Tema 4

Soluții

Exercițiul 1

a) Pentru ca $f(x)$ să fie densitate de probabilitate trebuie să verifice proprietățile $f(x) \geq 0$ și $\int f(x)\,dx = 1$. Avem

$$\int_{-\infty}^{\infty} f(x)\,dx = c\int_0^7 \ln\left(\frac{7}{x}\right)\,dx = 7c\int_0^{\infty} te^{-t}\,dt = 7c$$

deci $c = \frac{1}{7}$. Pentru funcția de repartiție avem

$$F(x) = \int_{-\infty}^{x} f(t)\,dt = \int_0^x \frac{1}{7}\ln\left(\frac{7}{t}\right)\,dt = \int_{\ln(7/x)}^{\infty} ue^{-u}\,du = \frac{x}{7}\left[1+\ln\left(\frac{7}{x}\right)\right], \quad x \in (0,7)$$

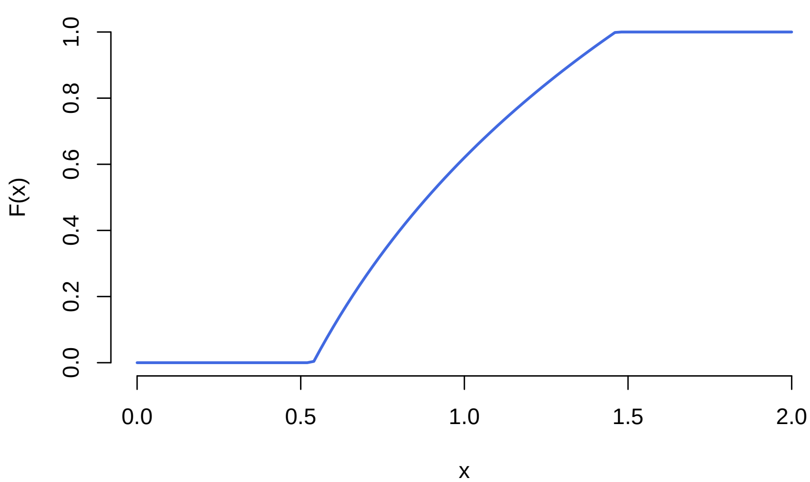
iar $\mathbb{P}(X > 3) = 1 - F(3) = \frac{4}{7} - \frac{3}{7}\ln\left(\frac{7}{3}\right) = 0.087$.

b) Din condiția $\int f(x)\,dx = 1$ decucem

$$\int_{1-c}^{1+c} \frac{1}{x}\,dx = \ln\frac{1+c}{1-c} = 1$$

ceea ce implică $c = \frac{e-1}{e+1}$. Pentru funcția de repartiție avem

$$F(x) = \int_{-\infty}^{x} f(t)\,dt = \int_{1-c}^{x} \frac{1}{t}\,dt = \ln\frac{x}{1-c}, \quad x \in (1-c, 1+c).$$



Exercițiul 2

a) Fie X nivelul de zgomot produs de o maşină de spălat luată la intamplare şi $\bar{X}_{10}$ media unui eşantion de talie 10. Presupunem că aproximarea gaussiană are loc pentru $n = 10$. Avem

$$\bar{X}_{10} \sim \mathcal{N}\left(44, \frac{5^2}{10}\right),$$

de unde $\mathbb{P}(\bar{X}_{10} > 48) \simeq \mathbb{P}\left(Z > \frac{48-44}{5/\sqrt{10}}\right) = \mathbb{P}(Z > 2.53) = 1 - 0.9943 = 0.0057$, unde $Z \sim \mathcal{N}(0,1)$. Observăm că această proabilitate este foarte mică.

b) Fie X greutatea unei persoane luate la intamplare şi $\bar{X}_{100}$ greutatea media a unui eşantion de 100 de persoane. Aplicand approximarea gaussiană (Teorema Limită Centrală) avem

$$\bar{X}_{100} \simeq \mathcal{N}\left(66.3, \frac{15.6^2}{100}\right),$$

de unde $\mathbb{P}(\bar{X}_{100} > \frac{7000}{100}) \simeq \mathbb{P}\left(Z > \frac{70-66.3}{15.6/\sqrt{100}}\right) = \mathbb{P}(Z > 2.37) = 0.0089$, unde $Z \sim \mathcal{N}(0,1)$.

Exercițiul 3

Fie X_i rezultatul obținut în urma celei de-a i-a aruncare cu banul, $X_i = 1$ dacă banul a picat cap și $X_i = 0$ dacă a picat pajură. Dacă notăm cu $\bar{X}_n = \frac{X_1+\cdots+X_n}{n}$ frecvența de apariție a capului în primele n aruncări atunci conform *Teoremei Limită Centrale* avem

$$\mathbb{P}\left(\sqrt{n}\frac{\bar{X}_n - \mu}{\sigma} \leq x\right) \approx \Phi(x)$$

unde $\mu = \mathbb{P}(X_1 = 1) = p$ și $\sigma^2 = Var(X_1) = p(1-p)$. Prin urmare

$$\mathbb{P}\left(\sqrt{n}\frac{|\bar{X}_n - p|}{\sqrt{p(1-p)}} \leq x\right) = \mathbb{P}\left(\sqrt{n}\frac{\bar{X}_n - p}{\sqrt{p(1-p)}} \leq x\right) + \mathbb{P}\left(\sqrt{n}\frac{\bar{X}_n - p}{\sqrt{p(1-p)}} \leq -x\right) \approx 2\Phi(x) - 1$$

deci

$$\mathbb{P}\left(|\bar{X}_n - p| \leq x\right) = \mathbb{P}\left(\sqrt{n}\frac{|\bar{X}_n - p|}{\sqrt{p(1-p)}} \leq x\sqrt{\frac{n}{p(1-p)}}\right) \approx 2\Phi\left(x\sqrt{\frac{n}{p(1-p)}}\right) - 1.$$

În cazul nostru $p = \frac{1}{2}$, $x = 0.01$ iar $\mathbb{P}\left(|\bar{X}_n - p| \leq x\right) \leq 0.6$ astfel n se determină rezolvând ecuația

$$2\Phi\left(0.01\sqrt{\frac{n}{0.5(1-0.5)}}\right) - 1 = 0.6$$

care implică $0.01\sqrt{\frac{n}{0.5(1-0.5)}} = 0.84$ adică $n = 1764$.

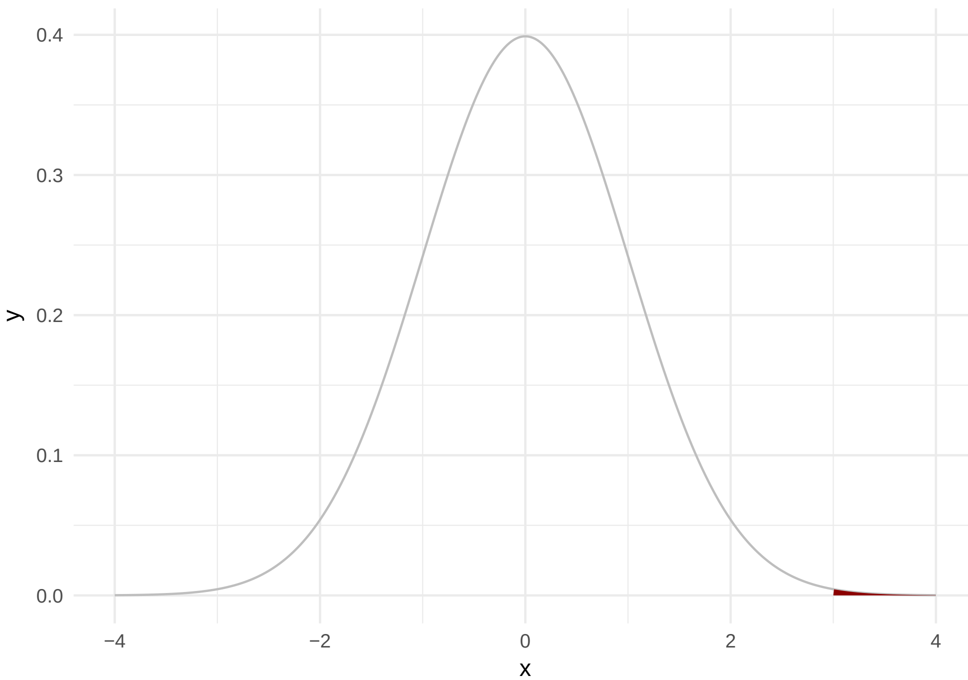
Exercițiul 4

a) Se observă cu uşurinţă că

$$
\begin{aligned}
H_n(x) &= \mathbb{P}(Y_n \leq x) = \mathbb{P}(X_1 \leq x, \cdots, X_n \leq x) \overset{indep.}{=} F(x)^n \\
h_n(x) &= \frac{d}{dx} H_n(x) = nf(x)F(x)^{n-1} \\
H_1(x) &= \mathbb{P}(Y_1 \leq x) = 1 - \mathbb{P}(Y_1 > x) = 1 - \mathbb{P}(X_1 > x, \cdots, X_n > x) \overset{indep.}{=} 1 - (1 - F(x))^n \\
h_1(x) &= \frac{d}{dx} H_1(x) = nf(x)\,(1 - F(x))^{n-1}
\end{aligned}
$$

b) Fie $X \sim \mathcal{N}(\mu, \sigma^2)$. Problema cere să găsim probabilitatea $\mathbb{P}(X > \mu + 3\sigma)$. Avem (vezi porţiunea roşie din figură)

$$\mathbb{P}(X > \mu + 3\sigma) = \mathbb{P}\left(\frac{X - \mu}{\sigma} > 3\right) = 1 - \mathbb{P}\left(\frac{X - \mu}{\sigma} \leq 3\right) = 0.00135$$



c) Fie $X_1, X_2, \ldots, X_n$ un e santion de talie $n = 100$ dintr-o populaţie normală $\mathcal{N}(\mu, \sigma^2)$ şi fie $Z_i = \mathbf{1}_{\{X_i > \mu + 3\sigma\}}$ variabilele Bernoulli care iau valoarea 1 atunci cand $X_i > \mu + 3\sigma$ şi 0 in rest. Problema revine la a determina probabilitatea

$$\mathbb{P}(Z_1 + \cdots Z_n = 1) \overset{i.i.d.}{=} \binom{n}{1} \mathbb{P}(Z_1 = 1)\mathbb{P}(Z_1 = 0)^{n-1} = n\mathbb{P}(X_1 > \mu + 3\sigma)\mathbb{P}(X_1 < \mu + 3\sigma)^{n-1} \simeq 0.11809$$

d) Fie $X_1, X_2, \ldots, X_n$ un e santion de talie $n = 100$ dintr-o populaţie normală $\mathcal{N}(0, 1)$. Problema ne cere să găsim valoarea lui x pentru care probabilitatea $\mathbb{P}(X_1 < x, X_2 < x, \cdots, X_n < x) = 0.99$. Prin urmare vrem să găsim pe x aşa incat $H_n(x) = 0.99$. Din punctul a) avem $H_n(x) = F(x)^n$ deci $x = F^{-1}(\sqrt[n]{0.99}) = 3.7177$.

e) Fie $X_1, X_2, \dots, X_n$ un e santion de talie $n = 50$ dintr-o populaţie normală $\mathcal{N}(10,1)$ ($n = 50$ reprezintă numărul de laboratoare iar X_i este concentraţia de crom din laboratorul i). Din datele problemei avem că laboratorul 1 a inregistrat cea mai mică valoare (6 mg/l) iar laboratorul 2 a inregistrat cea mai mare valoare (13 mg/l). Problema ne cere să evaluăm probabilitatea

$$\mathbb{P}(Y_1 \leq 6, Y_n \geq 13) = 1 - \mathbb{P}(\{Y_1 > 6\} \cup \{Y_n < 13\}) = 1 - \mathbb{P}(Y_1 > 6) - \mathbb{P}(Y_n < 13) + \mathbb{P}(Y_1 > 6, Y_n < 13).$$

Avem că $\mathbb{P}(Y_1 > 6) = \mathbb{P}(X_1 > 6, \cdots, X_n > 6) = (1 - F(6))^n$ iar $F(6) = \mathbb{P}(X_1 \leq 6) = \mathbb{P}\left(\frac{X_1 - 10}{1} \leq -4\right) \simeq 0.00003$ deci $\mathbb{P}(Y_1 > 6) \simeq 0.99871$.

De asemenea $\mathbb{P}(Y_n < 13) = F(13)^n$ iar cum $F(13) = \mathbb{P}(X_1 \leq 13) = \mathbb{P}\left(\frac{X_1 - 10}{1} \leq 3\right) \simeq 0.9986$ rezultă că $\mathbb{P}(Y_n < 13) \simeq 0.9346$.

In mod similar, $\mathbb{P}(Y_1 > 6, Y_n < 13) = \mathbb{P}(6 < X_1 < 13, \cdots, 6 < X_n < 13) = \mathbb{P}(6 < X_1 < 13)^n$ şi cum $\mathbb{P}(6 < X_1 < 13) = \mathbb{P}(X_1 < 13) - \mathbb{P}(X_1 \leq 6) \simeq 0.9986$ obţinem că $\mathbb{P}(Y_1 > 6, Y_n < 13) \simeq 0.9332$.

In concluzie avem că $\mathbb{P}(Y_1 \leq 6, Y_n \geq 13) \simeq 0.0001$.

Exercițiul 5

a) Pentru ca $f_{(X,Y)}$ să fie densitate trebuie ca $f_{(X,Y)}(x,y) \geq 0$ de unde $k \geq 0$ și

$$\iint_{\mathbb{R}^2} f_{(X,Y)}(x,y)\,dxdy = 1,$$

altfel spus $\int_0^1 \int_0^2 k(x+y+1)\,dxdy = 1$ de unde $k = \frac{1}{5}$.

b) Pentru a găsi densitățile marginale avem pentru X

$$f_X(x) = \int f_{(X,Y)}(x,y)\,dy = \int_0^2 \frac{x+y+1}{5} \mathbb{1}_{[0,1]}(x)\,dy = \frac{2x+4}{5} \mathbb{1}_{[0,1]}(x)$$

și pentru Y

$$f_Y(y) = \int f_{(X,Y)}(x,y)\,dx = \int_0^1 \frac{x+y+1}{5} \mathbb{1}_{[0,2]}(y)\,dx = \frac{2y+3}{10} \mathbb{1}_{[0,2]}(y).$$

c) Observăm că X și Y nu sunt independente deoarece $f_{(X,Y)}(x,y) \neq f_X(x) f_Y(y)$.

d) Funcția de repartiție a vectorului (X,Y) este dată de

$$F(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{(X,Y)}(u,v)\,dudv$$

prin urmare dacă $x \in (0,1]$ și $y \in (0,2]$ avem

$$F(x,y) = \int_0^x \int_0^y \frac{u+v+1}{5}\,dudv = \frac{xy}{10}(x+y+2),$$

dacă $x \in (1,\infty)$ și $y \in (0,2]$ atunci

$$F(x,y) = \int_0^1 \int_0^y \frac{u+v+1}{5}\,dudv = \frac{y}{10}(y+3),$$

dacă $x \in (0,1]$ și $y \in (2,\infty)$ atunci

$$F(x,y) = \int_0^x \int_0^2 \frac{u+v+1}{5}\, dudv = \frac{x}{5}(x+4),$$

iar dacă $x \in (1,\infty)$ și $y \in (2,\infty)$ atunci

$$F(x,y) = \int_0^1 \int_0^2 \frac{u+v+1}{5}\, dudv = 1.$$

Pentru a găsi funcțiile de repartiție marginale putem sau să folosim densitățile marginale sau să folosim relațiile

$$F_X(x) = \lim_{y\to\infty} F(x,y), \quad F_Y(y) = \lim_{x\to\infty} F(x,y).$$

Obținem

$$F_X(x) = \begin{cases} 0, & x \leq 0 \\ \frac{x}{5}(x+4), & x \in (0,1] \\ 1, & x > 1 \end{cases}$$

și

$$F_Y(y) = \begin{cases} 0, & y \leq 0 \\ \frac{y}{10}(y+3), & y \in (0,2] \\ 1, & y > 2 \end{cases}$$

e) Folosind definiția densității marginale avem

$$f_{X|Y}(x|y) = \frac{f_{(X,Y)}(x,y)}{f_Y(y)} = \frac{2(x+y+1)}{2y+3} \mathbb{1}_{[0,1]}(x)\mathbb{1}_{[0,2]}(y)$$

și

$$f_{Y|X}(y|x) = \frac{f_{(X,Y)}(x,y)}{f_X(x)} = \frac{x+y+1}{2x+4} \mathbb{1}_{[0,1]}(x)\mathbb{1}_{[0,2]}(y).$$

Exercițiul 6

Fie A_j evenimentul ca cel puțin o persoană din cele 110 să fie născută în ziua j, $1 \leq j \leq 365$ și fie $X_j = \mathbf{1}_{A_j}$. Variabila de interes X, numărul de zile de naștere distincte din grup, este

$$X = \sum_{j=1}^{365} X_j,$$

cu X_j repartizate Bernoulli de parametru $p = \mathbb{P}(A_j)$. Probabilitatea p se calculează din

$$\begin{aligned} p &= \mathbb{P}(A_j) = 1 - \mathbb{P}(A_j^c) \\ &= 1 - \mathbb{P}(\text{nicio persoană din grup nu este născută în ziua j}) = 1 - \left(\frac{364}{365}\right)^{110}. \end{aligned}$$

Putem observa că variabilele aleatoare X_j nu sunt independente (doar identic repartizate) dar folosind proprietatea de liniaritate a mediei avem

$$\mathbb{E}[X] = \sum_{j=1}^{365} \mathbb{E}[X_j] = 365p \approx 95.083.$$

Pentru calculul varianței avem

$$Var(X) = \sum_{j=1}^{365} Var[X_j] + \sum_{i\neq j} Cov[X_i, X_j]$$

care din simetrie devine

$$Var(X) = 365Var(X_1) + 2\binom{365}{2}Cov(X_1, X_2).$$

Cum $Var(X_1) = p(1-p)$, ne rămâne să calculăm covarianța $Cov(X_1, X_2)$. Aceasta din urmă verifică

$$Cov(X_1, X_2) = \mathbb{E}[X_1X_2] - \mathbb{E}[X_1]\mathbb{E}[X_2] = \mathbb{E}[X_1X_2] - p^2$$

iar $\mathbb{E}[X_1X_2] = \mathbb{P}(X_1X_2 = 1) = \mathbb{P}(A_1 \cap A_2)$. Prin complementare avem

$$\mathbb{P}(A_1^c \cup A_2^c) = \mathbb{P}(A_1^c) + \mathbb{P}(A_2^c) - \mathbb{P}(A_1^c \cap A_2^c) = 2\left(\frac{364}{365}\right)^{110} - \left(\frac{363}{365}\right)^{110}$$

de unde găsim că

$$Var(X) = 365p(1-p) + 365 \times 364 \times \left[1 - 2\left(\frac{364}{365}\right)^{110} + \left(\frac{363}{365}\right)^{110} - p^2\right] \approx 10.019.$$

Exercițiul 7

Fie X și Y cele două măsurători, iar conform ipotezei acestea sunt variabile aleatoare independente și $\mathcal{N}(0,1)$ repartizate. Fie de asemenea $M = \max(X,Y)$ și $L = \min(X,Y)$, cea mai mare și respectiv cea mai mică dintre valori. Cum pentru $x, y \in \mathbb{R}$ avem

$$\max(x,y) + \min(x,y) = x + y \quad \text{și} \quad \max(x,y) - \min(x,y) = |x - y|$$

deducem că

$$\begin{aligned} &\mathbb{E}[M] + \mathbb{E}[L] = \mathbb{E}[M + L] = \mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y] = 0 \\ &\mathbb{E}[M] - \mathbb{E}[L] = \mathbb{E}[M - L] = \mathbb{E}[|X - Y|]. \end{aligned}$$

Pentru a calcula $\mathbb{E}[|X-Y|]$ să observăm că $X - Y \sim \mathcal{N}(0,2)$ (deoarece X și Y sunt independente) prin urmare notând $X - Y = Z\sqrt{2}$, cu $Z \sim \mathcal{N}(0,1)$, obținem

$$\begin{aligned}\mathbb{E}[|X-Y|] = \sqrt{2}\mathbb{E}[|Z|] = \sqrt{2}\int_{-\infty}^{\infty}|z|\frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}\,dz \\ = 2\sqrt{2}\int_{0}^{\infty}z\frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}\,dz = 2\sqrt{2}\frac{1}{\sqrt{2\pi}} = \frac{2}{\sqrt{\pi}}.\end{aligned}$$

Astfel deducem că $\mathbb{E}[M] = \frac{1}{\sqrt{\pi}}$ și $\mathbb{E}[L] = -\frac{1}{\sqrt{\pi}}$ iar

$$Cov(M,L) = \mathbb{E}[ML] - \mathbb{E}[M]\mathbb{E}[L] = \mathbb{E}[XY] + \frac{1}{\pi} = \frac{1}{\pi}$$

deoarece $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y] = 0$. Cum corelația dintre M și L este definită prin

$$\rho(M,L) = \frac{Cov(M,L)}{\sqrt{Var(M)}\sqrt{Var(L)}},$$

trebuie să mai calculăm $Var(M)$ și $Var(L)$. Cum $M+L = X+Y$, luând varianța avem

$$Var(M+L) = Var(X+Y) = Var(X) + Var(Y) = 2$$

de unde

$$Var(M) + Var(L) + 2Cov(M,L) = 2$$

deci $Var(M) + Var(L) = 2\left(1-\frac{1}{\pi}\right)$. Cum $\max(X,Y) = -\min(-X,-Y)$ și cum, din simetria repartiției normale standard, $(-X,-Y)$ este repartizat la fel ca (X,Y) deducem că

$$Var(M) = Var(\max(X,Y)) = Var(-\min(-X,-Y)) = Var(\min(-X,-Y)) = Var(\min(X,Y)) = Var(L)$$

prin urmare $Var(M) = Var(L) = 1 - \frac{1}{\pi}$ de unde

$$\rho(M,L) = \frac{Cov(M,L)}{\sqrt{Var(M)}\sqrt{Var(L)}} = \frac{\frac{1}{\pi}}{1-\frac{1}{\pi}} = \frac{1}{\pi-1}.$$