Tema 2

Exercițiul 1[1]

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

Exercițiul 2

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

a) Determinaţi funcţia de repartiţie şi densitatea variabilelor m_n şi 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 $Z_n = n(1 - M_n)$. Arătaţi că $Z_n \xrightarrow{d} Z$, unde Z este o variabilă aleatoare a cărei funcţie de repartiţie este $F_Z(z) = 1 - e^{-z}$.

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

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

Considerăm variabilele aleatoare $\hat{\pi}_1 = \frac{4}{n}\sum_{i=1}^{n} X_i$ şi $\hat{\pi}_2 = \frac{4}{n}\sum_{i=1}^{n} Y_i$. Calculați media și varianța acestor variabile și stabiliți care este mai eficientă[2] în estimarea lui π.

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 $Var[N]$.

[1] Metoda de simulare prezentată în acest exercițiu se numește metoda Box-Muller
[2] Spunem că un estimator nedeplasat este mai eficient decat un altul dacă varianţa lui este mai mică

Exercițiul 5

Fie $(E_n)_{n\geq 1}$ un șir de variabile aleatoare independente, repartizate $Exp(\lambda)$, $S_n = \sum_{i=1}^{n} E_i$ pentru $n \geq 1$ și fie variabila aleatoare

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

a) Determinați repartiția lui S_n, $n \geq 1$.

b) Arătați că variabila aleatoare N este repartizată $Pois(\lambda)$.

Exercițiul 6

Fie $(X_n)_{n\geq 1}$ un şir de variabile aleatoare pozitive şi independente cu $\mathbb{E}[X_n] = c \in (0,1)$ pentru orice n. Dacă $Y_n = X_1 X_2 \cdots X_n$ atunci ară taţi că $Y_n \xrightarrow{P} 0$.

Tema 2

Soluții

Exercițiul 1 (Metoda Box-Muller)

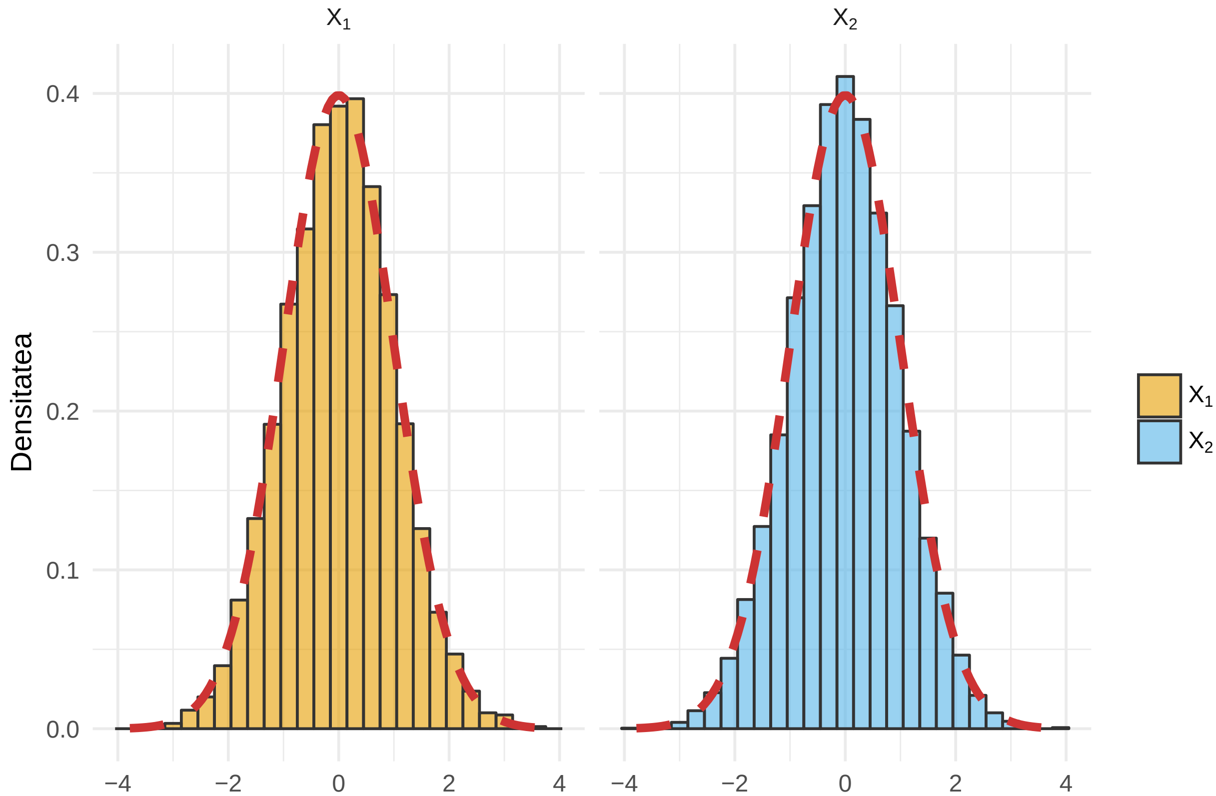
Fie $R = \sqrt{-2\log(U_2)}$ şi $\Theta = 2\pi U_1$, atunci $(X_1, X_2) = g(R, \Theta)$ cu $g(r, \theta) = (r\cos(\theta), r\sin(\theta))$, $g : [0, \infty) \times [0, 2\pi] \to \mathbb{R}^2$. Cum U_1 şi U_2 sunt independente obţinem că şi R şi Θ sunt independente (ca funcţii de v.a. independente). Mai 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).



Exercițiul 2

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 3

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

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

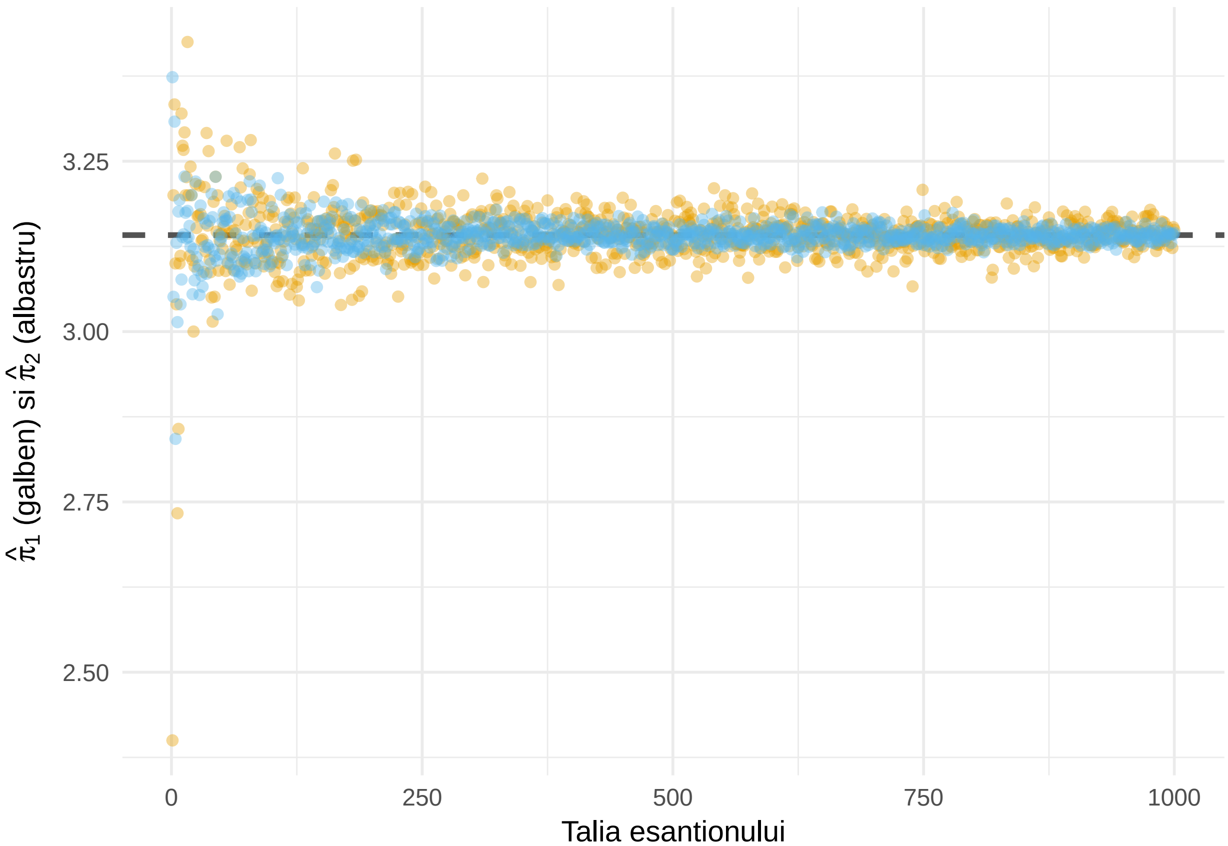
Prin aplicarea *Legii Numerelor Mari* rezultă că

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

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

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

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

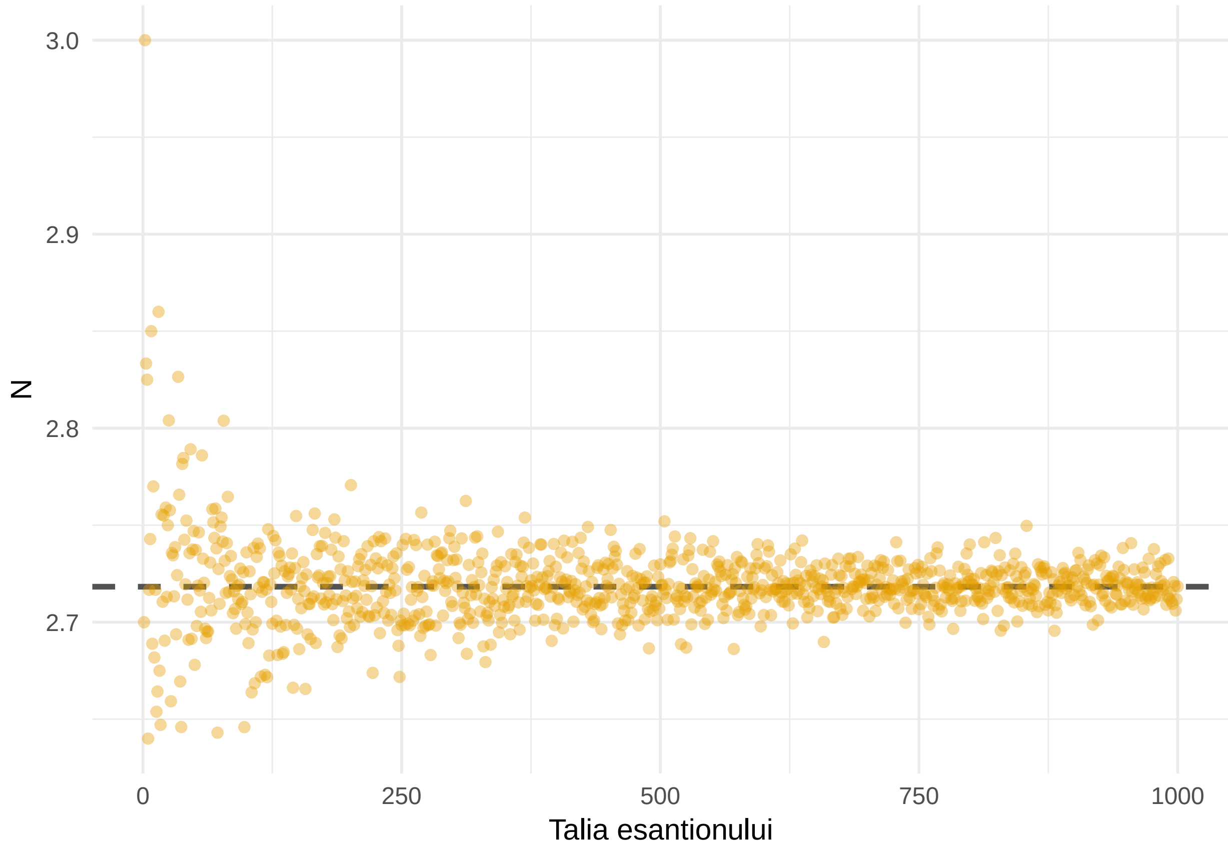
b) Pentru $n \geq 2$ să observăm că $\mathbb{P}(N = n) = \mathbb{P}(S_{n-1} < 1 \leq S_n)$ de unde

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

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 $Var[N] = e(3 - e)$.[1]



Exercițiul 5

a) Pentru a determina repartiția lui S_n vom folosi noțiunea de *funcție generatoare de moment*[2], 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 $E_1, \dots, E_n$ sunt independente deducem că funcția generatoare de moment a sumei S_n este

[1] 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.

[2] Problema se poate face și fără această noțiune, ținând seama de schimbarea de variabilă $\phi : (x_1, \dots, x_n) \to (s_1, \dots, s_n)$ 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 $\phi([0, \infty) = \{0 \leq s_1 \leq \dots \leq s_n\}$ ceea ce conduce la rezultatul dorit.

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

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

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

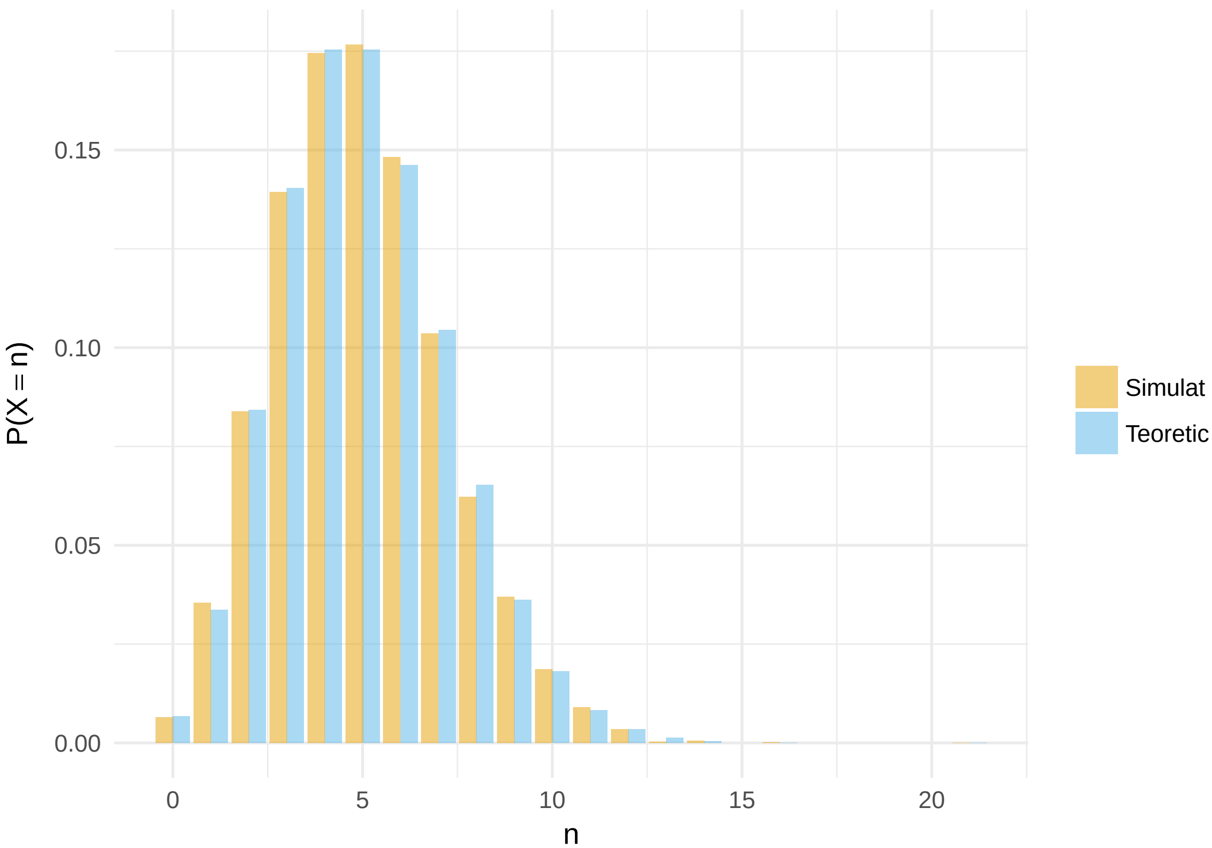
Cum cele două funcții generatoare de moment sunt egale și ținând cont de faptul că funcția generatoare caracterizează repartiția, deducem că $S_n \sim Gamma(n, \lambda)$.

b) 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

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



Exercițiul 6

Cum variabilele aleatoare $X_1, \dots, X_n$ sunt independente avem că $\mathbb{E}[Y_n] = \mathbb{E}[X_1] \cdot \mathbb{E}[X_2] \cdots \mathbb{E}[X_n] = c^n$, $c \in (0,1)$. Aplicand inegalitatea lui Markov obţinem, pentru $\varepsilon > 0$,

$$\mathbb{P}(|Y_n| > \varepsilon) \leq \frac{\mathbb{E}[|Y_n|]}{\varepsilon} = \frac{c^n}{\varepsilon} \overset{n \to \infty}{\to} 0,$$

unde am ţinut seama de faptul că v.a. sunt pozitive, deci $|Y_n| = Y_n$. Cum $\varepsilon > 0$ a fost ales arbitrar rezultă că $Y_n \xrightarrow{\mathbb{P}} 0$.