Metode numerice și statistică
Curs 9

lect. Ciprian Deliu
cdeliu@tuiasi.ro
moodle.deliu.ro

Universitatea Tehnică "Gh. Asachi" Iași
Facultatea de Hidrotehnică, Geodezie și Ingineria Mediului

2019

Variabile aleatoare

- După repetarea unui experiment de un număr mare de ori intervine o anumită regularitate în privința apariției unor rezultate ale acestuia, dar nu se poate preciza niciodată cu certitudine care anume dintre rezultate va apare într-o anumită probă
- Din acest motiv, conceptul de "**aleator**" trebuie înțeles în sensul că avem de a face cu experimente sau fenomene guvernate de legi *probabilistice* (atunci când există un anumit grad de incertitudine privind apariția unui rezultat sau reapariția lui) și nu de legi *deterministe* (când știm cu certitudine ce rezultat va apare sau nu).
- Pentru studierea unor astfel de experimente sau fenomene sunt importante următoarele două aspecte:
  1. rezultatele posibile ale experimentului, care pot constitui o mulțime finită, infinită și numărabilă, sau infinită și nenumărabilă.
  2. legea statistică (sau probabilitățile) după care apar rezultatele experimentului considerat.
- În înțeles mai larg, o mărime care ia valori aleatoare (la întâmplare) dintr-o mulțime oarecare posibilă se numește *variabilă aleatoare*.

Variabile aleatoare discrete

Definiție

Fie câmpul de probabilitate $(\Omega, \mathcal{K}, P)$. Se numește **variabilă aleatoare de tip discret** *o funcție $X : \Omega \to \mathbb{R}$ care verifică:*
1. *are o mulțime cel mult numărabilă de valori;*
2. $\{\omega \in \Omega | X(\omega) = x\} \in \mathcal{K}, \ \forall x \in \mathbb{R}$.

Notăm valorile pe care le ia X cu $x_1 < x_2 < \dots < x_n < \dots$, iar evenimentele corespunzătoare acestor valori cu $(X = x_i) = \{\omega \in \Omega | X(\omega) = x_i\}, \ i \in \mathbb{N}$.

Definiție

Numim **distribuția** *sau* **repartiția** *variabilei aleatoare discrete X un tablou de forma $X : \begin{pmatrix} x_i \\ p_i \end{pmatrix}_{i \in I}$ unde $p_i = P(X = x_i), i \in I$ sunt probabilitățile cu care X ia valorile $x_i, i \in I$, mulțimea I fiind finită sau cel mult numărabilă.*

Evenimentele $(X = x_i), \ i \in I$ formează un sistem complet de evenimente și
$$\sum_{i \in I} p_i = 1.$$

Funcție de repartiție

Definiție

Fie câmpul de probabilitate $(\Omega, \mathcal{K}, P)$ și variabila aleatoare discretă $X : \Omega \to \mathbb{R}$. Se numește **funcție de repartiție** *a lui X funcția $F : \mathbb{R} \to [0,1]$ definită prin*
$$F(x) = P(X \le x) = \sum_{x_i \le x} P(X = x_i), \ x \in \mathbb{R}.$$

Proprietăți:
1. F este funcție crescătoare pe $\mathbb{R}$:
$$\forall x_1, x_2 \in \mathbb{R}, x_1 < x_2 \Rightarrow F(x_1) \le F(x_2)$$
2. $\forall a, b \in \mathbb{R}, a < b$ avem $P(a < X \le b) = F(b) - F(a)$.
3. $\lim_{x \to -\infty} F(x) = 0, \ \lim_{x \to \infty} F(x) = 1$
4. F este continuă la stânga în orice punct $x \in \mathbb{R}$:
$$F(x - 0) = F(x), \ \forall x \in \mathbb{R}.$$

Variabile aleatoare continue

Definiție

Fie câmpul de probabilitate $(\Omega, \mathcal{K}, P)$. Se numește **variabilă aleatoare de tip continuu** *o funcție $X : \Omega \to \mathbb{R}$ care verifică:*
1. *are o mulțime nenumărabilă de valori (conține un interval de numere reale);*
2. $\{\omega \in \Omega | X(\omega) \le x\} \in \mathcal{K}, \ \forall x \in \mathbb{R}$.

Definiție

Fie câmpul de probabilitate $(\Omega, \mathcal{K}, P)$ și variabila aleatoare continuă $X : \Omega \to \mathbb{R}$. Se numește **funcție de repartiție** *a lui X funcția $F : \mathbb{R} \to [0,1]$ definită prin*
$$F(x) = P(X \le x) = \int_{-\infty}^{x} \rho(t) dt, \ x \in \mathbb{R}.$$
Funcția $\rho : \mathbb{R} \to \mathbb{R}$ se numește **densitate de probabilitate** *a variabilei aleatoare X.*

Proprietăți:
1. $\rho(x) \ge 0, \forall x \in \mathbb{R}$
2. $F'(x) = \rho(x), \forall x \in \mathbb{R}$
3. $P(a \le X < b) = \int_a^b \rho(t) dt$
4. $\int_{-\infty}^{\infty} \rho(t) dt = 1.$

Vectori aleatori bidimensionali

Definiție

Spunem că variabilele aleatoare discrete X și Y care au respectiv distribuțiile $X : \begin{pmatrix} x_i \\ p_i \end{pmatrix}_{i \in I}$ și $Y : \begin{pmatrix} y_j \\ q_j \end{pmatrix}_{j \in J}$ sunt **independente** *dacă*
$$P(X = x_i, Y = y_j) = P(X = x_i) \cdot P(Y = y_j), \ \forall (i,j) \in I \times J.$$

Teoremă

Fie variabilele aleatoare discrete X și Y care au respectiv distribuțiile $X : \begin{pmatrix} x_i \\ p_i \end{pmatrix}_{i \in I}$ și $Y : \begin{pmatrix} y_j \\ q_j \end{pmatrix}_{j \in J}$. Atunci variabilele aleatoare sumă $X + Y$, produs $X \cdot Y$ și cât $\frac{X}{Y}$ (dacă $y_j \neq 0, \forall j \in J$) au distribuțiile
$$X + Y : \begin{pmatrix} x_i + y_j \\ p_{ij} \end{pmatrix}_{(i,j) \in I \times J}, \ X \cdot Y : \begin{pmatrix} x_i y_j \\ p_{ij} \end{pmatrix}_{(i,j) \in I \times J}, \ \frac{X}{Y} : \begin{pmatrix} \frac{x_i}{y_j} \\ p_{ij} \end{pmatrix}_{(i,j) \in I \times J},$$
unde $p_{ij} = P(X = x_i, Y = y_j), \ \forall (i,j) \in I \times J$.

- $\sum_{j \in J} p_{ij} = p_i, \ \forall i \in I$ și $\sum_{i \in I} p_{ij} = q_j, \forall j \in J$.
- Dacă X și Y sunt independente, atunci $p_{ij} = p_i q_j, \ \forall (i,j) \in I \times J$.

Definiție

Fie câmpul de probabilitate $(\Omega, \mathcal{K}, P)$. Spunem că $U = (X, Y)$ este un **vector aleator bidimensional** *de tip discret dacă funcția $U : \Omega \to \mathbb{R}^2$ îndeplinește condițiile:*
1. *are o mulțime cel mult numărabilă de valori;*
2. $\forall (x,y) \in \mathbb{R}^2, \ (X = x, Y = y) = \{\omega \in \Omega | X(\omega) = x, Y(\omega) = y\} \in \mathcal{K}$.

Definiție

Numim **distribuția** *sau* **repartiția** *vectorului aleator discret (X, Y) tabloul:*

$X \setminus Y$	y_1	$\dots$	y_j	$\dots$
x_1	p_{11}	$\dots$	p_{1j}	$\dots$
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$
x_i	p_{i1}	$\dots$	p_{ij}	$\dots$
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$

unde (x_i, y_j) sunt valorile pe care le ia vectorul (X, Y), iar $p_{ij} = P(X = x_i, Y = y_j)$. Evident, $\sum_{(i,j) \in I \times J} p_{ij} = 1$.

Funcția de repartiție a unui vector aleator bidimensional

În mod analog se definesc vectorii aleatori bidimensionali de tip continuu.

Definiție

Se numește **funcție de repartiție** *a vectorului aleator bidimensional funcția $F : \mathbb{R}^2 \to [0,1]$ definită prin:*
$$F(x,y) = P(X \le x, Y \le y), \ \forall (x,y) \in \mathbb{R}^2.$$

Proprietăți:
1. dacă $a < b$ și $c < d$, atunci
$$P(a < X \le b, c < Y \le d) = F(b,d) - F(a,c)$$
2. F este monoton crescătoare în raport cu fiecare argument.
3. $\lim_{x \to -\infty} F(x,y) = \lim_{y \to -\infty} F(x,y) = \lim_{x,y \to -\infty} F(x,y) = 0$ și $\lim_{x,y \to \infty} F(x,y) = 1$.
4. F este continuă la stânga în raport cu fiecare argument.
5. Dacă cele două componente X și Y ale vectorului aleator au funcțiile de repartiție F_X și F_Y, atunci
$$F_X(x) = \lim_{y \to \infty} F(x,y) \text{ și } F_Y(y) = \lim_{x \to \infty} F(x,y)$$

În cazul continuu, funcția de repartiție a unui vector aleator (X, Y) se poate scrie sub forma
$$F(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} \rho(s,t) ds dt, \ \forall (x,y) \in \mathbb{R}^2.$$

Funcția $\rho : \mathbb{R}^2 \to \mathbb{R}$ se numește **densitate de probabilitate** a vectorului aleator (X, Y).

Proprietăți:
1. $\rho(x,y) \ge 0, \ \forall (x,y) \in \mathbb{R}^2$;
2. $\frac{\partial^2 F}{\partial x \partial y}(x,y) = \rho(x,y), \ \forall (x,y) \in \mathbb{R}^2$;
3. $P((X,Y) \in D) = \iint_D \rho(x,y) dx dy, \ D \subset \mathbb{R}^2$;
4. $\iint_{\mathbb{R}^2} \rho(x,y) dx dy = 1$;
5. $\rho_X(x) = \int_{-\infty}^{\infty} \rho(x,y) dy, \ \forall x \in \mathbb{R}$ și $\rho_Y(y) = \int_{-\infty}^{\infty} \rho(x,y) dx, \ \forall y \in \mathbb{R}$, unde ρ_X și ρ_Y sunt densitățile celor două componente X și Y ale vectorului aleator.

Definiție

Spunem că variabilele aleatoare de tip continuu X și Y sunt **independente** *dacă $F(x,y) = F_X(x) \cdot F_Y(y), \ \forall (x,y) \in \mathbb{R}^2$.*

Exerciții

1. Fie variabila aleatoare discretă $X : \begin{pmatrix} 1 & 2 & 3 & 4 \\ p^2 & \frac{7}{4}p & \frac{1}{3} & \frac{1}{6} \end{pmatrix}$. Care este $P(X \le 3)$?
2. Fie v. a. independente $X : \begin{pmatrix} 0 & 1 & 2 \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \end{pmatrix}$, $Y : \begin{pmatrix} -1 & 1 & 2 \\ \frac{1}{3} & \frac{1}{6} & \frac{1}{2} \end{pmatrix}$. Să se scrie variabilele aleatoare $2X,\ Y^2,\ X+Y,\ XY,\ \frac{X}{Y},\ \sqrt{X},\ 2X+3Y$.
3. Fie v. a. independente $X : \begin{pmatrix} -1 & 0 & 1 \\ p+\frac{1}{6} & q+\frac{1}{3} & \frac{1}{3} \end{pmatrix}$,
   $Y : \begin{pmatrix} -1 & 0 & 1 \\ \frac{1}{3} & 2p-q & 12p^2 \end{pmatrix}$.
   a) Să se scrie distribuția variabilei aleatoare $3XY$
   b) Pentru ce valori ale lui c avem $P(X+Y=c) > \frac{2}{9}$?
4. Să se scrie funcțiile de repartiție pentru variabilele aleatoare
   $X : \begin{pmatrix} -1 & 0 & 1 \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{6} \end{pmatrix}$, $Y : \begin{pmatrix} \frac{1}{2} & 1 & 2 \\ \frac{1}{6} & \frac{1}{2} & \frac{1}{3} \end{pmatrix}$.

Valoare medie

Definiție

Fie $(\Omega, \mathcal{K}, P)$ un câmp de probabilitate și variabila aleatoare discretă $X : \Omega \to \mathbb{R}$ cu distribuția $X : \begin{pmatrix} x_i \\ p_i \end{pmatrix}_{i \in I}$. Se numește valoare **medie** *a lui X caracteristica numerică*

$$E(X) = \sum_{i \in I} x_i p_i$$

Definiție

Fie $(\Omega, \mathcal{K}, P)$ un câmp de probabilitate și variabila aleatoare continuă $X : \Omega \to \mathbb{R}$ cu densitatea $\rho(x)$. Se numește valoare **medie** *a lui X caracteristica numerică*

$$E(X) = \int_{-\infty}^{\infty} x\rho(x)dx$$

1. $E(aX+b) = aE(X)+b,\ \forall a,b \in \mathbb{R}$;
2. $E(X+Y) = E(X)+E(Y)$;
3. X, Y independente $\Rightarrow E(XY) = E(X)E(Y)$.

Dispersie

Definiție

Fie $(\Omega, \mathcal{K}, P)$ un câmp de probabilitate și variabila aleatoare $X : \Omega \to \mathbb{R}$. Se numește **dispersia** *(varianța) lui X caracteristica numerică*

$$Var(X) = E\left[(X - E(X))^2\right]$$

iar $\sigma(X) = \sqrt{Var(X)}$ se numește **abatere medie pătratică**.

Dacă X este o variabilă aleatoare discretă, atunci

$$Var(X) = \sum_{i \in I} (x_i - E(X))^2 \cdot p_i.$$

Dacă X este o variabilă aleatoare continuă, atunci

$$Var(X) = \int_{-\infty}^{\infty} (x - E(X))^2 \rho(x)dx.$$

Proprietăți:

1. $Var(X) = E(X^2) - [E(X)]^2$;
2. $Var(aX+b) = a^2 Var(X),\ \forall a,b \in \mathbb{R}$;
3. X, Y independente $\Rightarrow Var(X+Y) = Var(X) + Var(Y)$.

Corelație

Definiție

Se numește **corelația** *sau* **covarianța** *variabilelor aleatoare X și Y caracteristica numerică*

$$C(X,Y) = E\left[(X - E(X))(Y - E(Y))\right].$$

- $C(X,Y) = E(XY) - E(X)E(Y)$; $C(X,X) = Var(X)$
- dacă X, Y sunt independente atunci $C(X,Y) = 0$ dar nu și reciproc.

Definiție

Se numește **coeficient de corelație** *al variabilelor aleatoare X și Y caracteristica numerică*

$$r(X,Y) = \frac{C(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}}.$$

- dacă X, Y sunt independente atunci $r(X,Y) = 0$ dar nu și reciproc;
- dacă $r(X,Y) = 0$, spunem că X, Y sunt *necorelate*;
- $|r(X,Y)| \le 1$;
- $r(X,Y) = 1 \Leftrightarrow Y = aX + b,\ a > 0$;

Exerciții

1. Să se calculeze media și dispersia variabilei aleatoare
   $X : \begin{pmatrix} 0 & 1 & 2 & 3 \\ \frac{1}{10} & \frac{1}{5} & \frac{3}{10} & \frac{1}{5} \end{pmatrix}$.
2. Fie variabila aleatoare $X : \begin{pmatrix} 1 & 3 & 4 \\ p_1 & p_2 & p_3 \end{pmatrix}$ având media $E(X) = \frac{8}{3}$ și dispersia $Var(X) = \frac{14}{9}$. Să se determine p_1, p_2, p_3.
3. Să se determine variabilele aleatoare independente
   $X : \begin{pmatrix} x & x+1 & x+2 & x+3 \\ p & 2p & 3p & 4p \end{pmatrix}$ și $Y : \begin{pmatrix} y & 2y & 3y \\ q & q^2 & q^2 \end{pmatrix}$ știind că $E(X) = 2$ și $E(Y) = 7$. Să se calculeze apoi media și dispersia variabilei aleatoare $2X + 3Y$.
4. Să se calculeze coeficientul de corelație al variabilelor aleatoare
   $X : \begin{pmatrix} -2 & 1 & 0 \\ 0.3 & 0.5 & 0.2 \end{pmatrix}$, $Y : \begin{pmatrix} -1 & 1 & 2 \\ 0.1 & 0.4 & 0.5 \end{pmatrix}$.