frac{F_X(x+h) - F_X(x)}{h} = F'_X(x) \tag{2.2.2}$$

atunci F'_X se numeşte densitatea de probabilitate a variabilei X şi

$$F'_X(x) = f_X(x), x \in \mathbb{R} \tag{2.2.3}$$

Proprietăţi 2.2.5. *(Proprietăţi ale funcţiei de repartiţie)*

1. Deoarece $F_X(x) = \displaystyle\int_{-\infty}^{x} f_X(t)dt$ rezultă $\displaystyle\int_{-\infty}^{\infty} f_X(t)dt = 1$
2. Deoarece F_X este crescătoare rezultă că $f_X(x) \geq 0, x \in \mathbb{R}$.

2.2.2 Caracteristici ale variabilelor aleatoare continue

Definiţia 2.2.6. *(Valoare medie a unei variabile aleatoare continue)*

Valoarea medie a v.a. continue X este

$$E[X] = \int_{-\infty}^{\infty} x dF_X(x) = \int_{-\infty}^{\infty} x f_X(x)dx \tag{2.2.4}$$

Definiţia 2.2.7. *(Momentul de ordin r al unei variabile aleatoare continue)*

Momentul de ordinul r al variabilei aleatoare continue X este

$$\mu_r(X) = E[X^r] = \int_{\mathbb{R}} x^r f_X(x)dx. \tag{2.2.5}$$

Definiţia 2.2.8. *(Momentul centrat de ordin r al unei variabile aleatoare continue)*

Momentul centrat de ordinul r al variabilei aleatoare continue X este

$$\mu_r(X) = E[(X - E[X])^r] = \int_{\mathbb{R}} (x - E[X])^r f_X(x)dx. \tag{2.2.6}$$

Pentru $r = 2$ obţinem dispersia variabilei aleatoare X:

$$D(X) = \mu_2(X) = E[(x - E[X])^2] = \int_{\mathbb{R}} (x - E[X])^2 f_X(x)dx. \tag{2.2.7}$$

$\sqrt{D(X)}$ se numeşte abaterea medie pătratică sau deviaţia standard.

Proprietăţile mediei şi dispersiei sunt aceleaşi ca şi pentru variabile aleatoare discrete.

Valoarea medie a unei v.a. este cel mai folosit indicator pentru locaţie sau poziţie. Există de asemenea şi modulul şi mediana.

Definiţia 2.2.9. *(Modulul unei variabile aleatoare)*

Modulul v.a. X, notat prin M_0, este acea valoare a v.a. X pentru care densitatea de probabilitate $f_X(x)$ are valoare maximă.

Definiţia 2.2.10. *(Mediana unei variabile aleatoare)*

Mediana v.a. X, notată prin Me, este acea valoare pentru care

$$\int_{-\infty}^{Me} f_X(x)dx = \int_{Me}^{\infty} f_X(x)dx = \frac{1}{2}. \tag{2.2.8}$$

Definiţia 2.2.11. *(Cuantila de ordinul α a unei variabile aleatoare)*

Cuantila de ordinul α a variabile aleatoare X este acea valoare q_α care verifică relaţia

$$\int_{-\infty}^{q_\alpha} f_X(x)dx = \alpha. \tag{2.2.9}$$

Exerciţiul 2.2.12. *(Test) Timpul de aşteptare la un ghişeu este o variabilă aleatoare T care urmează o repartiţie Exp(λ) cu timpul mediu de aşteptare de 15 minute. Să se afle probabilitatea ca un cetăţean să aştepte mai mult de 20 de minute.*

2.2.3 Funcţii de o variabilă aleatoare

Fie X o v.a. $X : \Omega \to \mathbb{R}$ care ia valori în $D \subset \mathbb{R}$ şi $\phi : D \to \mathbb{R}$, ϕ continuă. Atunci v.a. $Y = \phi(X) : \Omega \to \phi(X) \subset \mathbb{R}$ are repartiţia dată de

$$P(\{\omega \in \Omega | Y(\omega) \in A\} = P(\big\{\omega \in \Omega | X(\omega) \in \phi^{-1}(A)\big\})) \tag{2.2.10}$$

pentru orice $A \subset \phi(D)$ şi $\phi^{-1}(A) = \{z | \phi(z) \in A\}$.

Proprietăţi 2.2.13.

Dacă $X \sim N(m, \sigma^2)$ atunci $Y = e^X$ are densitatea de repartiţie:

$$f_Y(y) = \frac{1}{y\sigma\sqrt{2\pi}} e^{-\frac{(\ln y - m)^2}{2\sigma^2}}, y > 0 \tag{2.2.11}$$

Demonstraţie

$$F_Y(y) = P(\{\omega \in \Omega | Y(\omega) \leq y\}) = P(e^X \leq y) = P(X \leq \ln y) = F_X(\ln y) =$$
$$= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\ln y} e^{-\frac{(t-m)^2}{2\sigma^2}} dt.$$

Definiţia 2.2.14. *(Variabilă aleatoare lognormală)*

Variabila aleatoare Y având densitatea de probabilitate (2.2.11) se numeşte lognormală.

Proprietăţi 2.2.15.

Dacă $X \sim U(0,1)$ atunci variabila aleatoare

$$Y = \left(-\frac{1}{a} \ln(1 - X)\right)^{1/b}$$

cu $a, b > 0$ are densitatea de probabilitate

$$f(y) = aby^{b-1} e^{-ay^b}, \quad 0 < y < \infty.$$

Definiţia 2.2.16. *(Variabilă aleatoare Weibull)*

O variabilă aleatoare cu densitatea de probabilitate (2.2.15) se numeşte Weibull cu parametrii a şi b şi este notată $W(a, b)$.

Proprietăţi 2.2.17.

Fie X o variabilă aleatoare cu funcţia de repartiţie inversabilă $F(X)$ de tip continuu care este strict crescătoare. Atunci variabila aleatoare $Y = F(X)$ este repartizată uniform pe $(0, 1)$.

Demonstraţie

Funcţia de repartiţie a variabilei Y este

$$P(Y \leq y) = P(F(X) \leq y) = P(X \leq F^{-1}(y), 0 < y < 1) = F(F^{-1}(y)) = y,$$

$y \in (0, 1)$, ceea ce arată că $Y \sim U(0, 1)$.

Un rezultat pentru reciproca propoziţiei (2.2.17) este dat în următoarea propoziţie.

Proprietăţi 2.2.18.

Fie $Y \sim U(0,1)$ şi $F(x)$ o funcţie de repartiţie inversabilă, de tip continuu cu $F(a) = 0$ şi $F(b) = 1$ şi F strict crescătoare pe (a, b). Atunci v.a.

$$X = F^{-1}(Y) \tag{2.2.12}$$

este o v.a. continuă cu funcţia de repartiţie F.

Demonstraţie

Funcţia de repartiţie a v.a. X este

$$P(X \leq x) = P(F^{-1}(Y) \leq x) = P(Y \leq F(x))$$

deoarece F este strict crescătoare.

Cum $Y \sim U(0,1)$ rezultă $F_Y(y) = y$ pentru $y \in (0,1)$. Obţinem astfel

$$P(X \leq x) = F(x).$$

2.3 Exemple de repartiţii discrete

2.3.1 Repartiţia Bernoulli

Fie un eveniment aleator observabil A care are probabilitatea constantă $p = P(A) > 0$. Într-un experiment se poate produce A cu probabilitatea p sau evenimentul contrar A^C cu probabilitatea $q = 1 - p$. Un astfel de experiment se numeşte *probă Bernoulli*. Când se produce A spunem că s-a realizat un "succes", iar când A nu se produce spunem că avem un "eşec".

Observaţie 2.3.1.

Dacă $p = \dfrac{1}{2}$ atunci suntem în cazul particular al aruncării la întâmplare cu banul (tragerea la sorţi).

Asociem unei probe Bernoulli variabila aleatoare Z astfel încât $Z = 1$ dacă se produce A şi $Z = 0$ dacă se produce A^C. Variabila Z are repartiţia

$$Z : \begin{pmatrix} 0 & 1 \\ q & p \end{pmatrix}, E[Z] = p, V(Z) = pq = p(1-p). \tag{2.3.1}$$

Funcţia de repartiţie a lui Z este

$$F(x) = P(Z \leq x) = \begin{cases} 0, & \text{dacă} \quad x < 0 \\ q, & \text{dacă} \quad 0 \leq x < 1 \\ 1, & \text{dacă} \quad x \geq 1 \end{cases}. \tag{2.3.2}$$

2.3.2 Repartiţia binomială

Spunem că variabila discretă $X \in \mathbb{N}$ este o variabilă binomială $Binom(n, p), n \in \mathbb{N}^+, 0 < p < 1$ dacă $X =$ numărul de succese în n probe Bernoulli independente, adică

$$X = \sum_{i=1}^{n} Z_i \tag{2.3.3}$$

unde Z_i sunt variabile identic şi independent repartizate Bernoulli.

Repartiţia variabilei binomiale X este

$$X : \begin{pmatrix} 0 & 1 & 2 & \dots & x & \dots & n \\ q^n & npq^{n-1} & C_n^2 p^2 q^{n-2} & \dots & P(X=x) = C_n^x p^x q^{n-x} & \dots & p^n \end{pmatrix}, q = 1 - p. \tag{2.3.4}$$

Media şi dispersia variabilei X sunt date de formulele

$$\begin{aligned} E[X] &= E[(\sum_{i=1}^{n} Z_i)] = \sum_{i=1}^{n} E[Z_i] = np \\ V(X) &= V((\sum_{i=1}^{n} Z_i))) = \sum_{i=1}^{n} V[Z_i] = npq. \end{aligned} \tag{2.3.5}$$

Observaţie 2.3.2.

Variabila $Binom(n, p)$ are şi o interpretare în termeni de experiment cu urnă. Presupunem că într-o urnă avem A bile albe şi B bile negre, $A+B = N$. Presupunem că se realizează extrageri succesive din urnă şi după fiecare extragere se introduce bila extrasă la loc în urnă (experimentul cu bila întoarsă). Fie $p = A/N$ probabilitatea de a extrage o bilă albă într-o extracţie. De aici rezultă că X = numărul de bile albe în n extrageri succesive cu întoarcere, este o variabilă binomială $Binom(n, p)$.

Câteva exemple în care rezultatele unui experiment pot fi modelate folosind o variabilă aleatoare binomială.

1. Un medicament are probabilitatea de 0.90 de a vindeca o boală. Acesta este administrat la 100 de pacienţi, iar rezultatul pentru fiecare pacient este fie vindecat sau nevindecat. Dacă X este numărul de pacienţi vindecaţi, atunci X este o variabilă aleatoare binomială cu parametrii $(100, 0.9)$.

2. Institutul Naţional al sanătăţii mentale estimează că 20% dintre adulţii americani suferă de o afecţiune psihiatrică. Sunt selectaţi aleator 50 de adultţi. Dacă notăm cu X numărul de persoane cu afecţiuni, atunci X ia valori în concordanţă cu repartiţia binomială cu parametrii $(50, 0.2)$.

Exerciţiul 2.3.3.

Pentru exemplul 2, calculaţi probabilitatea ca cel mult 3 persoane selectate să sufere de o boală psihică.

2.3.3 Repartiţia geometrică

Variabila X are repartiţia $Geom(p), 0 < p < 1$ dacă X = numărul de eşecuri până la apariţia unui succes într-un şir oarecare de probe Bernoulli independente.

Repartiţia variabilei $Geom(p)$, X, este

$$X : \begin{pmatrix} 0 & 1 & 2 & \dots & x & \dots \\ p & pq & pq^2 & \dots & P(X=x) = pq^x & \dots \end{pmatrix}, q = 1-p. \tag{2.3.6}$$

Funcţia de repartiţie a variabilei X se poate calcula cu formula

$$F(x) = P(X \le x) = \sum_{i=0}^{x} pq^i = 1 - q^{x+1}, x = 0, 1, 2, \dots \tag{2.3.7}$$

Media şi dispersia variabilei X sunt date de formulele

$$\begin{aligned} E[X] &= \frac{q}{p} \\ V(X) &= \frac{q}{p^2}. \end{aligned} \tag{2.3.8}$$

Observaţie 2.3.4.

Interpretarea cu urnă din cazul repartiţiei $Binom(n,p)$ se poate adapta şi la cazul repartiţiei $Geom(p)$. Astfel, numărul X de bile negre extrase cu întoarcere până când se obţine o bilă albă, este o variabilă $Geom(p)$.

2.3.4 Repartiţia Pascal

Variabila X are repartiţia $Pascal(k,p), k \in \mathbb{N}^+, 0 < p < 1$ dacă X = numărul de eşecuri până la apariţia a k succese într-un şir oarecare de probe Bernoulli independente.

Repartiţia variabilei $Pascal(k,p)$, X, este

$$\begin{gathered} X : \begin{pmatrix} 0 & 1 & 2 & \dots & x & \dots \\ p^k & kp^kq & C_{k+1}^{k-1}p^kq^2 & \dots & P(X=x) = C_{x+k-1}^{k-1}p^kq^x & \dots \end{pmatrix}, \\ q = 1-p, x = 0, 1, 2, \dots . \end{gathered} \tag{2.3.9}$$

Media şi dispersia variabilei X sunt date de formulele

$$\begin{aligned} E[X] &= \frac{kq}{p} \\ V(X) &= \frac{kq}{p^2}. \end{aligned} \tag{2.3.10}$$

Observaţie 2.3.5.

Interpretarea cu urnă din cazul repartiţiei $Geom(p)$ se poate extinde şi la cazul repartiţiei $Pascal(k,p)$. Astfel, numărul X de bile negre extrase cu întoarcere până când se obţin k bile albe, este o variabilă $Pascal(k,p)$.

2.3.5 Repartiţia Poisson

O variabilă aleatoare X este o variabilă aleatoare Poisson cu parametrul $\lambda, \lambda > 0$ dacă are funcţia de probabilitate dată prin

$$P(X = x) = \frac{\lambda^x}{x!} e^{-\lambda}, \lambda > 0 \tag{2.3.11}$$

Media şi dispersia variabilei aleatoare Poisson X sunt

$$\begin{aligned} &E[X] = \lambda \\ &\text{şi} \\ &V(X) = \lambda. \end{aligned} \tag{2.3.12}$$

Repartiţia Poisson poate fi utilizată în numeroase aplicaţii. Câteva situaţii în care o variabilă aleatoare discretă poate avea o distribuţie Poisson sunt:

- numărul erorilor de tipografie dintr-o pagină;
- numărul concediilor dintr-o firmă în decursul unei luni
- numărul defectelor de-a lungul unui fir.

Observaţie 2.3.6.

Repartiţia Poisson este adesea utilizată pentru a aproxima repartiţia binomială. Pentru valori mari ale lui n şi mici ale lui p (deci valori moderate pentru np), numărul de succese apărute înn n probe poate fi aproximat de variabila aleatoare Poisson cu parametrul $\lambda = np$.

Exemplul 2.3.7.

Presupunem că numărul erorilor dintr-o pagină a unui manuscris urmează o distribuţie Poisson cu parametrul $\lambda = 0.25$. Calculăm probabilitatea ca o pagină să aibă cel puţin 2 erori precum urmează:

$$P(X \geq 2) = 1 - (P(X = 0) + P(X = 1)) = 1 - e^{-0.25} - 0.25 \cdot e^{-0.25} \approx 0.0265$$

2.4 Exemple de repartiţii continue

Vezi Capitolul 2 din [Martinez, Martinez (2002)].

2.5 Vectori aleatori

2.5.1 Funcţia de repartiţie. Densitatea de repartiţie.

Definiţia 2.5.1. *(Vector aleator)*

Un vector aleator multidimensional $X = (X_1, X_2, \ldots, X_k)$ este definit de funcţia $X : \Omega \to \mathbb{R}^k$ care asociază evenimentului $\omega \in \Omega$ vectorul

$$X(\omega) = (X_1(\omega), X_2(\omega), \ldots, X_k(\omega)) \,. \tag{2.5.1}$$

Astfel, $X : \Omega \to \mathbb{R}$ este vector aleator bidimensional dacă pentru orice pereche $(A_1, A_2) \in \mathbb{R}^2, (A_1, A_2) : A_1 \in (-\infty, x_1], A_2 \in (-\infty, x_2]$, evenimentul

$$\{\omega \in \Omega | X_1(\omega) \in A_1 \text{ şi } X_2(\omega) \in A_2\}$$

aparţine mulţimii $\mathcal{B}$.

Definiţia 2.5.2. *(Funcţie de repartiţie a unui vector aleator)*

Fie $X = (X_1, X_2, \ldots, X_k) : \Omega \to \mathbb{R}^k$ un vector aleator. Funcţia

$$F_X : \mathbb{R}^k \to [0, 1]$$

definită prin

$$F_X(x) = F_X(x_1, x_2, \ldots, x_k) = P(\{\omega \in \Omega | X_1(\omega) \le x_1, \ldots, X_k(\omega) \le x_k\})$$

se numeşte funcţia de repartiţie a vectorului X.

Proprietăţi 2.5.3.

1. $\lim\limits_{x_j \to -\infty} F_X(x_1, x_2, \ldots, x_k) = 0, \forall 1 \le j \le k$
2. $\lim\limits_{(x_1, \ldots, x_k) \to (\infty, \ldots, \infty)} F_X(x_1, x_2, \ldots, x_k) = 1$
3. $\lim\limits_{x_j \to \infty} F_X(x_1, \ldots, x_j, \ldots, x_k) = F_{(x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_k)}(x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_k)$

Definiţia 2.5.4. *(Densitate de probabilitate a unui vector aleator)*

Dacă există $f_X : \mathbb{R}^k \to \mathbb{R}$ astfel încât

$$F_X(x_1, x_2, \ldots, x_k) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_k} f_X(u_1, u_2, \ldots, u_k) du_1 du_2 \ldots du_k \tag{2.5.2}$$

atunci f se numeşte densitatea de repartiţie a vectorului X.

Proprietăţi 2.5.5.

1. $\displaystyle\int_{\mathbb{R}^k} f_X(x_1, \ldots, x_k) dx_1 \ldots dx_k = 1$
2. $f_X(x_1, \ldots, x_k) = \dfrac{\partial^k F(x_1, \ldots, x_k)}{\partial x_1, \ldots, \partial x_k}$
3. $P(X_1 \in A_1, X_2 \in A_2, \ldots, X_k \in A_k) = \displaystyle\int_{A_1 \times \ldots \times A_k} f_X(x_1, \ldots, x_k) dx_1 \ldots dx_k$

Capitolul 3

Noţiuni de probabilităţi şi statistică - Partea a III-a

3.1 Model probabilist

Fie X o v.a. cu densitatea $f(x, \theta), x \in \mathbb{R}, \theta \in \mathbb{R}$.

Definiţia 3.1.1. *(Model probabilist unidimensional)*

Mulţimea densităţilor de probabilitate $f(x, \theta), \theta \in \Theta \subset \mathbb{R}$, ce conţine parametrul necunoscut θ se numeşte model probabilist.

$$\{f(x, \theta) | x \in \mathbb{R}, \theta \in \Theta\}. \tag{3.1.1}$$

Fie $X = (X_1, X_2, \ldots, X_n)$ un vector aleator cu densitatea

$$f(x_1, x_2, \ldots, x_n; \theta), (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n, \theta \in \mathbb{R}^k.$$

Definiţia 3.1.2. *(Model probabilist multidimensional)*

Mulţimea densităţilor de probabilitate $f(x_1, x_2, \ldots, x_n; \theta)$ cu parametrul $\theta \in \Theta \subset \mathbb{R}^k$ se numeşte model probabilist.

$$\{f(x_1, x_2, \ldots, x_n; \theta) | \theta \in \Theta\}. \tag{3.1.2}$$

3.2 Noţiunea de selecţie. Model statistic.

Experimentul aleator se realizează pentru colectarea de date necesară pentru a obţine informaţii privind un anumit fenomen de interes. Pe baza datelor se emit concluzii care, în general, ies din sfera experimentului particular. Cercetătorii generalizează concluziile experimentului pentru clasa tuturor experimentelor similare. Acesta este cadrul statisticii inferenţiale. Problema acestui demers este că nu putem garanta corectitudinea concluziilor obţinute. Totuşi, folosind tehnici statistice, putem măsura şi administra gradul de incertitudine al rezultatelor.

Statistica inferenţială este o colecţie de metode care permit cercetătorilor să observe o submulţime a obiectelor de interes şi folosind informaţia obţinută pe baza acestor observaţii să facă afirmaţii sau inferenţe privind întreaga populaţie. Câteva dintre aceste metode sunt estimarea parametrilor unei populaţii, verificarea ipotezelor statistice şi estimarea densităţii de probabilitate.

Populaţia ţintă este definită ca fiind întreaga colecţie de obiecte sau indivizi despre care vrem să obţinem anumite informaţii. Populaţia ţintă trebuie binedefinită indicând ce constituie membrii acesteia (de exemplu, populaţia unei zone geografice, o anumită firmă care construieşte componente hardware, etc.) şi caracteristicile populaţiei (starea de sănătate, numărul de defecţiuni, etc.).

În majoritatea cazurilor este imposibil sau nerealist să observăm întreaga populaţie. Astfel, cercetătorii măsoară numai o parte a populaţiei ţintă, denumită eşantion. Pentru a face inferenţe privind întreaga populaţie, folosind informaţia obţinută pe baza populaţiei eşantion, este foarte important ca mulţimea eşantion să fie reprezentativă relativ la întreaga populaţie.

Definiţia 3.2.1. *(Mulţime de selecţie)*

O selecţie este o mulţime de variabile aleatoare $X_1, X_2, \ldots, X_n$ având aceeaşi densitate de probabilitate $f(x, \theta)$.

Datele observate (datele eşantionului) nu sunt considerate ca o mulţime de numere; ele reprezintă rezultatul unui experiment. Deoarece selecţia este o mulţime de variabile aleatoare asociate unui model probabilist, selecţia trebuie să aibă o repartiţie, pe care o vom numi repartiţie de selecţie.

Definiţia 3.2.2. *(Repartiţia unei selecţii)*

Repartiţia selecţiei $X_1, X_2, \ldots, X_n$ este definită ca fiind repartiţia vectorului $X = (X_1, X_2, \ldots, X_n)$, notată prin $f(x_1, x_2, \ldots, x_n; \theta)$.

Cea mai folosită formă de selecţie este selecţia aleatoare şi este bazată pe ideea experimentului aleator.

Definiţia 3.2.3. *(Selecţie aleatoare)*

Spunem că $X_1, X_2, \ldots, X_n$ este o selecţie aleatoare asupra v.a. X care are densitatea de probabilitate $f(x; \theta)$ dacă $X_1, X_2, \ldots, X_n$ sunt v.a. independente şi identic repartizate ca X.

$X_1, X_2, \ldots, X_n$ se numesc variabile de selecţie.

În cazul selecţiei aleatoare, densitatea de probabilitate comună a variabilelor de selecţie este

$$f(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} f(x_i; \theta)$$

O selecţie aleatoare poate fi construită prin repetarea unui experiment aleator de n ori.

Realizarea selecţiei aleatoare se notează prin $(x_1, x_2, \ldots, x_n)$ şi mulţimea tuturor realizărilor definesc spaţiul observaţiilor $S \equiv \mathbb{R}^n$.

Definiţia 3.2.4. *(Model statistic)*

Modelul probabilist

$$\{f(x;\theta), \theta \in \Theta)\}$$

împreună cu selecţia $X = (X_1, X_2, \ldots, X_n)$ definesc modelul statistic.

3.3 Statistici. Momente de selecţie

Definiţia 3.3.1. *(Statistică)*

Statistica este o funcţie $t_n : S \to \Theta \subset \mathbb{R}^k$ care nu depinde de nici un parametru.

Cele mai utilizate statistici sunt momentele de selecţie.

Definiţia 3.3.2. *(Moment de selecţie de ordinul r)*

Momentul de selecţie de ordinul r, notat m'_r, este definit prin

$$m'_r = \frac{1}{n}\sum_{j=1}^{n} X_j^r. \tag{3.3.1}$$

Momentul de selecţie de ordin unu, m'_1, se notează prin $\overline{X}_n$ şi se numeşte media de selecţie,

$$m'_1 = \overline{X}_n = \frac{1}{n}\sum_{j=1}^{n} X_j. \tag{3.3.2}$$

Proprietăţi 3.3.3.

Avem

$$E[m'_r] = E[X^r] = \mu'_r$$

şi în particular,

$$E[\overline{X}_n] = \mu, D(\overline{X}_n) = \frac{\sigma^2}{n}.$$

Definiţia 3.3.4. *(Convergenţă în probabilitate)*

Şirul de v.a. $(X_n)_n$ converge în probabilitate la v.a. X dacă

$$\lim_{n\to\infty} P\left(\{\omega \in \Omega, |X_n(\omega) - X(\omega)| < \epsilon\}\right) = 1. \tag{3.3.3}$$

Proprietăţi 3.3.5.

Avem relaţia

$$\overline{X}_n \xrightarrow[n\to\infty]{P} E[X] = \mu$$

Demonstraţie.
Rezultă din aplicarea inegalităţii lui Cebîşev descrisă în cele ce urmează. Pentru fiecare v.a. X de pătrat integrabilă şi fiecare $a > 0$, obţinem

$$P(|X - E[X]| \geq a) \leq \frac{D(X)}{a^2}. \tag{3.3.4}$$

Definiţia 3.3.6. *(Moment centrat de selecţie de ordin r)*

Momentul centrat de selecţie de ordin r se notează prin m_r şi este definit de

$$m_r = \frac{1}{n}\sum_{j=1}^{n}(X_j - \overline{X}_n)^r. \tag{3.3.5}$$

Cel mai utilizat moment centrat de selecţie este momentul de ordinul doi, numit dispersia de selecţie

$$m_2 = \frac{1}{n}\sum_{j=1}^{n}(X_j - \overline{X}_n)^2. \tag{3.3.6}$$

3.4 Estimarea parametrilor

Definiţia 3.4.1. *(Estimator. Estimaţie)*

Se numeşte estimator, variabila aleatoare

$$t_n(X) : \Omega \to \Theta, \tag{3.4.1}$$

unde $X = (X_1, X_2, \ldots, X_n)$, Ω este spaţiul de selecţie; $t_n(x), x \in S$, S spaţiul observaţiilor se numeşte estimaţie.

Definiţia 3.4.2. *(Estimator consistent)*

Un estimator $t_n = t_n(X)$ se numeşte consistent pentru θ dacă

$$\lim_{n\to\infty} P(|t_n - \theta| < \epsilon) = 1 \tag{3.4.2}$$

şi notăm $t_n \xrightarrow{P} \theta$.

Consistenţa unui estimator reprezintă o proprietate asimptotică a estimatorului. Un estimator bun pentru parametrul θ trebuie să aibă o repartiţie cu o valoare centrală în vecinătatea lui θ. Definiţia următoare cere ca estimatorul să aibă o valoare centrală în vecinătatea lui θ nu numai pentru valori mari ale lui n, ci pentru orice n.

Definiţia 3.4.3. *(Estimator nedeplasat)*

Estimatorul t_n se numeşte nedeplasat pentru θ dacă

$$E[t_n] = \theta. \tag{3.4.3}$$

Bibliografie

[Craiu (1998)] M. Craiu (1998), *Statistică matematică: teorie şi probleme*, Editura Matrix Rom, Bucureşti

[Martinez, Martinez (2002)] W. L. Martinez, A. R. Martinez (2002), *Computational Statistics Handbook with MATLAB*, Chapman & Hall/CRC, Boca Raton London New York Washington, D.C.

[Văduva (2004)] I. Văduva (2004), *Modele de simulare: note de curs*, Editura Universităţii din Bucureşti, Bucureşti