1.8 Variabile aleatoare continue

Definiţia 1.8.1 *Numim variabilă aleatoare continuă pe spaţiul de probabilitate* $(\Omega, \mathcal{F}, P)$ *o variabilă aleatoare* $X : \Omega \to \mathbb{R}$ *pentru care funcţia de distribuţie corespunzătoare* $F_X : R \to [0,1]$, $F_X(a) = P(X < a)$ *este o funcţie continuă.*

Exemplul 1.8.2 *timpul de oprire al unui tren în staţie, temperatura într-o anumită zi, etc. sunt variabile aleatoare continue.*

Observaţia 1.8.3 *Am văzut că funcţia de distribuţie a unei variabile aleatoare discrete* $X = \begin{pmatrix} x_1 & x_2 & \cdots \\ p_1 & p_2 & \cdots \end{pmatrix}$ *este discontinuă în punctele* x_i *(are un salt egal cu* p_i *în aceste puncte), şi deci o variabilă aleatoare discretă nu este o variabilă aleatoare continuă.*

Observaţia 1.8.4 *Cerinţa ca funcţia de distribuţie* F_X *să fie o funcţie continuă este echivalentă cu* $P(X = a) = 0$ *pentru orice* $a \in \mathbb{R}$.

Într-adevăr, dacă F_X *este o funcţie continuă, atunci avem*

$$\begin{aligned} P(X = a) &= P(X \le a) - P(X < a) \\ &= \lim_{\alpha \searrow a} P(X < a) - P(X < a) \\ &= \lim_{\alpha \searrow a} F_X(\alpha) - F_X(a) \\ &= F_X(a) - F_X(a) \\ &= 0, \end{aligned}$$

deoarece F *este continuă în punctul* a*, şi reciproc.*

Observaţia 1.8.5 *Faptul că pentru o variabilă aleatoare continuă avem* $P(X = a) = 0$ *oricare ar fi* $a \in \mathbb{R}$ ***nu*** *înseamnă că* X *nu poate lua valoarea* a *(cum este cazul variabilelor aleatoare discrete). Dacă* X *nu ar putea lua valoarea* a *(oricare ar fi* $a \in \mathbb{R}$*), ar însemna că* X *nu ar putea lua nici o valoare reală!*

Definiţia 1.8.6 *Numim densitate de probabilitate a unei variabile aleatoare continue* X *o funcţie* $f = f_X : \mathbb{R} \to [0, +\infty)$ *integrabilă cu proprietatea că*

$$P(X \in B) = \int_B f(x)\, dx,$$

pentru orice mulţime boreliană $B \in \mathcal{B}$.

Observaţia 1.8.7 *Pentru* $B = (-\infty, +\infty) \in \mathcal{B}$*, obţinem*

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

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

Reciproc, se poate arăta că dacă $f : \mathbb{R} \to [0, +\infty)$ este o funcţie integrabilă ce verifică această relaţie, atunci există o variabilă aleatoare continuă având funcţia f ca densitate de probabilitate.

Observaţia 1.8.8 *Am văzut că pentru o variabilă aleatoare continuă avem $P(X = a) = 0$, şi deci*

$$P(a \le X \le b) = P(a < X \le b) = P(a \le X < b) = P(a < X < b).$$

Alegând $B = (a, b) \in \mathcal{B}$ în definiţia anterioară, rezultă că toate aceste valori egale sunt date de:

$$\left.\begin{array}{l} P(a \le X \le b) \\ P(a < X \le b) \\ P(a \le X < b) \\ P(a < X < b) \end{array}\right\} = \int_a^b f(x)\,dx,$$

pentru orice $a, b \in \mathbb{R}$ cu $a < b$.

Observaţia 1.8.9 *Pentru $B = (-\infty, a) \in \mathcal{B}$ în definiţia anterioară avem*

$$F(a) = P(X < a) = P(X \in (-\infty, a)) = \int_{-\infty}^{a} f(x)\,dx,$$

şi deci funcţia de distribuţie F dă aria de sub graficul funcţiei f (a densităţii variabilei aleatoare X).

Are loc următoarea:

Propoziţia 1.8.10 *Dacă variabila aleatoare X are o densitate de probabilitate f_X continuă, atunci ea este egală cu derivata funcţiei de distribuţie F_X a lui X:*

$$F_X'(a) = f(a), \qquad a \in \mathbb{R}.$$

Demonstraţie. Pentru $a \in \mathbb{R}$ fixat, avem

$$\begin{aligned} \frac{F(a+\Delta a) - F(a)}{\Delta a} &= \frac{1}{\Delta a}\left(\int_{-\infty}^{a+\Delta a} f(x)\,dx - \int_{-\infty}^{a} f(x)\,dx\right) \\ &= \frac{1}{\Delta a}\int_a^{a+\Delta a} f(x)\,dx \\ &= \frac{1}{\Delta a} f(a^*)\,\Delta a \\ &= f(a^*), \end{aligned}$$

conform teoremei de medie, unde a^* este un punct intermediar între a şi $a + \Delta a$.

Trecând la limită cu $\Delta a \to 0$, obţinem

$$\begin{aligned} F'(a) &= \lim_{\Delta a \to 0} \frac{F(a+\Delta a) - F(a)}{\Delta a} \\ &= \lim_{\Delta a \to 0} f(a^*) \\ &= f(a), \end{aligned}$$

deoarece f este continuă în punctul a (şi deci $a^* \to a$). ■

Observaţia 1.8.11 *Valoarea $f(a)$ a densităţii unei variabile aleatoare X în punctul a **nu** este probabilitatea ca X să ia valoarea a (am văzut că pentru o variabilă aleatoare continuă avem $P(X=a)=0$, oricare ar fi $a \in \mathbb{R}$).*

Interpretarea ce se poate da este următoarea

$$P\left(a - \frac{\varepsilon}{2} < X < a + \frac{\varepsilon}{2}\right) = \int_{a-\frac{\varepsilon}{2}}^{a+\frac{\varepsilon}{2}} f(x)\, dx \approx f(a) \cdot \varepsilon,$$

şi deci probabilitatea ca X să ia valori conţinute într-un interval de lungime $\varepsilon > 0$ mică ce conţine pe a este $f(a) \cdot \varepsilon$, adică este proporţională cu $f(a)$ ($f(a)$ arată cât este de probabil ca X să ia o valoare apropiată de a, în sensul de mai sus).

1.8.1 Caracteristici ale variabilelor aleatoare continue

Pentru o variabilă aleatoare X având densitatea f, definim:

- Media/valoarea aşteptată a lui X prin

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

- Dispersia /abaterea pătratică a lui X prin

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

  dacă integralele respective există şi sunt finite.

  Mai general, dacă $g : \mathbb{R} \to \mathbb{R}$ este o funcţie măsurabilă, definim

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

  În particular, pentru $g(x) = x^r$, respectiv pentru $g(x) = (x - M(X))^r$, obţinem

- Momentul de ordin r:

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

- Momentul centrat de ordin r:

$$\mu_r(X) = M\left((X - M(X))^r\right) = \int_{-\infty}^{+\infty} (x - M(X))^r f(x)\, dx.$$

1.8.2 Variabile aleatoare continue clasice

1. Variabila aleatoare uniformă

Numim variabilă aleatoare uniform distribuită pe intervalul $[a, b]$ o variabilă aleatoare având densitatea

$$f(x) = \begin{cases} \frac{1}{b-a}, & a \le x \le b \\ 0, & \text{în rest} \end{cases}$$

Observaţia 1.8.12 *Să observăm că funcţia $f(x)$ este într-adevăr o densitate de probabilitate, doarece este integrabilă, ne-negativă ($f(x) \ge 0$, oricare ar fi $x \in \mathbb{R}$) şi avem*

$$\int_{-\infty}^{+\infty} f(x)\, dx = \int_a^b \frac{1}{b-a} dx = \left.\frac{x}{b-a}\right|_a^b = \frac{b}{b-a} - \frac{a}{b-a} = 1.$$

Densitatea $f(x)$ determină în mod unic funcţia de distribuţie corespunzătoare $F(x)$:

$$F(x) = \begin{cases} 0, & x \le a \\ \frac{x-a}{b-a}, & a < x \le b \\ 1, & b < x \end{cases}$$

Pentru o variabilă aleatoare uniformă pe intervalul $[a, b]$, avem

$$P(X \le a) = \int_{-\infty}^{a} f(x)\, dx = \int_{-\infty}^{0} 0 dx = 0,$$

şi

$$P(X \ge a) = \int_b^{+\infty} f(x)\, dx = \int_b^{+\infty} 0 dx = 0,$$

iar pentru $a \le \alpha < \beta \le b$ avem

$$P(\alpha \le X \le \beta) = \int_\alpha^\beta f(x)\, dx = \int_\alpha^\beta \frac{1}{b-a} dx = \frac{\beta - \alpha}{b-a},$$

ceea ce arată că X ia valori (numai) în intervalul $[a, b]$, probabilitatea ca X să ia valori într-un subinterval al lui $[a, b]$ fiind proporţională cu lungimea acestui subinterval.

Media şi dispersia unei variabile aleatoare uniforme pe intervalul $[a, b]$ sunt

$$M(X) = \int_{-\infty}^{+\infty} x f(x)\, dx = \int_a^b x \cdot \frac{1}{b-a} dx = \frac{1}{b-a}\left(\frac{b^2}{2} - \frac{a^2}{2}\right) = \frac{a+b}{2},$$

respectiv

$$\sigma^2(X) = M\left((X - M(X))^2\right) = \int_a^b \left(x - \frac{a+b}{2}\right)^2 \cdot \frac{1}{b-a} dx = \frac{(b-a)^2}{12}.$$

Observaţia 1.8.13 *Dacă X este o variabilă aleatoare uniformă pe intervalul $[a, b]$, atunci $Y = c_1 X + c_2$ este o variabilă aleatoare uniformă pe intervalul $[ac_1 + c_2, bc_1 + c_2]$.*

În particular, dacă X este o variabilă aleatoare uniformă pe intervalul $[a, b]$, atunci $Y = \frac{X-a}{b-a}$ este o variabilă aleatoare uniformă pe intervalul $[0, 1]$.

2. Variabila aleatoare normală

Definiţia 1.8.14 *Numim variabilă aleatoare normală cu parametrii μ şi σ^2 ($\sigma > 0$) o variabilă aleatoare X având densitatea*

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}},$$

şi notăm $X \in \mathcal{N}\left(\mu, \sigma^2\right)$.

Observaţia 1.8.15 *Funcţia f este într-adevăr o densitatea de probabilitate, deoarece este integrabilă, ne-negativă ($f(x) \geq 0$ pentru orice $x \in \mathbb{R}$) şi verifică relaţia $\int_{-\infty}^{+\infty} f(x)\,dx = 1$ (vezi seminar).*

Graficul funcţiei $f(x)$ are forma de clopot, ce are un maxim pentru $x = \mu$ şi puncte de inflexiune pentru $x = \mu \pm \sigma$, graficul fiind simetric faţă de dreapta $x = \mu$.

Media şi dispersia unei variabile aleatoare normale cu parametrii μ şi σ^2 sunt:

– media: $M(X) = \int_{-\infty}^{+\infty} x f(x)\,dx = \mu$

– dispersia: $\sigma^2(X) = M\left((X - M(X))^2\right) = \int_{-\infty}^{+\infty} (x-\mu)^2 f(x)\,dx = \sigma^2$,

şi deci media şi dispersia sunt chiar parametrii μ şi σ^2 ai variabilei aleatoare normale $X \in \mathcal{N}\left(\mu, \sigma^2\right)$.

Are loc următoarea:

Propoziţia 1.8.16 *Dacă $X \in \mathcal{N}\left(\mu, \sigma^2\right)$ este o variabilă aleatoare normală cu parametrii μ şi σ^2, atunci $Y = \alpha X + \beta \in \mathcal{N}\left(\alpha\mu + \beta, \alpha^2\sigma^2\right)$.*

În particular, dacă $X \in \mathcal{N}\left(\mu, \sigma^2\right)$, atunci $Y = \frac{X-\mu}{\sigma} \in \mathcal{N}(0, 1)$ este o variabilă aleatoare cu medie 0 şi dispersie 1 (numită variabilă aleatoare normală standard*).*

Demonstraţie. Să calculăm funcţia de distribuţie F_Y a variabilei aleatoare Y:

$$\begin{aligned} F_Y(a) &= P(Y < a) = P(\alpha X + \beta < a) \\ &= P\left(X < \frac{a-\beta}{\alpha}\right) \\ &= F_X\left(\frac{a-\beta}{\alpha}\right), \end{aligned}$$

adică funcţia de distribuţie a variabilei aleatoare X calculată în punctul $\frac{a-\beta}{\alpha}$ (am folosit aici faptul că $\alpha > 0$).

Derivând în raport cu a, şi folosind faptul că derivata funcţiei de distribuţie este densitatea de probabilitate (adică $F'_Y = f_Y$ şi $F'_X = f_X$), obţinem:

$$\begin{aligned} f_Y(a) &= F'_Y(a) = \left(F_X\left(\frac{a-\beta}{\alpha}\right)\right)' = \frac{1}{\alpha}F'_X\left(\frac{a-\beta}{\alpha}\right) \\ &= \frac{1}{\alpha}f_X\left(\frac{a-\beta}{\alpha}\right) \\ &= \frac{1}{\alpha}\cdot\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{\left(\frac{a-\beta}{\alpha}-\mu\right)^2}{2\sigma^2}} \\ &= \frac{1}{\sqrt{2\pi}\,(\alpha\sigma)}e^{-\frac{(a-(\alpha\mu+\beta))^2}{2(\alpha\sigma)^2}}, \end{aligned}$$

adică $Y \in \mathcal{N}(\alpha\mu+\beta, \alpha\sigma)$. ■

Observaţia 1.8.17 *Pentru o variabilă aleatoare normală standard $X \in \mathcal{N}(0,1)$, valorile funcţiei de distribuţie*

$$F_X(a) = P(X < a) = \int_{-\infty}^{a}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx \overset{not}{=} \Phi(a)$$

se găsesc tabelate (sau pot fi calculate folosind diverse programe/aplicaţii pe calculator).

Dacă $Y \in \mathcal{N}(\mu,\sigma^2)$, din propoziţia anterioară rezultă că

$$F_Y(a) = P(Y < a) \overset{Y=\sigma X+\mu}{=} P\left(X < \frac{a-\mu}{\sigma}\right) = \Phi\left(\frac{a-\mu}{\sigma}\right),$$

şi deci aceste valori se pot calcula folosind funcţia Φ definită anterior.

1.8.3 Exerciţii

1. O variabilă aleatoare are densitatea

$$f(x) = \begin{cases} c\left(4x - x^2\right), & 0 < x < 2 \\ 0, & \text{în rest} \end{cases}$$

   (a) Să se determine valoarea constantei c

   (b) Să se determine probabilitatea $P(X > 1)$

   (c) Care este intervalul valorilor posibile ale lui X?

2. Numărul de ore de funcţionare al unei componente electronice este o variabilă aleatoare cu densitatea

$$f(x) = \begin{cases} ce^{+\frac{x}{100}}, & x \geq 0 \\ 0, & \text{în rest} \end{cases}$$

(a) Să se determine valoarea constantei c

(b) Care este probabilitatea ca piesa să funcţioneze între 50 şi 100 de ore?

(c) Care este probabilitatea ca piesa să funcţioneze mai puţin de 100 de ore?

3. Durata de viaţă (în ore) a unei componente electronice este o variabilă aleatoare cu densitatea

$$f(x) = \begin{cases} 0, & x \le 100 \\ \frac{c}{x^2}, & x > 100 \end{cases}$$

   (a) Să se determine valoarea constantei c

   (b) Care este probabilitatea ca din 5 astfel de piese, exact 2 să trebuiască a fi înlocuite în primele 5 ore de funcţionare? (presupunem că piesele se defectează independent unele de altele)

4. Dacă X este o variabilă aleatoare uniformă pe intervalul $(0, 10)$, să se calculeze următoarele probabilităţi:

   (a) $P(X < 3)$

   (b) $P(X > 6)$

   (c) $P(3 < X < 8)$

5. Autobuzele sosesc în staţie începând cu ora 7 la intervale de 15 minute. Un călător soseşte în staţie la o oră ce este uniform distribuită între ora $7:00$ şi ora $7:30$. Să se determine:

   (a) Probabilitatea de a aştepta cel puţin 5 minute sosirea primului autobuz

   (b) Probabilitatea de a aştepta cel puţin 10 minute sosirea primului autobuz.

6. Timpul necesar completării unui formular de către un angajat este de o variabilă aleatoare X uniform distribuită între 1.5 şi 2.2 minute.

   (a) Care este probabilitatea ca angajatul să completeze formularul în mai puţin de 2 minute?

   (b) Să se determine funcţia de distribuţie corespunzătoare F_X şi să se reprezinte grafic.

7. Timpul necesar asamblării unui produs are o densitate

$$f(x) = \begin{cases} 0.1, & 30 \le x \le 40 \\ 0, & \text{în rest} \end{cases}$$

(a) Să se determine proporţia pieselor ce necesită cel puţin 35 minute pentru asamblare

(b) Ce durată de timp de asamblare este depăşită pentru 90% din piese?

8. Dacă X este o variabilă aleatoare normală $\mathcal{N}(3, 9)$ cu medie $\mu = 3$ şi dispersie $\sigma^2 = 9$, să se determine următoarele probabilităţi:

   (a) $P(2 < X < 5)$

   (b) $P(X > 0)$

   (c) $P(|X - 3| > 6)$

   Observaţie: pentru calculul acestor probabilităţi este nevoie de un tabel de valori a funcţiei de distribuţie normală, sau a unui computer.

9. Notele obţinute în urma unei examinări se consideră a avea o distribuţie normală $\mathcal{N}(\mu, \sigma^2)$. Există regula de a da evaluări prin litere, A, B, C, D şi F după următoarea: A dacă nota este mai mare decât $\mu + \sigma$, B dacă nota este în intervalul $(\mu, \mu + \sigma)$, C dacă nota este în intervalul $(\mu - \sigma, \mu)$, D dacă nota este în intervalul $(\mu - 2\sigma, \mu - \sigma)$, respectiv F dacă nota este mai mică decât $\mu - 2\sigma$.

   Să se determine procentele de note A, B, C, D şi F acordate conform acestei reguli.

   Observaţie: pentru acest calcul este nevoie de un tabel de valori a funcţiei de distribuţie normală, sau a unui computer. Răspuns: A $\to$ 15.87%, B $\to$ 34.13%, C $\to$ 34.13%, D $\to$ 13.59%, respectiv F $\to$ 2.28%.

10. Se aruncă o monedă de $n = 40$ de ori, şi se notează cu X numărul de feţe stemă obţinute.

    (a) Să se determine probabilitatea $P(X = 20)$

    (b) Să se aproximeze această probabilitate folosind aproximaţia prin variabilă aleatoare normală.

11. Presupunem că durata unei convorbiri telefonice este o variabilă aleatoare cu parametrul $\lambda = \frac{1}{10}$. Dacă o persoană soseşte la o cabină telefonică exact când o altă persoană a început convorbirea, să se determine următoarele probabilităţi:

    (a) $P(X > 10)$

    (b) $P(10 < X < 20)$

    (c) Să se determine t astfel încât $P(X > t) = \frac{1}{2}$

    Răspuns: a) 0.368 b) 0.233 c) $t = 10 \ln 2 \approx 0.693$