# Proiect la Statistică

Grupele 311,312, 321,322

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

1. Generaţi 100 000 de valori dintr-o variabilă aleatoare folosind **metoda transformării inverse** pentru repartiţiile definite mai jos:

a) *Repartiţia logistică* are densitatea de probabilitate $f(x) = \frac{1}{\beta} \frac{e^{-(x-\mu)/\beta}}{(1+e^{-(x-\mu)/\beta})^2}$ şi funcţia de repartiţie $F(x) = \frac{1}{1+e^{-(x-\mu)/\beta}}$ .

b) *Repartiţia Cauchy* are densitatea de probabilitate $f(x) = \frac{1}{\pi\sigma} \frac{1}{1+\left(\frac{x-\mu}{\sigma}\right)^2}$ şi funcţia de repartiţie $F(x) = \frac{1}{2} + \frac{1}{\pi}\arctan\left(\frac{x-\mu}{\sigma}\right)$.

Comparaţi rezultatele obţinute cu valorile date de funcţiile **rlogis** şi respectiv **rcauchy(**funcţiile de repartiţie predefinite în **R** pentru repartiţiile logistică şi respectiv Cauchy**)**. Ilustraţi grafic aceste rezultate.

2. Folosiţi **metoda respingerii** pentru a genera observaţii din densitatea de probabilitate definită prin $f(x) \propto \exp(-(x-2)^2/8)(\sin(2x)^2 - 2\cos(x)^2\sin(3x)^2 + 5)$ parcurgând paşii următori:

(**OBS**: Notaţia "$\propto$" înseamnă că $f(x)$ *este proporţional cu* expresia din dreapta)

a) Reprezentaţi grafic $f(x)$ şi arătaţi că aceasta este mărginită de $Mg(x)$ unde $g(x)$ este densitatea de probabilitate a repartiţiei normale de medie 2 şi dispersie 4. Determinaţi o valoare potrivită pentru constanta $M$ , chiar dacă nu este optimă.
(**Indiciu**: Folosiţi funcţia *optimise* din **R**)

b) Generaţi 100000 de observaţii din densitatea de mai sus folosind metoda respingerii.

c) Deduceţi, pornind de la rata de acceptare a acestui algoritm, o aproximare a *constantei de normalizare* a lui $f(x)$, apoi comparaţi histograma valorilor generate cu reprezentarea grafică a lui $f(x)$ normalizată.

3. **<u>Metoda Monte Carlo pentru aproximarea unor integrale</u>**

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

*Principiul metodei Monte Carlo este de a aproxima expresia de mai sus cu media de selecţie* $\overline{h_n} = \frac{1}{n}\sum_{j=1}^{n} h(x_j)$ *pornind de la un eşantion* $\left(X_1, X_2, \ldots X_n\right)$ *din densitatea* $f$*, întrucât aceasta converge a.s. către* $\mathrm{E}_f(h(X))$*, conform legii numerelor mari. Mai mult, atunci când* $h^2(X)$ *are medie finită viteza de convergenţă a lui* $\overline{h_n}$ *poate fi determinată întrucât convergenţa este de ordin* $O(\sqrt{n})$ *iar varianţa aproximării este* $\mathrm{var}(\overline{h_n}) = \frac{1}{n}\int_{\chi}(h(x) - \mathrm{E}_f(h(X)))^2 f(x)dx$*, cantitate care poate fi de asemenea aproximată prin* $v_n = \frac{1}{n^2}\sum_{j=1}^{n}(h(x_j) - \overline{h_n})^2$.

*Mai precis, datorită teoremei limită centrală, pentru un* $n$ *suficient de mare expresia*
$\frac{\overline{h_n} - \mathrm{E}_f(h(X))}{\sqrt{v_n}}$ *poate fi aproximată cu o normală standard, ceea ce conduce la posibilitatea construirii unui test de convergenţă şi a unor margini pentru aproximarea lui* $\mathrm{E}_f(h(X))$.

**Cerinţă:**

Pentru funcţia $h(x) = (1 - x^2)^{\frac{3}{2}}$ construiţi o aproximare a integralei acesteia pe intervalul [0,1] după cum urmează:

Valoarea integralei poate fi văzută ca fiind media funcţiei $h(X)$ unde $X$ este repartizată uniform pe [0,1]. Urmărind algoritmul dat de metoda Monte Carlo construiţi programul **R** care determină aproximarea acestei integrale. Comparaţi rezultatul obţinut cu cel analitic. Ataşaţi reprezentările grafice pe care le consideraţi utile pentru a putea observa eficienţa metodei.

4. Construiţi două funcţii în **R** ***frcpois*** şi respectiv ***frcexp*** care să calculeze **marginea inferioară Rao-Cramer (MIRC)** pentru varianţa estimatorilor parametrilor repartiţiilor ***Poisson*** şi respectiv ***Exponenţială*** pentru un eşantion de dimensiune n (generaţi voi un asemenea eşantion într-o manieră corespunzătoare şi folosiţi-l în apelul funcţiei!).

5. a) Construiți o funcție care calculează media unei variabile aleatoare, primind ca parametri de intrare tipul variabilei aleatoare(discretă sau continuă) și densitatea/funcția de masă, după caz. Comparați cu rezultatul teoretic.

b) Construiți o funcție care calculează dispersia unei variabile aleatoare, folosindu-vă de funcția care calculează media. Comparați cu rezultatul teoretic.

c) Construiți o funcție care calculează funcția generatoare de momente pentru variabila aleatoare X pentru un interval dat(discretizat de un vector t, care este parametru de intrare pentru funcție).

d) Aproximați media și dispersia variabilei aleatoare X folosind funcția generatoare de momente de la c). Comparați cu rezultatele obținute la a) și b).

6. a) Pentru repartiţiile ***logistică*** şi respectiv ***Cauchy***(*vezi* problema 1) construiţi *funcţia de verosimilitate* pentru parametrul $\mu$ considerând că parametrii $\beta$ şi respectiv $\sigma$ sunt cunoscuţi(alegeţi valori potrivite pentru aceştia).

b) Reprezentaţi grafic funcţiile de verosimilitate pentru cele două cazuri şi folosind funcţia ***optimise*** determinaţi o estimaţie pentru $\mu$ în baza unui eşantion de dimensiune 1000 pe care l-aţi construit în prealabil. Explicaţi modul în care aţi generat valorile din eşantion. Comentaţi şi interpretaţi rezultatele.