Tema 2

Soluții

Exercițiul 1

Fie $X_1, X_2, \dots, X_n$ variabile aleatoare i.i.d. repartizate $\mathcal{U}([0,1])$.

a) Determinați funcția de repartiție și densitatea variabilelor m_n și M_n, unde $m_n = \min(X_1, X_2, \dots, X_n)$ iar $M_n = \max(X_1, X_2, \dots, X_n)$.

b) Fie $Z_n = n(1 - M_n)$. Arătați că $Z_n \xrightarrow{d} Z$, unde Z este o variabilă aleatoare a cărei funcție de repartiție este $F_Z(z) = 1 - e^{-z}$.

a) Pentru $M_n = \max(X_1, X_2, \dots, X_n)$ observăm că pentru $x \in (0,1)$

$$F_{M_n}(x) = \mathbb{P}\left(M_n \leq x\right) = \mathbb{P}(X_1 \leq x,\, X_2 \leq x, \dots,\, X_n \leq x) \overset{indep.}{=} \prod_{i=1}^{n} \mathbb{P}(X_i \leq x) = F(x)^n = x^n.$$

Dacă $x < 0$ atunci $F_{M_n}(x) = 0$ iar dacă $x \geq 1$ avem $F_{M_n}(x) = 1$. In mod similar pentru $m_n = \min(X_1, X_2, \dots, X_n)$ şi $x \in (0,1)$ rezultă că

$$\begin{aligned} F_{m_n}(x) &= \mathbb{P}\left(m_n \leq x\right) = 1 - \mathbb{P}(m_n > x) = 1 - \mathbb{P}(X_1 > x,\, X_2 > x, \dots,\, X_n > x) \\ &\overset{indep.}{=} 1 - \prod_{i=1}^{n} \mathbb{P}(X_i > x) = 1 - [1 - F(x)]^n = 1 - (1-x)^n. \end{aligned}$$

Pentru a calcula densitatea v.a. m_n şi M_n este suficient să derivăm expresiile de mai sus şi obţinem $f_{m_n}(x) = n(1-x)^{n-1}$ şi $f_{M_n}(x) = nx^{n-1}$ pentru $x \in [0,1]$ şi 0 in rest.

b) Fie $Z_n = n(1 - M_n)$. Pentru calculul funcţiei de repartiţie avem

$$F_{Z_n}(z) = \mathbb{P}(Z_n \leq z) = \mathbb{P}\left(M_n \geq 1 - \frac{z}{n}\right) = 1 - \left(1 - \frac{z}{n}\right)^n,\ z > 0.$$

Cum $\left(1 - \frac{z}{n}\right)^n \to e^{-z}$ pentru $n \to \infty$ rezultă că

$$\lim_{n\to\infty} F_{Z_n}(z) = 1 - e^{-z}$$

această limită reprezentand funcţia de repartiţie a unei v.a. repartizată exponenţial de parametru 1.

Exercițiul 2

Fie U_1, U_2 două variabile aleatoare independente repartizate uniform $\mathcal{U}([0,1])$.

a) Arătați că variabilele

$$X_1 = \cos(2\pi U_1)\sqrt{-2\log(U_2)}, \quad X_2 = \sin(2\pi U_1)\sqrt{-2\log(U_2)}$$

sunt variabile aleatoare independente repartizate normal $\mathcal{N}(0,1)$.

b) Deduceți că reprezentarea în coordonate polare (R, Θ) a lui (X_1, X_2) verifică

$$R^2 \sim \mathcal{E}\left(\frac{1}{2}\right) \quad \text{și} \quad \Theta \sim \mathcal{U}([0, 2\pi])$$

a) Considerăm schimbarea de variabilă

$$\begin{aligned} g &: (0,1)^2 \to \mathbb{R}^2 \setminus \{(x,0) \,|\, x \geq 0\} \\ &(u_1, u_2) \mapsto (\sqrt{-2\log(u_1)}\cos(2\pi u_2, \sqrt{-2\log(u_1)}\sin(2\pi u_2)) \end{aligned}$$

Observăm că g este un difeomorfism de clasă $\mathcal{C}^1$ între mulțimile deschise $(0,1)^2$ și $\mathbb{R}^2 \setminus \{(x,0) \,|\, x \geq 0\}$ cu

$$\begin{aligned} g^{-1} &: \mathbb{R}^2 \setminus \{(x,0) \,|\, x \geq 0\} \to (0,1)^2 \\ &(x,y) \mapsto \left(\exp\left(-\frac{x^2+y^2}{2}\right), \frac{1}{2} + \frac{1}{\pi}\arctan\frac{y}{x + \sqrt{x^2+y^2}}\right) \end{aligned}$$

Plecând de la densitatea cuplului (U_1, U_2) putem determina densitatea vectorului (X_1, X_2) în urma aplicării teoremei de schimbare de variabilă:

$$f_{(X_1,X_2)}(x_1,x_2) = f_{(U_1,U_2)}(g^{-1}(x_1,x_2))\left|\det J_{g^{-1}}(x_1,x_2)\right|.$$

Avem

$$\det J_{g^{-1}}(x_1,x_2) = \begin{vmatrix} -x_1\exp\left(-\frac{x_1^2+x_2^2}{2}\right) & -x_2\exp\left(-\frac{x_1^2+x_2^2}{2}\right) \\ -\frac{x_2}{2\pi(x_1^2+x_2^2)} & \frac{x_1}{2\pi(x_1^2+x_2^2)} \end{vmatrix} = -\frac{1}{2\pi}\exp\left(-\frac{x_1^2+x_2^2}{2}\right)$$

și în plus

$$f_{(U_1,U_2)}(g^{-1}(x_1,x_2)) = \mathbf{1}_{\left\{\exp\left(-\frac{x_1^2+x_2^2}{2}\right)\in[0,1]\right\}} \times \mathbf{1}_{\left\{\frac{1}{2}+\frac{1}{\pi}\arctan\frac{x_2}{x_1+\sqrt{x_1^2+x_2^2}}\in[0,1]\right\}} = 1$$

Astfel găsim că pentru $(x_1,x_2) \in \mathbb{R}^2 \setminus \{(x,0) \,|\, x \geq 0\}$

$$f_{(X_1,X_2)}(x_1,x_2) = \frac{1}{2\pi}\exp\left(-\frac{x_1^2+x_2^2}{2}\right) = \left(\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x_1^2}{2}\right)\right)\left(\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x_2^2}{2}\right)\right)$$

ceea ce arată că variabilele aleatoare X_1 și X_2 sunt independente și repartizate $\mathcal{N}(0,1)$.

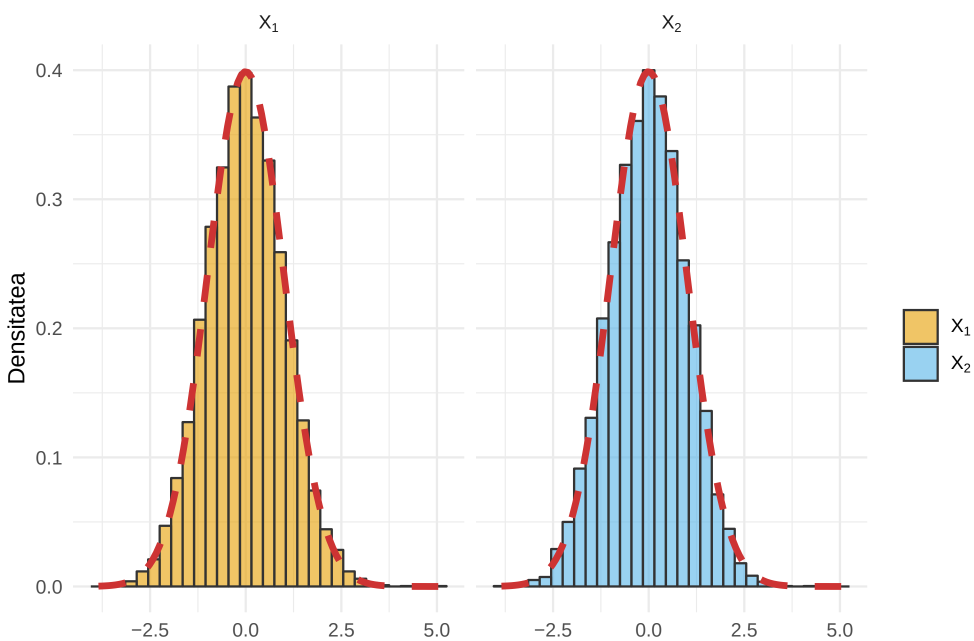
O altă soluție ar fi să notăm cu $R = \sqrt{-2\log(U_2)}$ şi $\Theta = 2\pi U_1$, atunci $(X_1, X_2) = g(R, \Theta)$ cu $g(r,\theta) = (r\cos(\theta), r\sin(\theta))$, $g : [0,\infty) \times [0, 2\pi] \to \mathbb{R}^2$. Cum U_1 şi U_2 sunt independente obţinem că şi R şi Θ sunt independente (ca funcţii de v.a. independente). Mai mult, cum $U_1 \sim \mathcal{U}([0,1])$ avem că $\Theta \sim \mathcal{U}([0,2\pi])$ iar din $R = h(U_2)$ cu $h(u) = \sqrt{1-2\log(u)}$ rezultă

$$f_R(r) = f_{U_2}(h^{-1}(r)) \left|\frac{d}{dr}h^{-1}(r)\right| = |r|e^{-\frac{r^2}{2}}$$

Obţinem astfel că

$$\begin{aligned} f_{(X_1,X_2)}(x_1,x_2) &= f_{(R,\Theta)}\left(g^{-1}(x_1,x2)\right)\left|\det J_{g^{-1}}\right| = f_{(R,\Theta)}\left(\sqrt{x_1^2+x_2^2}, \arctan\frac{x_2}{x_1}\right)\frac{1}{\sqrt{x_1^2+x_2^2}} \\ &= f_R\left(\sqrt{x_1^2+x_2^2}\right) f_\Theta\left(\arctan\frac{x_2}{x_1}\right)\frac{1}{\sqrt{x_1^2+x_2^2}} = \sqrt{x_1^2+x_2^2}e^{-\frac{x_1^1+x_2^2}{2}}\frac{1}{2\pi}\frac{1}{\sqrt{x_1^2+x_2^2}} \\ &= \left(\frac{1}{\sqrt{2\pi}}e^{-\frac{x_1^2}{2}}\right)\left(\frac{1}{\sqrt{2\pi}}e^{-\frac{x_2^2}{2}}\right) \end{aligned}$$

unde am folosit faptul că determinantul Jacobian-ului este $\det J_{g^{-1}} = \frac{1}{\sqrt{x_1^2+x_2^2}}$. Astfel densitatea cuplului (X_1, X_2) se poate scrie ca un produs de densităţi care depind de x_1 şi respectiv x_2 ceea ce conduce la concluzia problemei (densităţile din factorizare sunt tocmai densităţile normalei standard).



b) Din punctul precedent avem $R^2 = -2\log U_1$ și $\Theta = 2\pi U_2$. Am văzut că $\Theta \sim \mathcal{U}([0,2\pi])$. Cum funcția cuantilă F^{-1} a repartiției $\mathcal{E}\left(\frac{1}{2}\right)$ este $F^{-1}(u) = -2\log u$ avem concluzia.

Exercițiul 3

Fie $U_{i1}, U_{i2}, V_i,\ i \in \{1,2,\dots,n\}$, variabile aleatoare independente repartizate unifom $\mathcal{U}([0,1])$. Definim variabilele aleatoare

$$X_i = \begin{cases} 1, & U_{i1}^2 + U_{i2}^2 < 1 \\ 0, & \text{altfel} \end{cases} \qquad \text{şi} \quad Y_i = \sqrt{1 - V_i^2},\ \ i \in \{1,2,\dots,n\}$$

Considerăm variabilele aleatoare $\hat{\pi}_1 = \frac{4}{n}\sum_{i=1}^{n} X_i$ și $\hat{\pi}_2 = \frac{4}{n}\sum_{i=1}^{n} Y_i$. Calculați media și varianța acestor variabile și stabiliți care este mai eficientă[1] în estimarea lui π.

Pentru $\hat{\pi}_1$: observăm că v.a. X_i sunt v.a. de tip Bernoulli cu

$$\begin{aligned}
\mathbb{P}(X_i = 1) &= \mathbb{P}\left(U_{i1}^2 + U_{i2}^2 < 1\right) = \iint_{\{u^2+u^2<1\}\cap[0,1]^2} f_{(U_{i1},U_{i2})}(u,v)\,dudv \\
&\overset{indep.}{=} \iint_{\{u^2+u^2<1\}\cap[0,1]^2} f_{U_{i1}}(u) f_{U_{i2}}(v)\,dudv = \int_0^1\int_0^{\sqrt{1-u^2}} 1\,dvdu = \int_0^1 \sqrt{1-u^2}\,du \\
&\overset{u=\sin\alpha}{=} \int_0^{\frac{\pi}{2}} \cos^2\alpha\,d\alpha = \int_0^{\frac{\pi}{2}} \frac{\cos 2\alpha + 1}{2}\,d\alpha = \frac{\pi}{4} + \frac{1}{2}\int_0^{\frac{\pi}{2}} \cos 2\alpha\,d\alpha = \frac{\pi}{4}
\end{aligned}$$

O altă variantă de calcul pentru $\mathbb{P}(X_i = 1)$ era să observam că această probabilitate se exprima şi ca raportul dintre aria mulţimii $\{(u,v) \in [0,1]^2\,|, u^2 + u^2 < 1\}$ şi cea a pătratului $[0,1]^2$, deci tot $\frac{\pi}{4}$.

Dacă $T = \sum_{i=1}^{n} X_i$ atunci $T \sim \mathcal{B}\left(n, \frac{\pi}{4}\right)$ de unde avem că media este $\mathbb{E}[T] = \frac{n\pi}{4}$ iar varianţa

$$\mathbb{V}[T] = n\frac{\pi}{4}\left(1 - \frac{\pi}{4}\right).$$

Cum $\hat{\pi}_1 = \frac{4}{n}T$ deducem că $\mathbb{V}[\hat{\pi}_1] = \frac{4\pi-\pi^2}{n}$. Din *Legea Numerelor Mari* obţinem că $\hat{\pi}_1 = \frac{4}{n}\sum_{i=1}^{n} X_i \overset{a.s.}{\to} 4\mathbb{E}[X_1] = 4\mathbb{P}(X_1 = 1) = \pi$.

Pentru $\hat{\pi}_2$, să observăm pentru inceput că media lui Y_1 este

$$\mathbb{E}[Y_1] = \int_0^1 \sqrt{1-u^2}\,du = \frac{\pi}{4}$$

iar varianţa lui Y_1 este

$$\mathbb{V}[Y_1] = \mathbb{E}[Y_1^2] - \mathbb{E}^2[Y_1] = \int_0^1 1 - u^2\,du - \frac{\pi^2}{16} = \frac{2}{3} - \frac{\pi^2}{16}.$$

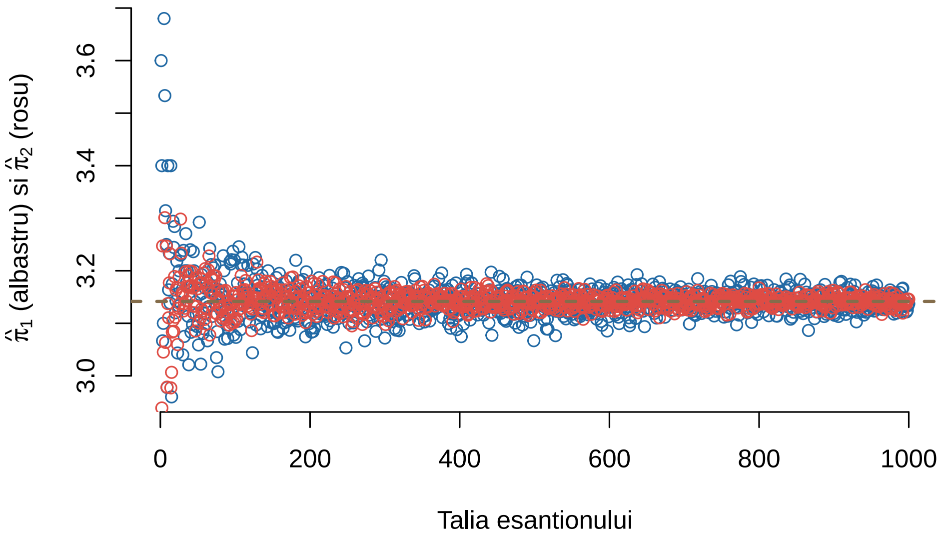
Prin aplicarea *Legii Numerelor Mari* rezultă că

$$\hat{\pi}_2 = \frac{4}{n}\sum_{i=1}^{n} Y_i \overset{a.s.}{\to} 4\mathbb{E}[Y_1] = \pi$$

iar varianţa lui $\hat{\pi}_2$ este

$$\mathbb{V}[\hat{\pi}_2] = \frac{16}{n^2}\sum_{i=1}^{n}\mathbb{V}[Y_i] = \frac{16}{n}\left(\frac{2}{3} - \frac{\pi^2}{16}\right).$$

Pentru a vedea care dintre cei doi estimatori este mai eficient trebuie să verificăm care are varianţa mai mică. Cum $\frac{32}{3} < 12 < 4\pi$ rezultă că $\mathbb{V}[\hat{\pi}_2] < \mathbb{V}[\hat{\pi}_1]$ deci al doilea estimator este mai eficient.



Exercițiul 4

Fie $U_1, U_2, \dots, U_n$ variabile aleatoare independente și repartizate $\mathcal{U}([0,1])$ și $S_n = \sum_{i=1}^n U_i$. Dacă variabila aleatoare N este definită prin

$$N = \min\{k \,|\, S_k > 1\}$$

atunci:

a) Arătați că dacă $0 \leq t \leq 1$ atunci $\mathbb{P}(S_k \leq t) = \frac{t^k}{k!}$.

b) Determinați $\mathbb{E}[N]$ și $Var[N]$.

a) Pentru a calcula probabilitatea $\mathbb{P}(S_k \leq t)$ cu $0 < t < 1$ să ne reamintim că dacă X și Y sunt două variabile aleatoare independente cu densitățile f_X și f_Y atunci densitatea sumei $Z = X + Y$ (convoluția) este dată de

$$f_Z(z) = \int f_X(z-t) f_Y(t)\, dt.$$

Fie f_n densitatea variabilei aleatoare S_n pentru $n \geq 1$. Avem, pentru $0 < x < 1$, că $f_1(x) = 1$ și pentru a calcula densitatea f_{n+1} a variabilei aleatoare S_{n+1} să observăm că $S_{n+1} = S_n + U_{n+1}$ cu S_n și U_{n+1} variabile aleatoare independente, de unde aplicând formula pentru densitatea sumei deducem că

$$f_{n+1}(x) = \int_0^x f_n(t)\, dt, \quad n \geq 1.$$

Prin inducție rezultă că $f_n(x) = \frac{x^{n-1}}{(n-1)!}$ pentru $0 < x < 1$ de unde

$$\mathbb{P}(S_n \leq t) = \int_0^t f_n(x)\,dx = \int_0^t \frac{x^{n-1}}{(n-1)!}\,dx = \frac{t^n}{n!}.$$

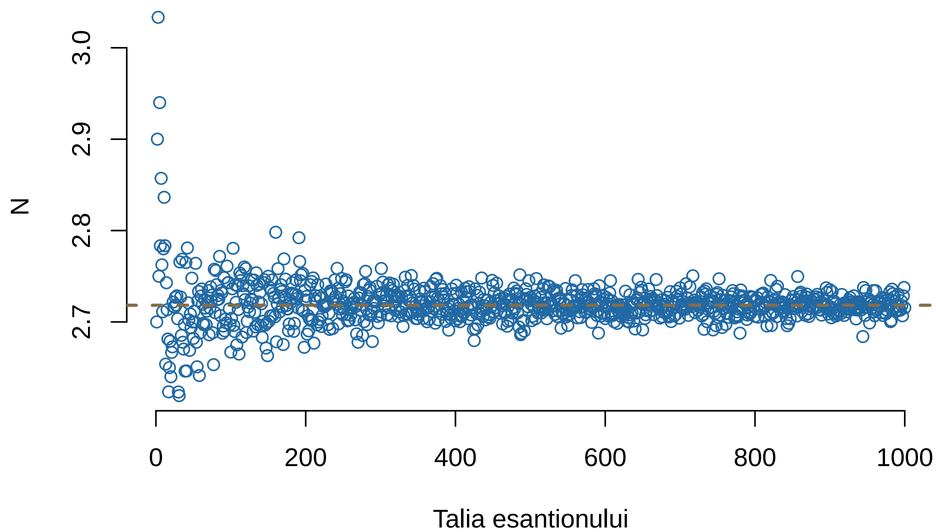
b) Pentru $n \geq 2$ să observăm că $\mathbb{P}(N = n) = \mathbb{P}(S_{n-1} < 1 \leq S_n)$ de unde

$$\mathbb{P}(N = n) = \mathbb{P}(S_{n-1} < 1) - \mathbb{P}(S_n < 1) = \frac{1}{(n-1)!} - \frac{1}{n!} = \frac{n-1}{n!}.$$

Pentru medie avem

$$\mathbb{E}[N] = \sum_{n=2}^{\infty} n\mathbb{P}(N = n) = \sum_{n=2}^{\infty} n\frac{n-1}{n!} = \sum_{n=2}^{\infty} \frac{1}{(n-2)!} = e.$$

În mod similar se poate arăta că $Var[N] = e(3-e)$.[2]



Exercițiul 5

Fie $(E_n)_{n\geq 1}$ un șir de variabile aleatoare independente și repartizate $\mathcal{E}(\lambda)$.

a) Pentru $n \geq 1$ definim

$$f_n(x) = e^{-\lambda x}\frac{\lambda^n x^{n-1}}{(n-1)!}\mathbf{1}_{\{x\geq 0\}}, \quad x \in \mathbb{R}.$$

Arătați că f_n este o densitate de repartiție pentru orice $n \geq 1$. Repartiția a cărei densitate este f_n se numește repartiția Gamma de parametrii $n \geq 1$ și λ și se notează cu $\Gamma(n, \lambda)$.

[2] Această metodă de a estima e este discutată în lucrarea: Russell, K.G. *Estimating the value of e by simulation*, The American Statistician, Vol. 45, Nr. 1, pp 66-68, 1991.

b) Fie $S_n = \sum_{i=1}^{n} E_i$ pentru $n \geq 1$. Arătați că S_n este repartizată $\Gamma(n, \lambda)$.

c) Considerăm variabila aleatoare

$$N = \max\{n \geq 1 \,|\, S_n \leq 1\}$$

cu convenția $N = 0$ dacă $X_1 > 1$. Arătați că variabila aleatoare N este repartizată $Pois(\lambda)$.

a) Prin inducție vom verifica că f_n este o densitate de repartiție. Să observăm că $f_n \geq 0$ prin urmare este suficient să arătăm că $\int_{\mathbb{R}} f_n(x)\,dx = 1$.

Pentru $n = 1$ avem

$$\int_{\mathbb{R}} f_1(x)\,dx = \int_0^\infty \lambda e^{-\lambda x}\,dx = 1.$$

Presupunem proprietatea adevărată pentru n și arătăm pentru $n+1$:

$$\int_{\mathbb{R}} f_{n+1}(x)\,dx = \left[-\frac{\lambda^n x^n}{n!}e^{-\lambda x}\right]_0^\infty + \int_0^\infty n\lambda^n \frac{x^{n-1}}{n!}e^{-\lambda x}\,dx = \int_{\mathbb{R}} f_n(x)\,dx = 1.$$

b) Pentru a determina repartiția lui S_n vom folosi noțiunea de *funcție generatoare de moment*[3], i.e. $M_E(t) = \mathbb{E}[e^{tE}]$.

Se poate calcula cu ușurință că

$$M_{E_i}(t) = \frac{\lambda}{\lambda - t}, \qquad t < \lambda$$

și cum variabilele aleatoare $E_1, \ldots, E_n$ sunt independente deducem că funcția generatoare de moment a sumei S_n este

$$M_{S_n}(t) = \mathbb{E}[e^{tS_n}] = \prod_{i=1}^{n} M_{E_i}(t) = \left(\frac{\lambda}{\lambda - t}\right)^n, \quad t < \lambda.$$

Știm că dacă $X \sim Gamma(n, \lambda)$ atunci $f_X(x) = \frac{\lambda^n}{(n-1)!}x^{n-1}e^{-\lambda x}$ iar funcția generatoare de moment este

$$M_X(t) = \left(\frac{\lambda}{\lambda - t}\right)^n, \quad t < \lambda.$$

Cum cele două funcții generatoare de moment sunt egale și ținând cont de faptul că funcția generatoare caracterizează repartiția, deducem că $S_n \sim Gamma(n, \lambda)$.

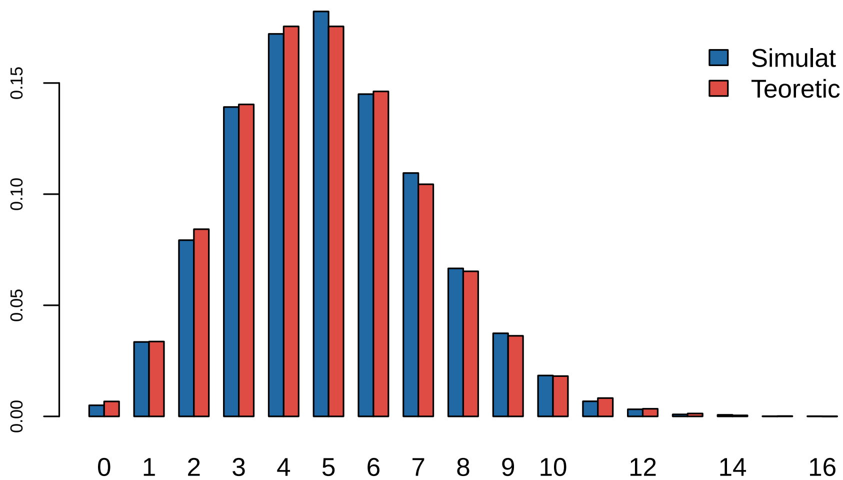
c) Pentru a demonstra că $N \sim Pois(\lambda)$ este suficient să calculăm $\mathbb{P}(N = n)$. Avem

$$\mathbb{P}(N = n) = \mathbb{P}\left(S_n \leq 1 < S_{n+1}\right) = \int_0^1 \mathbb{P}(E_{n+1} \geq 1 - u \,|\, S_n = u) f_{S_n}(u)\,du,$$

unde f_{S_n} este densitatea lui S_n de la punctul a). Ținând seama că E_{n+1} și S_n sunt independente și cum $\mathbb{P}(E_{n+1} \geq 1 - u) = e^{-\lambda(1-u)}$ avem că

[3] Problema se poate face și fără această noțiune, ținând seama de schimbarea de variabilă $\phi : (x_1, \ldots, x_n) \to (s_1, \ldots, s_n)$ cu $s_n = \sum_{k=1}^{n} x_k$ a cărei inversă ϕ^{-1} este dată prin $x_1 = s_1$ și $x_k = s_k - s_{k-1}$. Determinantul matricii Jacobiene asociate lui ϕ^{-1} este 1 iar imaginea $\phi([0, \infty) = \{0 \leq s_1 \leq \cdots \leq s_n\}$ ceea ce conduce la rezultatul dorit.

$$\mathbb{P}(N=n) = \int_0^1 \mathbb{P}(E_{n+1} \geq 1-u \,|\, S_n = u) f_{S_n}(u)\, du = \int_0^1 e^{-\lambda(1-u)} \frac{\lambda^n}{(n-1)!} x^{n-1} e^{-\lambda x}\, du$$
$$= \frac{e^{-\lambda}\lambda^n}{(n-1)!} \int_0^1 u^{n-1}\, du = \frac{e^{-\lambda}\lambda^n}{n!}.$$



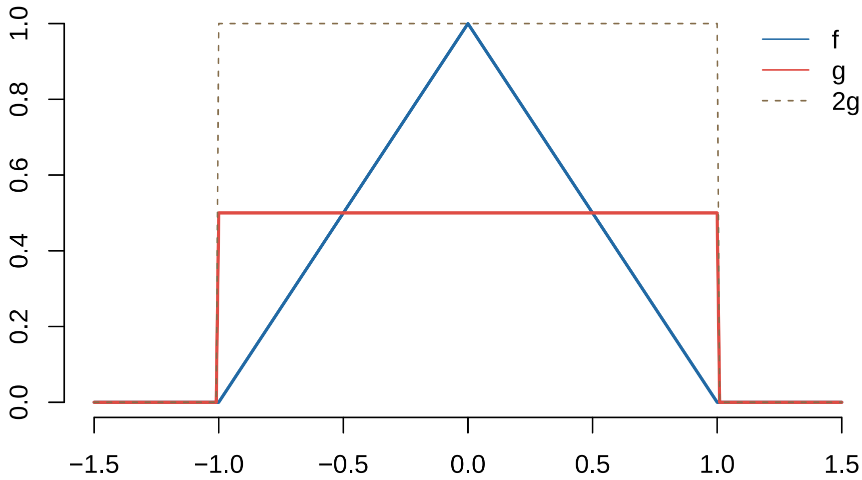
Exercițiul 6

Folosind metoda respingerii, propuneți o metodă de simulare pentru observații independente din densitatea de repartiție $f : x \mapsto (1-|x|)^+$.

Observăm că pentru toți $x \in \mathbb{R}$ avem

$$f(x) \leq \mathbf{1}_{\{x\in[-1,1]\}} = 2g(x),$$

unde $g(x) = \frac{1}{2}\mathbf{1}_{\{x\in[-1,1]\}}$ este densitatea repartiției uniforme pe $[-1, 1]$.



Pentru a simula din repartiția f procedăm astfel

1. simulăm X repartizată $\mathcal{U}[-1,1]$ ($X = 2V-1$ cu $V \sim \mathcal{U}(0,1)$)
2. simulăm $U \sim \mathcal{U}(0,1)$
3. repetăm procedeul până când $2Ug(X) < f(X)$.