Laboratorul 10 - 2017

I. Generarea de numere pseudo-aleatoare ce urmează o distribuţie continuă dată

Fie X o variabilă aleatoare ce are funcţia de repartiţie F. Din teorie se ştie că F este continuă şi monoton crescătoare. Presupunem că F *este inversabilă*, adică există F^{-1}: pentru orice $y \in (0,1)$ există un unic $x \in \mathbb{R}$ astfel încât $F(x) = y$, ceea ce este echivalent cu $F^{-1}(y) = x$.

Procedeul de generare a numerelor aleatoare $Y(i)$, $i = \overline{1,N}$, *care au aceeaşi distribuţie ca* X *este*:

- Se citeşte numărul N, se defineşte funcţia F^{-1}.
- Se generează N numere aleatoare uniform distribuite pe [0,1]: $U(i)$, $i = \overline{1,N}$.
- Pentru fiecare $i = \overline{1,N}$: $Y(i) = F^{-1}(U(i))$.
- Se afişează: $Y(i), i = \overline{1,N}$.

Verificarea procedeului: Fie variabila aleatoare $U \sim Unif[0,1]$ şi definim variabila aleatoare $Y = F^{-1}(U)$. Arătăm că Y are aceeaşi funcţie de repartiţie ca X: pentru orice $y \in \mathbb{R}$ are loc

$$F_Y(y) = P(Y \leq y) = P(F^{-1}(U) \leq y) = P(U \leq F(y)) = F(y),$$

deci Y are aceeaşi distribuţie ca X, pentru că $F_Y = F$.

Exemplu: pentru $X \sim Exp(\lambda)$, atunci

$$F(x) = 1 - e^{-\lambda x},\ x > 0 \Longrightarrow F^{-1}(y) = -\frac{\ln(1-y)}{\lambda},\ y \in (0,1)$$

adică $-\frac{1}{\lambda}\ln(U) \sim Exp(\lambda)$, dacă $U \sim \mathcal{U}[0,1]$.

II. Generarea de numere pseudo-aleatoare care urmează distribuţia *Gamma*

Metoda 1: Fie $n \in \mathbb{N}, \beta > 0$ date. Considerăm n variabile aleatoare independente $X_1, \dots, X_n \sim Exp(\frac{1}{\beta})$

$$\Longrightarrow X_1 + \dots + X_n \sim Gamma(n, \beta)$$

adică $-\beta\ln(U_1) - \dots - \beta\ln(U_1) = -\beta\ln(U_1 \cdot \dots \cdot U_n) \sim Gamma(n,\beta)$, dacă $U_i \sim \mathcal{U}[0,1]$, $i = 1, 2..., n$ sunt variabile aleatoare independente.

Generarea de numere pseudo-aleatoare care urmează distribuţia $Gamma(\alpha,\beta)$ cu $\alpha > 0, \beta > 0$:

Metoda 2 (Marsaglia-Tsang):

▶ În cazul în care $X \sim Gamma(\alpha, 1)$ cu $\alpha \geq 1$:

Pasul 1: Fie $d = \alpha 1/3,\ c = 1/\sqrt{9d}$.

Pasul 2: Se generează $Z \sim N(0,1), U \sim \mathcal{U}(0,1)$ variabile aleatoare independente.

Pasul 3: Dacă $Z > 1/c$ şi $ln(U) < \frac{1}{2}Z^2 + d(1-V) + d \cdot ln(V)$, cu $V = (1+cZ)^3$, atunci $X = d \cdot V$, altfel se continuă cu Pasul 2.

▶ În cazul în care $X \sim Gamma(\alpha, 1)$ cu $0 < \alpha < 1$:

Generăm mai întâi $Y \sim Gamma(\alpha + 1, 1)$ şi $V \sim \mathcal{U}(0,1)$ independente, atunci $X = Y \cdot V^{\frac{1}{\alpha}} \sim Gamma(\alpha, 1)$.

▶ În cazul în care $X \sim Gamma(\alpha, \beta)$ cu $0 < \alpha, 0 < \beta, \beta \neq 1$:

Generăm mai întâi $Y \sim Gamma(\alpha, 1)$ şi atunci $X = \frac{Y}{\beta} \sim Gamma(\alpha,\beta)$.

Sursa: G. Marsaglia, W. Tsang, *A simple method for generating gamma variables.* ACM Transactions on Mathematical Software, 26(3):363-372, 2000.

III. Generarea de numere pseudo-aleatoare care urmează distribuţia *normală*

Metoda 3 (Box-Muller): Fie $U_1, U_2 \sim \mathcal{U}[0,1]$ variabile aleatoare independente.

$$\Longrightarrow\ Y_1 = \sqrt{-2\log U_1}\cos(2\pi U_2) \sim N(0,1),\ \ Y_2 = \sqrt{-2\log U_1}\sin(2\pi U_2) \sim N(0,1)$$

variabile aleatoare independente.

Metoda 4 (Metoda polară Marsaglia): Fie $V_1, V_2 \sim \mathcal{U}[-1,1]$ variabile aleatoare independente.
Fie $S = V_1^2 + V_2^2$. Dacă $0 < S \leq 1$, atunci

$$Z_1 = V_1\sqrt{\frac{-2ln(S)}{S}} \sim N(0,1),\ \ Z_2 = V_2\sqrt{\frac{-2ln(S)}{S}} \sim N(0,1)$$

sunt variabile aleatoare independente.

IV. Generarea de numere pseudo-aleatoare care urmează distribuţia χ^2 (*Chi-pătrat*)

Metoda 5:

$X_1, \ldots, X_n \sim N(0,1)$ variabile aleatoare independente $\Longrightarrow\ X_1^2 + \cdots + X_n^2 \sim \chi^2(n)$.

Metoda 6:

$$X \sim \chi^2(n)\ \iff\ X \sim Gamma\Big(\frac{n}{2}, 2\Big)$$

V. Generarea de numere pseudo-aleatoare care urmează distribuţia *Student*

Metoda 7:

$X \sim N(0,1), Y \sim \chi^2(n)$ variabile aleatoare independente $\Longrightarrow\ \dfrac{X}{\sqrt{\frac{Y}{n}}} \sim Student(n)$

Metoda 8 (Bailey):

Pasul 1: Fie $V_1, V_2 \sim \mathcal{U}[-1,1]$ variabile aleatoare independente.

Pasul 2: Fie $S = V_1^2 + V_2^2$. Dacă $S > 1$, atunci se reia Pasul 1, altfel $T = V_1\sqrt{\dfrac{n(S^{-\frac{2}{n}} - 1)}{S}} \sim Student(n)$.

Sursa: R.W. Bailey, *Polar generation of random variates with the T-distribution.* Mathematics of Computation 62(206), 779781, 1994.

P1. Realizaţi programe care generează N valori aleatoare care urmează distribuţiile prezentate la I,II,III, IV şi V. Estimaţi valorile medii corespunzătoare.

P2. Fie X variabila aleatoare, care indică timpul trecut (exprimat în minute) între două comenzi succesive de pizza la un anumit restaurant italian. Se presupune că X urmează legea exponenţială cu media de 4 minute ($\Longrightarrow$ parametrul este $\frac{1}{4}$). Simulaţi $N(= 100, 1000)$ valori aleatoare pentru X.
a) Care este probabilitatea ca între două comenzi succesive să nu treacă 2 minute?
b) Care este probabilitatea ca niciun client să nu comande pizza într-un interval de 5 minute?
Comparaţi rezultatele obţinute prin simulare cu cele teoretice.

P3. Un anumit motor pentru un vapor de croazieră de ultima generaţie funcţionează cu ajutorul unei pompe, şi două pompe sunt montate de rezervă, astfel: după ce prima pompă s-a defectat se foloseşte cea de-a doua, iar după defectarea celei de-a doua pompe se folosesşte cea de-a treia. Se ştie că o astfel de pompă funcţionează în medie 20 de zile până la defectare şi urmează distribuţia exponenţială ($\lambda = \frac{1}{20}$). Ce distribuţie are variabila aleatoare care indică timpul de funcţionare până la defectarea motorului (adică niciuna din cele 3 pompe nu e funcţională)? Să se simuleze valori pentru această variabilă aleatoare, folosind distribuţia continuă uniformă. Care este probabilitatea ca motorul să funcţioneze mai mult de 60 de zile fără să se defecteze?

P4. Durata de funcţionare a unui anumit dispozitiv este o variabilă aleatoare, care urmează legea normală. În medie un astfel de dispozitiv funcţionează 100 de ore, cu o deviaţie standard de 2 ore.
1) Simulaţi $N(= 100, 1000)$ valori aleatoare. Care este probabilitatea ca un astfel de dispozitiv să funcţioneze mai mult de 99 de ore?
2) $k = 20$ astfel de dispozitive sunt legate a) în serie; b) în paralel. Care este durata medie de funcţionare a unui astfel de sistem?

P5. Presupunem că erorile coordonatelor în spaţiul 3 dimensional atunci când încercăm să localizăm o ţintă, sunt variabile aleatoare normal distribuite cu media 0 şi deviaţia standard 2 (cm). Să se estimeze probabilitatea ca distanţa dintre punctul ales şi ţinta (reală) să nu fie mai mult de 3 (cm). Comparaţi rezultatele simulărilor cu cele teoretice.

Observaţie: Comenzi Octave/Matlab care generează valori pseudo-aleatoare ce urmează o anumită distribuţie: $rand, exprnd, normrnd, randn, gamrnd, randg, chi2rnd, trnd.$