Laboratorul 6

Coeficientul de corelație (Pearson) al variabilelor aleatoare X și Y este

$$\rho(X,Y) = \frac{\text{cov}(X,Y)}{\sqrt{V(X)V(Y)}},$$

dacă cov(X,Y), V(X), V(Y) există și $V(X)V(Y) \neq 0$, unde $cov(X,Y) = E\Big((X-E(X))(Y-E(Y))\Big)$ este **covarianța** lui X și Y.

- $\triangleright \rho(X,Y) \in [-1,1]$ măsoară dependența liniară între variabilele aleatoare X și Y.
- \triangleright Dacă X şi Y sunt independente, atunci cov(X,Y)=0, adică X şi Y sunt necorelate.
- \triangleright Dacă $|\rho(X,Y)|=1$, atunci există $a,b\in\mathbb{R}$ astfel încât Y=aX+b sau X=aY+b.

Coeficientul de corelație (Pearson) de selecție al datelor de selecție x_1, \ldots, x_n (nu toate egale) și y_1, \ldots, y_n (nu toate egale) este

$$\bar{\rho} = \frac{\sum_{i=1}^{n} (x_i - \bar{x}_n)(y_i - \bar{y}_n)}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x}_n)^2 \sum_{i=1}^{n} (y_i - \bar{y}_n)^2}},$$

unde $\bar{x}_n = \frac{1}{n}(x_1 + \dots + x_n)$ și $\bar{y}_n = \frac{1}{n}(y_1 + \dots + y_n)$ sunt mediile de selecție.

- 1. i) Fie $X \sim N(1,1)$ și $Y \sim N(3,1)$ variabile aleatoare independente. Simulați $n \in \{100, 500, 1000\}$ valori pentru X, respectiv Y, apoi estimați următoarele valori: $V(X), \ V(Y), \ E(X \cdot Y), \ |\rho(X,Y)|, \ P(X < 2, Y > 1)$. Comparați rezultatele obținute cu valorile teoretice.
- ii) Fie $X \sim N(1,1)$ și Y = 3X+1 variabile aleatoare independente. Simulați $n \in \{100, 500, 1000\}$ valori pentru X, respectiv Y, apoi estimați următoarele valori: $V(X), V(Y), E(X \cdot Y), |\rho(X,Y)|, P(X < 2, Y > 1)$. Comparați rezultatele obținute cu valorile teoretice.

Funcții utile: var, mean, corrcoef, normcdf.

Metode Monte Carlo pentru integrare numerică

Fie $g:[a,b]\to[0,\infty)$ o funcție continuă și M>0 astfel încât $g(x)\leq M$, oricare ar fi $x\in[a,b]$. Considerăm următoarele metode pentru aproximarea integralei $\int_{a}^{b} g(x) dx$ folosind valori aleatoare.

Metoda 1:

- Considerăm $(U_n)_n$ şir de v.a. independente uniform distribuite pe [a,b] şi notăm $X_n = g(U_n)$.
- $(X_n)_n$ satisface LTNM, adică

$$\frac{1}{n} \sum_{k=1}^{n} X_k \xrightarrow{a.s.} E(X_1) = \frac{1}{b-a} \int_a^b g(t)dt.$$

• În simulări:

$$\int_a^b g(t)dt \approx (b-a)\frac{1}{n} \left(g(u_1) + \dots + g(u_n)\right), \text{ pentru } n \text{ suficient de mare},$$

unde u_1, \ldots, u_n sunt valori aleatoare generate independent conform distribuției uniforme pe intervalul [a,b].

Metoda 2:

- Fie (X,Y) un vector aleator care are distribuția uniformă pe $[a,b] \times [0,M]$.
- Folosind probabilitatea geometrică (a se revedea Laboratorul 2), avem:

$$P\left((X,Y) \text{ este sub graficul lui } g\right) = \frac{\text{aria subgraficului lui } g}{\text{aria dreptunghiului } [a,b] \times [0,M]} = \frac{1}{(b-a)M} \int_a^b g(t) dt.$$

• În simulări:

$$\int_a^b g(t)dt \approx \frac{\#\{k \in \{1,...,n\}: y_k \leq g(x_k)\}}{n} (b-a)M \text{ , pentru } n \text{ suficient de mare,}$$

unde $(x_1, y_1), \ldots, (x_n, y_n)$ sunt perechi de numere aleatoare generate independent conform distribuției uniforme pe dreptunghiul $[a, b] \times [0, M]$.

- 2. Implementați în Matlab cele două metode Monte Carlo pentru integrarea numerică a unei funcții continue $g:[a,b]\to[0,\infty)$. Testați programele realizate cu funcțiile:

funcții continue
$$g: [a, b] \to [0, \infty)$$
. Testați programele realizi \mathbf{i}) $g_1: [-2, 2] \to \mathbb{R}, g_1(x) = e^{-x^2}, x \in [-2, 2];$
 \mathbf{ii}) $g_2: [-1, 3] \to \mathbb{R}, g_2(x) = |\sin(e^x)|, x \in [-1, 3].$
 \mathbf{iii}) $g_3: [-1, 2] \to [0, \infty), g_3(x) = \begin{cases} \frac{x^2}{1+x^2}, & x \in [-1, 0]\\ \sqrt{2x-x^2}, & x \in (0, 2]. \end{cases}$

Comparați rezultatele obținute cu rezultatele date de funcția integral.