Proiect

Notă: Raportul poate fi scris în *Microsoft Word* sau LATEX(pentru ușurință recomand folosirea pachetului *rmarkdown* din *R* - mai multe informații găsiți pe site la secțiunea *Link-uri utile*). Toate simulările, figurile și codurile folosite trebuie incluse în raport. Se va folosi doar limbajul *R*.

1 Problema 1

Considerăm următoarele distribuții: $\mathrm{Bin}(n,p)$, $\mathrm{Pois}(\lambda)$, $\mathrm{Exp}(\lambda)$, $\mathcal{N}(\mu,\sigma^2)$

1. Generați $N = 1000$ de realizări independente din fiecare repartiție și calculați media și varianța eșantionului.
2. Ilustrați grafic funcțiile de masă, respectiv funcțiile de densitate pentru fiecare din repartițiile din enunțul problemei. Considerați câte 5 seturi de parametrii diferiți pentru fiecare repartiție și suprapuneți graficele pe aceeași figură pentru fiecare rapetiție. Adăugați și legenda.
3. Pentru seturile de parametrii de la punctul anterior trasați funcțiile de repartiție pentru fiecare repartiție (tot suprapuse) și adăugați legenda corespunzătoare.

Scopul următoarelor subpuncte este de a evalua acuratețea unor aproximări ale funcției de repartiție a binomialei $\mathcal{B}(n,p)$. Vom compara următoarele patru aproximări (cu excepția aproximării Camp-Paulson, celelalte trei au fost văzute la curs):

a) *Aproximarea Poisson*

$$F_{n,p}(k) \approx F_\lambda(k) = \sum_{x=0}^{k} e^{-\lambda}\frac{\lambda^x}{x!}, \quad \lambda = np$$

b) *Aproximarea Normală* (rezultată din Teorema Limită Centrală)

$$F_{n,p}(k) \approx \Phi\left(\frac{k-np}{\sqrt{np(1-p)}}\right)$$

c) *Aproximarea Normală cu factor de corecție*

$$F_{n,p}(k) \approx \Phi\left(\frac{k+0.5-np}{\sqrt{np(1-p)}}\right)$$

d) *Aproximarea Camp-Paulson*[1]

$$F_{n,p}(k) \approx \Phi\left(\frac{c-\mu}{\sigma}\right)$$

pentru $c = (1-b)r^{\frac{1}{3}}$, $\mu = 1-a$ și $\sigma^2 = a + br^{\frac{2}{3}}$ unde $a = \frac{1}{9(n-k)}$, $b = \frac{1}{9(k+1)}$ și respectiv $r = \frac{[(k+1)(1-p)]}{[p(n-k)]}$.

Se cere:

[1] A se vedea articolul lui Camp, B. H. - *Approximation to the Point Binomial*, Annals of Mathematical Statistics, 22, pp. 130-131, 1951.

4. Pentru fiecare $n \in \{25, 50, 100\}$ și fiecare $p \in \{0.05, 0.1\}$ să se afișeze un tabel cu șase coloane (`k`, `Binomiala`, `Poisson`, `Normala`, `Normala Corecție`, `Camp-Paulson`) în care să apară aproximările de mai sus pentru funcția de repartiție și de masă a binomialei, pentru $k \in \{1, 2, \dots, 10\}$.

5. Pentru a cuantifica acuratețea aproximărilor de mai sus vom folosi ca metrică, *eroarea maximală absolută* dintre două funcții de repartiții F și H (numită și distanța Kolmogorov) dată de formula

$$d_K(F(k), H(k)) = \max_k |F(k) - H(k)|\,.$$

Pentru fiecare $n \in \{25, 50, 100\}$ ilustrați pe același grafic erorile maximale absolute (folosind diferite culori și simboluri pentru puncte) dintre funcția de repartiție binomială și cele patru aproximări de mai sus considerând $0.01 \leq p \leq 0.5$. Ce observați ?

2 Problema 2

Obiectivul acestui exercițiu este de a simula un vector aleator (X_1, X_2) repartizat uniform pe discul unitate $D(1)$ (discul de centru $(0, 0)$ și de rază 1). Densitatea acestuia este

$$f_{(X_1,X_2)}(x_1, x_2) = \frac{1}{\pi}\mathbf{1}_{D(1)}(x_1, x_2).$$

Pentru aceasta vom folosi două metode. O primă metodă este metoda de simulare prin acceptare și respingere. Această metodă este des utilizată pentru generarea unei v.a. repartizate uniform pe o mulțime oarecare E. Metoda constă în generarea unei v.a. X repartizată uniform pe o mulțime $F \supset E$ mai simplă decât E, apoi de a testa dacă X se află în E sau nu. În caz afirmativ, păstrăm X altfel generăm o nouă realizare a lui X pe F.

1. Justificați teoretic că putem simula un vector (cuplu) aleator repartizat uniform pe pătratul $[-1, 1]^1$ plecând de la două v.a. independente repartizate uniform pe segmentul $[-1, 1]$.

2. Prin metoda acceptării și respingerii simulați $N = 1000$ de puncte independente repartizate uniform pe discul unitate $D(1)$. Reprezentați grafic punctele (X_i, Y_i) din interiorul discului unitate cu albastru, și pe celelalte cu roșu.

3. Calculați media aritmetică a distanței care separă cele N puncte de origine. Comparați rezultatul cu media teoritică a variabilei corespunzătoare.

O a doua metodă de simulare a unui punct (X, Y) repartizat uniform pe $D(1)$ constă în folosirea schimbării de variabilă în coordonate polare: $X = R\cos(\Theta)$ și $Y = R\sin(\Theta)$.

4. Plecând de la densitatea cuplului (X, Y), găsiți densitatea v.a. R și Θ.[2]

5. Simulați $N = 1000$ de puncte prin această metodă și ilustrați grafic aceste puncte (incluzand conturul cercului).

[2] *Indicație:* Aici puteți folosi următorul rezultat bazat pe formula de schimbare de variabilă în cazul multidimensional: Fie $\mathbf{X} = (X_1, \dots, X_n)$ un vector aleator cu densitatea $f_{\mathbf{X}}$ și $g : \mathbb{R}^n \to \mathbb{R}^n$ o funcție diferențiabilă de clasă $\mathcal{C}^1$, injectivă și cu Jacobianul nenul. Atunci vectorul aleator $\mathbf{Y} = g(\mathbf{X})$ are densitatea $f_{\mathbf{Y}}(y) = f_{\mathbf{X}}\left(g^{-1}(y)\right)\left|\det J_{g^{-1}}(y)\right|$ dacă $y \in Im(g)$ și 0 altfel