Curs 6: Variabile aleatoare continue: Definiție. Proprietăți

Conf.dr. Maria Jivulescu

Departamentul de Matematică
UPT

Cuprins

Variabile aleatoare continue

- Definiție. Exemple
- Funcția de repartiție.
- Medie. Dispersie

Variabile aleatoare: Definiție

O variabilă aleatoare (v.a.) este o funcție ce atribuie un număr real fiecărui rezultat din spațiul tuturor realizărilor unui experiment aleator.

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

Variabilă aleatoare discretă: are valori într-o mulțime finită sau infinit numărabilă

- număr de biți transmiși cu eroare într-un canal de comunicație

Variabilă aleatoare continuă: poate lua orice valoare dintr-un interval din. $\mathbb{R}$ (mărginit sau nu)

- timpul de execuție a unui program
- durata de viață a unei componente electronice
- frecvența de acces în traficul pe WEB
- dimensiunea pachetelor de date în FTP (File Transfer Protocol)

Densitate de probabilitate

Definiție

Se numește **densitate de probabilitate** o funcție $f : \mathbb{R} \to \mathbb{R}$ cu proprietățile:

- $f(x) \geq 0, \quad x \in \mathbb{R}.$
- f este integrabilă pe $\mathbb{R}$ și $\int\limits_{-\infty}^{\infty} f(x)\,dx = 1$

O variabilă aleatoare continuă este o variabilă aleatoare pentru care distribuția de probabilitate este definită de o densitate de probabilitate (p.d.f.), f_X.

Densitatea de probabilitate

Ce putem face cu p.d.f?

Ne interesează: $P(X \in I)$, unde
$I = [a, b], (a, b), (a, b], (a, b); (-\infty, b], (-\infty, b), [a, \infty), (a, \infty)$
Relația dintre p.d.f. și probabilitatea unui eveniment:

$$P(X \in I) = \int_I f_X(x)dx$$

Interpretare geometrică: $\int_I f_X(x)\, dx =$ aria domeniului de sub graficul lui f

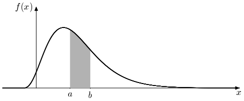


Figure: Aria domeniului hașurat reprezintă $P(a \leq X \leq b)$

Observație: $P(X = a) = P(X \in [a, a]) = \int_a^a f_X(x)dx = 0.$

Funcția de repartiție

Fie X o variabilă aleatoare.
Funcția $F_X : \mathbb{R} \to [0, 1]$, $F_X(x) = P(X \leq x)$ (probabilitatea ca variabila aleatoare X să ia valori mai mici sau egale cu x), se numește funcția de repartiție a variabilei X("cumulative distribution function"(c.d.f.))

$$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f(x)\, dx$$

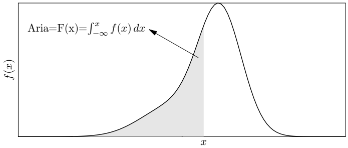


Figure: Ilustrarea semnificației geometrice a valorii funcției de repartiție într-un punct.

Proprietăți

Pentru o variabilă aleatoare continuă, funcția de repartiție este:

- continuă
- nedescrecătoare
- $\lim\limits_{x \to -\infty} F_X(x) = 0$
- $\lim\limits_{x \to \infty} F_X(x) = 1$

Observație:

$$\begin{aligned} P(X \in I) &= \int_a^b f(x)\,dx = \int\limits_{-\infty}^{b} f(x)\,dx - \int\limits_{-\infty}^{a} f(x)\,dx = \\ &= F(b) - F(a) \end{aligned}$$

În cazul în care una dintre extremitățile intervalului este $\pm\infty$ notăm $F(-\infty) = \lim\limits_{x \to -\infty} F(x) = 0, F(\infty) = \lim\limits_{x \to \infty} F(x) = 1$

Funcția de repartiție

Exemplu: Fie X variabile aleatoare, cu distribuția de probabilitate dată de densitatea f:

$$f(x) = \begin{cases} 0 & \text{dacă } x < 0 \\ \frac{1}{2}e^{-x/2} & \text{dacă } x \geq 0 \end{cases}$$

$$F(x) = \int\limits_{-\infty}^{x} f(x)\,dx = \begin{cases} \int\limits_{-\infty}^{x} 0\,dx = 0 & \text{dacă } x < 0 \\ \int\limits_{-\infty}^{0} 0\,dx + \int\limits_{0}^{x} \frac{1}{2}e^{-x/2}\,dx = 1 - e^{-x/2} & \text{dacă } x \geq 0 \end{cases}$$

Observație importantă: Fie I este un interval arbitrar din $[0, \infty)$. $P(X \in I)$ se poate calcula în două moduri:
1) $P(X \in I) = \int_I \frac{1}{2}e^{-x/2}dx$;
2) $P(X \in I) = F(b) - F(a) = e^{-a/2} - e^{-b/2}$, unde $a < b$ sunt extremitățile intervalului I.

Densitatea de probabilitate din functia de repartiție

- Dacă f_X este densitatea de probabilitate a variabilei aleatoare X, atunci funcția sa de repartiție este:

$$F_X(x) = \int_{-\infty}^{x} f_X(x)dx$$

- Dacă densitatea de probabilitate, f, a unei variabile aleatoare, X, este continuă, atunci funcția sa de repartiție, F, este derivabilă și

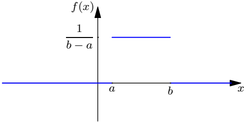
$$F'(x) = f(x), \forall x \in \mathbb{R}$$

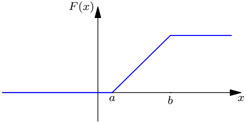
.

Distribuția uniformă

V. a. X, continuă ce are densitatea de probabilitate $f(x) = \begin{cases} \frac{1}{b-a} & \text{dacă } x \in [a,b) \\ 0 & \text{dacă } x \notin (a,b) \end{cases}$ se numește v.a. **uniform distribuită** pe intervalul $[a,b)$($X \sim \text{Unif}[a,b)$.)

Funcția de repartiție: $F(x) = \begin{cases} 0, & \text{dacă} \quad x < a \\ \frac{x-a}{b-a} & \text{dacă} \quad a \leq x < b \\ 1 & \text{dacă} \quad x \geq b \end{cases}$





https://mathlets.org/mathlets/probability-distributions/

Distribuția uniformă

Proprietate: Alegând orice subinterval $[x, x+L)$ de lungime L din $[a, b)$, avem $P(x \leq X < x+L) == \frac{L}{b-a}$, adică această probabilitate nu depinde de capetele intervalului, ci doar de lungimea lui.

Valorile lui X sunt "uniform distribuite" în subintervalele din $[a, b]$ de aceeași lungime.

Cazul a=0, b=1: de interes pentru algoritmii de generare de numere (pseudo)aleatoare folosind v.a. U, $U \sim$Unif[0,1).

Densitatea de probabilitate/funcția de repartiție este:

$$f_U(x) = \begin{cases} 1 & \text{dacă } x \in [0,1) \\ 0 & \text{în rest,} \end{cases} \quad , \quad F_U(x) = \begin{cases} 0 & \text{dacă } x < 0 \\ x & \text{dacă } x \in [0,1) \\ 1 & \text{dacă } x \geq 1 \end{cases}$$

Probabilitatea ca U, $U \sim$Unif[0,1), să ia valori într-un subinterval $[c, d) \subset [0,1)$ **este egală cu lungimea,** $d - c$, **a intervalului:**

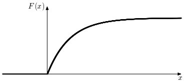
$$P(c \leq U < d) = F(d) - F(c) = d - c$$

Distribuția exponențială

V. a. X ce are densitatea de probabilitate f, definită prin:
$f(x) = \begin{cases} 0 & \text{dacă } x < 0 \\ \frac{1}{\theta} e^{-x/\theta} & \text{dacă } x \geq 0 \end{cases}$, $\theta > 0$ se numește v. a. cu **distribuție exponențială, de parametru** θ. ($X \sim \text{Exp}(\theta)$)

Funcția de repartiție a unei variabile $X \sim \text{Exp}(\theta)$ este:

$$F(x) = \begin{cases} 0 & \text{dacă } x < 0 \\ 1 - e^{-x/\theta} & \text{dacă } x \geq 0 \end{cases},$$



V. a. exponențial distribuite se folosesc ca modele pentru:

- Durata servirii unui client, de către un server dintr-un sistem coadă;
- Intervalul de timp dintre două sosiri consecutive ale clienților la coadă;
- Durata de viată a componentelor electronice;

Fie X o variabilă aleatoare continuă ce are densitatea de probabilitate f. Dacă funcția g, $g(x) = xf(x)$, este integrabilă pe $\mathbb{R}$, atunci integrala sa

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

se numește *valoarea medie* a variabilei.

Media distribuției exponențiale

Fie X o v.a. exponențial distribuită de parametru, $\theta > 0$:

$$f(x) = \begin{cases} \frac{1}{\theta} e^{-x/\theta} & \text{dacă } x > 0 \\ 0 & \text{în rest} \end{cases}$$

Avem: $M(X) = \int_{-\infty}^{\infty} x f(x)\, dx = \int_{-\infty}^{0} x \cdot 0\, dx + \int_{0}^{\infty} x \frac{1}{\theta} e^{-x/\theta}\, dx = \theta.$

Observație: Există variabile aleatoare care nu au valoare medie.
De exemplu, X-o v.a. cu distribuția de probabilitate Cauchy:
$f(x) = \frac{1}{\pi} \frac{1}{1+x^2}, \quad x \in \mathbb{R}$, integrala $\int_{-\infty}^{\infty} x f(x)\, dx$ este divergentă.

Transformări ale unei variabile aleatoare

Fie X v.a. cu densitatea de probabilitate f.
Dacă $g : \mathbb{R} \to \mathbb{R}$ este o funcție continuă sau cu un număr finit de puncte de discontinuitate de speța întâi, atunci $Y = g(X)$ este o variabilă aleatoare.
Dacă $\int_{-\infty}^{\infty} |g(x)| f(x) dx < \infty$, atunci $Y = g(X)$ are media

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

(LOTUS=law of unconscious statistician)

Dispersia unei variabile aleatoare continue X ce are media $m = M(X)$ este, ca și în cazul variabilelor discrete, numărul notat $\sigma^2(X)$ sau $D^2(X)$, și egal cu:

$$\sigma^2(X) = M((X-m)^2)$$

Notând cu g funcția definită prin $g(X) = (X-m)^2$, avem

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

Proprietăți:

- $M(aX+b) = aM(X) + b, \forall a, b \in \mathbb{R}$
- $\sigma^2(aX+b) = a^2\sigma^2(X), \forall a, b \in \mathbb{R}$
- $\sigma^2(X) = M(X^2) - (M(X))^2$

Observație: Fie X o v.a. de medie m și abatere standard σ. Variabila aleatoare asociată, $Y = \frac{X-m}{\sigma}$, se numește v.a. **standardizată**. Avem $M(Y) = 1, \sigma^2(Y) = 0$.

Concluzii

Variabile aleatoare discrete

- **Mulțime de valori:** D_x-finită sau numărabilă
- **Reprezentare** tablou de distribuție
  $X = \begin{pmatrix} x_i \\ p_i \end{pmatrix}, \quad \sum_{i=1}^{n} p_i = 1$
  sau p.m.f
  $p(x_i) = P(X = x_i) = p_i.$
- **Calcul prob**: $P(X = x_i) = p_i$
- **Funcția de repartiție:**
  $F_X(x) := P(X \leq x)$
  $F_X(x) = \sum_{i: x_i < x} p_i$

Variabile aleatoare continue

- interval din $\mathbb{R}$
- p.d.f : $f(x) \geq 0, \quad x \in \mathbb{R}$,
  $\int_{-\infty}^{\infty} f(x)\, dx = 1$
- $P(X \in I) = \int_I f(x) dx$
- $F_X(x) = \int_{-\infty}^{x} f(x) dx$
- $P(X \in I) = \int_a^b f(x) dx = F_X(b) - F_X(a),$

Concluzii

Variabile aleatoare discrete

- **Media** $M(X)= \sum_{i=1}^{n} x_i p_i$
- **Dispersie**

  $\sigma^2(X) = \sum_{i=1}^{n} (x_i - m)^2 p_i$

Variabile aleatoare continue

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

$$\sigma^2(X) = M(X^2) - (M(X))^2$$