

Curs 4

Variabile aleatoare

4.1 Caracteristici numerice ale variabilelor aleatoare discrete

Definiția 4.1.1 Se spune că o variabilă aleatoare discretă X , având matricea de repartiție de forma (??) admite medie și dispersie dacă seriile $\sum_{n \geq 1} x_i p_i$ și $\sum_{n \geq 1} x_i^2 p_i$ sunt convergente. În acest caz se definesc **media** lui X

$$M[X] = \sum_{n \geq 1} x_i p_i = m \text{ (suma seriei),}$$

momentul de ordin doi

$$M[X^2] = \sum_{n \geq 1} x_i^2 p_i$$

și **dispersia**

$$D[X] = M[X^2] - [M[X]]^2 = \sum_{n \geq 1} x_i^2 p_i - \left(\sum_{n \geq 1} x_i p_i \right)^2.$$

Dacă X are un număr finit de valori, atunci evident X are medie și dispersie și sumele anterioare sunt finite.

Definiția 4.1.2 Se spune că o variabilă aleatoare discretă X , având matricea de repartiție de forma (??) admite **moment de ordin k** dacă seria $\sum_{n \geq 1} x_i^k p_i$ este convergentă. În acest caz se definesc **momentul de ordin k** al lui X

$$M[X^k] = \sum_{n \geq 1} x_i^k p_i$$

și **momentul centrat de ordin k** al v.a. X

$$M[(X - M[X])^k] = \sum_{n \geq 1} (x_i - M[X])^k p_i.$$

Definiția 4.1.3 Se numește **abarere media pătratică** $\sigma = \sqrt{D[X]}$.

Exercițiul 1 Demonstrați că momentul centrat de ordin doi este chiar dispersia, adică $D[X] = \sum_{n \geq 1} (x_i - M[X])^2 p_i$.

Definiția 4.1.4 Variabila aleatoare $X - M[X]$ se numește **abaterea de la medie** a variabilei aleatoare X .

Propoziția 4.1.5 Valoarea medie a abaterii oricărei variabile aleatoare este nulă.

Exercițiul 2 Variabilele aleatoare discrete simple X_1 și X_2 , având repartițiile:

$$X_1, X_2 : \begin{pmatrix} 0 & 1 \\ q & p \end{pmatrix},$$

au valoarea medie

$$M[X_i] = 1p + 0q = p, \quad i = 1, 2,$$

iar

$$X_1 + X_2 : \begin{pmatrix} 0 & 1 & 2 \\ q^2 & 2pq & p^2 \end{pmatrix},$$

$$X_1 X_2 : \begin{pmatrix} 0 & 1 \\ 2pq + q^2 & p^2 \end{pmatrix}$$

au media

$$M[X_1 + X_2] = 2pq + 2p^2 = 2p, \quad M[X_1 X_2] = p^2.$$

Calculăm și dispersiile:

$$M[X_i^2] = p, \quad D[X_i] = p^2 - p = pq,$$

$$M[(X_1 + X_2)^2] = 1^2 2pq + 2^2 p^2, \quad D[X_1 + X_2] = 2pq + 4p^2 - 4p^2 = 2pq,$$

$$D[X_1 X_2] = M[(X_1 X_2)^2] - (M[X_1 X_2])^2 = p^2 - p^4 = p^2 q^2.$$

Observația 4.1.6 Denumirea de medie este îndreptățită dacă ținem seama de sensul ei practic. Presupunem că am repetat de N ori o experiență care ne-a condus la variabila aleatoare X . Dacă valoarea x_1 este luată de n_1 ori, valoarea x_2 de n_2 ori, ..., valoarea x_m de n_m ori, unde $n_1 + n_2 + \dots + n_m = N$, atunci suma valorilor luate de variabila aleatoare X este $n_1 x_1 + n_2 x_2 + \dots + n_m x_m$, iar media aritmetică a valorilor luate de X va fi

$$\frac{n_1 x_1 + n_2 x_2 + \dots + n_m x_m}{N} = \frac{n_1}{N} x_1 + \frac{n_2}{N} x_2 + \dots + \frac{n_m}{N} x_m.$$

Dar $\frac{n_1}{N} = p_1$ reprezintă raportul dintre numărul cazurilor în care s-a luat valoarea x_1 și numărul total de experimentări, adică p_1 . La fel $\frac{n_2}{N} = p_2, \dots, \frac{n_m}{N} = p_m$. Media lui X ne arată la ce valoare putem să ne așteptăm pentru media aritmetică a unui mare număr de valori ale lui X , obținute în urma repetării experienței date.

Proprietăți ale mediei

1. Valoarea medie a unei constante este egală cu constanta,

$$X : \begin{pmatrix} c \\ 1 \end{pmatrix}, \quad M[X] = c.$$

2. Valoarea medie a unei variabile aleatoare care admite medie este cuprinsă între cea mai mică și cea mai mare dintre valorile posibile ale variabilei aleatoare. Fie $a = \inf_i x_i$ și $A = \sup_i x_i$. Atunci

$$M[X] = \sum_{i \geq 1} x_i p_i \geq a \sum_{i \geq 1} p_i = a,$$

$$M[X] = \sum_{i \geq 1} x_i p_i \leq A \sum_{i \geq 1} p_i = A, \quad \Leftrightarrow a \leq M[X] \leq A.$$

3. Dacă X, Y sunt variabile aleatoare discrete care admit medie atunci

$$M[X + Y] = M[X] + M[Y]$$

Într-adevăr,

$$\begin{aligned} M[X + Y] &= \sum_{i \geq 1} \sum_{j \geq 1} (x_i + y_j) p_{ij} = \sum_{i \geq 1} x_i \sum_{j \geq 1} p_{ij} + \sum_{j \geq 1} y_j \sum_{i \geq 1} p_{ij} \\ &= \sum_{i \geq 1} x_i p_i + \sum_{j \geq 1} y_j q_j = M[X] + M[Y]. \end{aligned}$$

4. Dacă X, Y sunt variabile aleatoare discrete care admit medie sunt independente atunci media produsului este egală cu produsul mediilor variabilelor considerate.

$$M[XY] = M[X]M[Y].$$

Fie

$$M[XY] = \sum_{i \geq 1} \sum_{j \geq 1} x_i y_j p_{ij}.$$

Dacă variabilele sunt independente, se produc simplificări importante deoarece putem scrie $p_{ij} = p_i q_j$ și atunci

$$M[XY] = \sum_{i \geq 1} \sum_{j \geq 1} x_i y_j p_i q_j = \left(\sum_{i \geq 1} x_i p_i \right) \left(\sum_{j \geq 1} y_j q_j \right) = M[X]M[Y].$$

5. Dacă X este o variabilă aleatoare discretă care admite medie și $a \in \mathbb{R}$ o constantă, atunci au loc relațiile:

$$M[a + X] = a + M[X], \quad M[aX] = aM[X].$$

Proprietăți ale dispersiei

1. Dispersia unei constante este nulă deoarece, ținând seama de proprietățile mediei,

$$D[c] = M[c^2] - (M[c])^2 = 0.$$

2. $D[aX] = a^2 D[X]$ deoarece

$$D[aX] = \sum_{i \geq 1} p_i (ax_i - am)^2 = \sum_{i \geq 1} p_i a^2 (x_i - m)^2 = a^2 \sum_{i \geq 1} p_i (x_i - m)^2 = a^2 D[X].$$

3. Dacă variabilele aleatoare X_1, X_2, \dots, X_k sunt independente, atunci dispersia sumei este egală cu suma dispersiilor.

$$D[X_1 + X_2 + \dots + X_k] = D[X_1] + D[X_2] + \dots + D[X_k].$$

Demonstrăm pentru două variabile aleatoare. Într-adevăr, fie $Y = X_1 + X_2$ atunci

$$\begin{aligned} D[Y] &= M[Y^2] - (M[Y])^2 = M[(X_1 + X_2)^2] - (M[X_1 + X_2])^2 \\ &= M[X_1^2 + X_2^2 + 2X_1X_2] - (M[X_1] + M[X_2])^2 \\ &= M[X_1^2] + M[X_2^2] + 2M[X_1X_2] - (M[X_1])^2 - 2M[X_1]M[X_2] - (M[X_2])^2 \\ &= M[X_1^2] - (M[X_1])^2 + M[X_2^2] - (M[X_2])^2 = D[X_1] + D[X_2] \end{aligned}$$

4. Două variabile aleatoare care diferă printr-o constantă au dispersiile egale.

Mai general, combinând Proprietățile 2 și 3, avem

$$D\left[\sum_{i=1}^k a_i X_i\right] = \sum_{i=1}^k a_i^2 D[X_i].$$

În particular,

$$D[X - Y] = D[X] + D[Y].$$

Arătăm că dispersia unei variabile aleatoare X este, într-un anumit sens, cea mai bună valoare care caracterizează împrăștierea valorilor x_1, x_2, \dots, x_n sau, cu alte cuvinte, $M[X]$ este punctul cel mai potrivit față de care trebuie să măsurăm devierile acestor valori. Acest lucru rezultă din următoarea propoziție.

Propoziția 4.1.7 Dacă X este o variabilă aleatoare care ia valorile $x_1, x_2, \dots, x_n, \dots$ cu probabilitățile $p_1, p_2, \dots, p_n, \dots$, iar m este media, atunci

$$D[X] = \sum_{i \geq 1} p_i (x_i - m)^2 \leq \sum_{i \geq 1} p_i (x_i - x)^2, \quad \forall x \in \mathbb{R}.$$

Demonstrație. Putem scrie $x_i - x = (x_i - m) + (m - x)$, $i = \overline{1, n}$ și

$$\begin{aligned} p_i (x_i - x)^2 &= \sum_{i \geq 1} p_i [(x_i - m) + (m - x)]^2 \\ &= \sum_{i \geq 1} p_i (x_i - m)^2 + 2(m - x) \sum_{i \geq 1} p_i (x_i - m) + (m - x)^2 \\ &= \sum_{i \geq 1} p_i (x_i - m)^2 + 2(m - x)(m - m) + (m - x)^2 = D[X] + (m - x)^2 \leq D[X]. \quad \square \end{aligned}$$

Definiția 4.1.8 Fie X și Y două variabile aleatoare care admit medie și dispersie. **Covarianța** (sau **corelația**) lui X și Y este numărul real

$$\text{cov}(X, Y) = M[(X - M[X])(Y - M[Y])].$$

În cazul în care $\text{cov}(X, Y) = 0$ se spune că X și Y sunt **necorelate**.

Teorema 4.1.9 Fie X și Y două variabile aleatoare care admit medie. Atunci:

a) $\text{cov}(X, Y) = M[XY] - M[X]M[Y]$.

b) dacă X și Y sunt independente, atunci X și Y sunt necorelate. Reciproca nu este adevărată.

Demonstrație. a) $\text{cov}(X, Y) = M[XY - YM[X] - XM[Y] + M[X]M[Y]]$. Notând

$$M[X] = m_1, M[Y] = m_2,$$

atunci

$$\begin{aligned} \text{cov}(X, Y) &= M[XY - Ym_1 - Xm_2 + m_1m_2] \\ &= M[XY] - m_1M[X] - m_2M[Y] + m_1m_2 = M[XY] - M[X]M[Y]. \end{aligned}$$

b) Dacă X și Y sunt independente, atunci $M[XY] = M[X]M[Y] \Rightarrow \text{cov}(X, Y) = 0$. \square

Teorema 4.1.10 Fie X și Y două variabile aleatoare care admit medie și dispersie. Atunci: $D[X + Y] = D[X] + D[Y] + 2 \text{cov}(X, Y)$.

Demonstrație. Avem

$$\begin{aligned} D[X + Y] &= M[(X + Y - M[X + Y])^2] = M[((X - M[X]) + (Y - M[Y]))^2] = \\ &= M[(X - M[X])^2 + (Y - M[Y])^2 + 2(X - M[X])(Y - M[Y])] \\ &= M[(X - M[X])^2] + M[(Y - M[Y])^2] + 2M[(X - M[X])(Y - M[Y])] \\ &= D[X] + D[Y] + 2 \text{cov}(X, Y). \quad \square \end{aligned}$$

Teorema 4.1.11 (Inegalitatea lui Cebâșev) Fie X o variabilă aleatoare care admite medie și dispersie finite. Atunci, oricare ar fi $\varepsilon > 0$, are loc inegalitatea

$$P(\{|X - m| < \varepsilon\}) \geq 1 - \frac{D[X]}{\varepsilon^2}, \quad m = M[X]. \quad (4.1)$$

Demonstrație. O vom face în cazul continuu. □

O formă echivalentă a teoremei este: dacă media distribuției X este m și abaterea medie pătratică este $\sigma = \sqrt{D[X]}$, probabilitatea ca valorile variabilei aleatoare să fie mai departate de medie cu cel puțin $k\sigma$ este cel mult $\frac{1}{k^2}$.

$$P(\{|X - m| \geq k\sigma\}) < \frac{1}{k^2}. \quad (4.2)$$

Dacă luăm $\varepsilon = k\sigma$ această inegalitatea (4.1) se obține $P(\{|X - m| < k\sigma\}) \geq 1 - \frac{1}{k^2}$ sau $P(\{m - k\sigma < X < m + k\sigma\}) \geq 1 - \frac{1}{k^2}$. Dar evenimentele $\{|X - m| < k\sigma\}$ și $\{|X - m| \geq k\sigma\}$ sunt contrare, rezultă (4.2).

De exemplu, pentru $k = 3$ obținem $P(\{m - 3\sigma < X < m + 3\sigma\}) \geq 1 - \frac{1}{9} = \frac{8}{9} = 0.88889$. Deci, cu o probabilitate cuprinsă între 0.88889 și 1, orice variabilă ia valori cuprinse în intervalul $(m - 3\sigma, m + 3\sigma)$.

Exercițiul 3 Numărul clienților care vizitează un showroom sâmbăta dimineață este o variabilă aleatoare cu media $m = 18$ și $\sigma = 2, 5$. Cu ce probabilitate putem presupune că numărul vizitatorilor va fi cuprins între 8 și 28 de clienți.

Deoarece $8 < X < 28 \Rightarrow -10 < X - 18 < 10 \Rightarrow \varepsilon = 10$ și folosind relația (4.1) obținem: $P(\{|X - 18| < 10\}) \geq 0.9375$, deci probabilitatea ca numărul vizitatorilor să fie cuprins între 8 și 28 de clienți este de cel puțin 0.9375.

Avantajul inegalității lui Cebâșev este că se poate aplica oricărei distribuții pentru care media și abaterea medie pătratică există.

Dezavantajul este că această inegalitate dă o margine inferioară a probabilității ca valorile variabilei aleatoare să fie cuprinse într-un interval de forma $(m - \varepsilon, m + \varepsilon)$.