

# Proiect

## Grupele 311, 321

**Notă:** Rezolvarea problemelor de mai jos va fi realizată în **R** (scripturile trebuie să fie comentate) și va fi însoțită de un document text (.pdf sau .docx) care să conțină comentarii și concluzii, acolo unde sunt cerute.

**Punctaj:** 1. 0.5p , 2. 0.25p, 3. 0.25p 4. 0.25p 5. 0.25p

1. Generați 10000 de variabile aleatoare folosind **metoda transformării inverse** pentru repartițiile definite mai jos:

a) Repartiția logistică are densitatea de probabilitate  $f(x) = \frac{e^{-\frac{x-\mu}{\beta}}}{\beta \left(1 + e^{-\frac{x-\mu}{\beta}}\right)^2}$  și funcția de repartiție

$$F(x) = \frac{1}{1 + e^{-\frac{x-\mu}{\beta}}}.$$

b) Repartiția Cauchy are densitatea de probabilitate  $f(x) = \frac{1}{\pi\sigma} \frac{1}{1 + \left(\frac{x-\mu}{\sigma}\right)^2}$  și funcția de repartiție

$$F(x) = \frac{1}{2} + \frac{1}{\pi} \arctan\left(\frac{x-\mu}{\sigma}\right).$$

Comparați rezultatele obținute cu valorile date de funcțiile **rlogis** și respectiv **rcauchy** (funcțiile de repartiție predefinite în R pentru repartițiile logistică și respectiv Cauchy). Ilustrați grafic aceste rezultate.

2. Folosiți **metoda respingerii** pentru a genera observații din densitatea de probabilitate definită prin  $f(x) \propto e^{-\frac{x^2}{2}} [\sin(6x)^2 + 3 \cos(x)^2 \sin(4x)^2 + 1]^1$  parcurgând pașii următori:

- a) Reprezentați grafic  $f(x)$  și arătați că aceasta este mărginită de  $Mg(x)$  unde  $g(x)$  este densitatea de probabilitate a repartiției normale standard. Determinați o valoare potrivită pentru constanta  $M$ , chiar dacă nu este optimă<sup>2</sup>.
- b) Generați 2500 de observații din densitatea de mai sus folosind metoda respingerii.
- c) Deduceți, pornind de la rata de acceptare a acestui algoritm, o aproximare a *constantei de normalizare* a lui  $f(x)$ , apoi comparați histograma valorilor generate cu reprezentarea grafică a lui  $f(x)$  normalizată.

### 3. Metoda Monte Carlo pentru aproximarea unor integrale

Punctul de plecare al metodei Monte Carlo pentru aproximarea unei integrale este nevoia de a evalua expresia  $\mathbb{E}_f[h(X)] = \int_{\chi} h(x)f(x) dx$ , unde  $\chi$  reprezintă mulțimea de valori a variabile aleatoare  $X$  (care este, de obicei, suportul densității  $f$ ).

Principiul metodei Monte Carlo este de a aproxima expresia de mai sus cu media de selecție  $\bar{h}_n = \frac{1}{n} \sum_{j=1}^n h(X_j)$  pornind de la un eșantion  $X_1, X_2, \dots, X_n$  din densitatea  $f$ , întrucât aceasta converge a.s. către  $\mathbb{E}_f[h(X)]$ , conform legii numerelor mari. Mai mult, atunci când  $h(X)^2$  are medie finită viteza de convergență a lui  $\bar{h}_n$  poate fi determinată întrucât convergența este de ordin  $\mathcal{O}(\sqrt{n})$  iar varianța aproximării este  $Var(\bar{h}_n) = \frac{1}{n} \int_{\chi} (h(x) - \mathbb{E}_f[h(X)])^2 f(x) dx$ , cantitate care poate fi de asemenea aproximată prin  $v_n = \frac{1}{n^2} \sum_{j=1}^n (h(X_j) - \bar{h}_n)^2$ .

Mai precis, datorită teoremei limită centrală, pentru un  $n$  suficient de mare expresia

$$\frac{\bar{h}_n - \mathbb{E}_f[h(X)]}{\sqrt{v_n}}$$

<sup>1</sup>Notăția  $\propto$  înseamnă că  $f(x)$  este proporțională cu expresia din dreapta

<sup>2</sup>**Indiciu:** Folosiți funcția **optimise** din R

poate fi aproximată cu o normală standard, ceea ce conduce la posibilitatea construirii unui test de convergență și a unor margini pentru aproximarea lui  $\mathbb{E}_f[h(X)]$ .

**Cerință:**

Pentru funcția  $h(x) = (\cos(50x) + \sin(20x))^2$  construiți o aproximare a integralei acesteia pe intervalul  $[0,1]$  după cum urmează: valoarea integralei poate fi văzută ca fiind media funcției  $h(X)$  unde  $X$  este repartizată uniform pe  $[0,1]$ . Urmărind algoritmul dat de metoda Monte Carlo construiți programul R care determină aproximarea acestei integrale. Comparați rezultatul obținut cu cel analitic și cu cel numeric obținut folosind funcția `integrate`. Atașați reprezentările grafice pe care le considerați utile pentru a putea observa eficiența metodei.

4. Pornind de la un set de date din cele oferite de R (*fiecare student își alege singur setul de date*) realizați o analiză de tipul “statistică descriptivă” a acestora (medie, quartile, histogramă, boxplot, boxplot comparativ, identificare de outlieri, etc.). Documentați corespunzător acest proces și explicați ce concluzii puteți trage în urma acestei analize asupra datelor. Includeți în fișier și o descriere a datelor și sursa lor.
5. Construiți o funcție în R care să calculeze **marginea inferioară Rao-Cramer (MIRC)** pentru varianța estimatorilor parametrilor repartițiilor *Poisson* și *Exponențială* pentru un eșantion de talie  $n$  (generați voi eșantionul într-o manieră corespunzătoare și folosiți-l în apelul funcției). Generalizați funcția astfel încât să poată fi folosită și pentru alte repartiții.