Laboratorul 8 - 2017

Generarea de numere pseudo-aleatoare ce urmează o distribuţie discretă dată (metoda inversei)

Se dau $(x_1, \ldots, x_n)$ (valorile) şi $(p_1, \ldots, p_n)$ (probabilităţile lor). Realizaţi un program care generează N numere pseudo-aleatoare, care urmează *distribuţia discretă*

$$X \sim \begin{pmatrix} x_1 & x_2 & \ldots & x_n \\ p_1 & p_2 & \ldots & p_n \end{pmatrix},$$

folosind numere aleatoare uniform distribuite pe [0,1].

Procedeul de generare al numerelor aleatoare $Y(i)$, $i = \overline{1,N}$, *este*:

• Se citesc valorile $x_1, x_2, \ldots, x_n$ şi probabilităţile corespunzătoare $p_1, p_2, \ldots, p_n$, precum şi numărul N. Fie $p_0 = 0$.

• Se generează N numere aleatoare uniform distribuite pe [0,1]: $U(i)$, $i = \overline{1,N}$.

• Pentru fiecare $i = \overline{1,N}$: $Y(i) = x_k$ dacă şi numai dacă

$$p_0 + p_1 + \cdots + p_{k-1} < U(i) \leq p_0 + p_1 + \cdots + p_k, \ k \in \{1, \ldots, n\}.$$

• Se returnează: $Y(i), i = \overline{1,N}$.

Verificarea procedeului: deoarece U urmează legea uniformă, avem pe baza procedeului de mai sus: $P(\text{“se generează } x_k\text{”}) = P(p_0 + p_1 + \cdots + p_{k-1} < U \leq p_0 + p_1 + \cdots + p_k) = p_k, \ k = 1, \ldots, n$, deci numerele generate urmează legea de distribuţie discretă dată.

P1. Conform statisticilor medicale 46% din oameni au grupa sanguină **0**, 40% au grupa sanguină **A**, 10% au grupa sanguină **B** şi 4% au grupa sanguină **AB**. Simulaţi de $N(= 100, 1000)$ ori stabilirea grupei sanguine a unei persoane alese aleator şi afişaţi frecvenţa de apariţie a fiecărei grupe sanguine. Comparaţi rezultatele obţinute cu cele teoretice.

P2. Un pachet de cărţi de joc conţine 4 aşi. Se extrag fără retrunare 4 cărţi de joc. Fie X variabila aleatoare, care indică câţi aşi au fost extraşi. Folosind

a) *randsample*

b) *hygernd*

c) metoda prezentată mai sus

să se simuleze valori aleatoare ale lui X.

Comparaţi rezultatele obţinute cu cele teoretice.

P3. Se ştie că 5% din produsele unei anumite companii sunt defecte. Produsele sunt testate pe rând până la detectarea primului produs defect. Fie X variabila aleatoare, care indică numărul de produse testate până la detectarea primului produs defect. Folosind metoda prezentată mai sus, să se simuleze valori aleatoare pentru X. Care este probabilitatea de a se verifica mai mult de 5 produse, până la detectarea primului produs defect? $(P(X > 5))$

Comparaţi rezultatele obţinute prin simulare cu cele teoretice.

Indicaţie pentru P3: $X \sim Geo(p)$, $p = 0.05$, $P(X = k) = p(1-p)^k, k \in \{0, 1, 2, ...\}$, fie $U \sim \mathcal{U}[0, 1]$

$$\sum_{j=0}^{k} P(X = k) = \sum_{j=0}^{k} p(1-p)^k = 1 - (1-p)^{k+1}$$

Calcule pentru metoda inversei:

$$X = 0 \ \Leftrightarrow U \leq P(X = 0)$$

$$X = k \ (k \geq 1) \ \Leftrightarrow \ P(X = 0) + ... + P(X = k-1) < U \leq P(X = 0) + ... + P(X = k-1) + P(X = k+1)$$

$$X = k \ (k \geq 1) \ \Leftrightarrow \ 1 - (1-p)^k < U \leq 1 - (1-p)^{k+1}$$

$$X = k \ (k \geq 1) \ \Leftrightarrow k < \frac{ln(1-U)}{ln(1-p)} \leq k+1$$

Simularea de la metoda inversei revine la calculul valorilor $X = \left\lceil \frac{ln(1-U)}{ln(1-p)} \right\rceil - 1$, adică $V \sim \mathcal{U}[0,1] \ \Rightarrow \ X = ceil(ln(V)/ln(1-p)) - 1 \sim Geo(p)$ pentru că $V \sim \mathcal{U}[0,1] \Leftrightarrow 1 - V \sim \mathcal{U}[0,1]$. În Octave/Matlab *log* corespunde logaritmului natural.

P4. Realizaţi un program care generează N valori aleatoare care urmează *legea Poisson* cu parametrul $\lambda > 0$, folosind metoda inversei. Estimaţi pe baza simulărilor valoarea medie pentru această variabilă aleatoare.

$$X \sim \begin{pmatrix} j \\ \frac{\lambda^j}{j!} e^{-\lambda} \end{pmatrix}_{j=0,1,2,...} \Rightarrow E(X) = \sum_{j=0}^{\infty} j P(X = j) = \lambda$$

I. Metoda inversei: Fie $U \sim \mathcal{U}[0,1]$, $P_0 = e^{-\lambda}$, $P_{j+1} = P_j \cdot \frac{\lambda}{j+1}$

$$X = k \ (k = 0, 1, 2...) \Leftrightarrow k \text{ este cel mai mic indice a.î. } U \leq \sum_{j=0}^{k} P_j$$

II. Metoda proceselor Poisson: Fie $U_1, U_2, ... \sim \mathcal{U}[0,1]$

$$X = k - 1 \ (k = 1, 2...) \Leftrightarrow k \text{ este cel mai mic indice a.î. } U_1 \cdot U_2 \cdot \ldots \cdot U_k < e^{-\lambda}.$$

Aplicaţie: Numărul de clienţi care sună la o centrală telefonică în decurs de o oră urmează distribuţia Poisson cu media de 20 clienţi pe oră ($\Rightarrow \lambda = 20$). Simulaţi numărul de apeluri în centrală şi afişaţi frecvenţa de apariţie a exact k clienţi, unde $k \in \{5, 10, 15\}$. Comparaţi rezultatele obţinute prin simulare cu cele teoretice.

P5. Realizaţi un program care generează N valori aleatoare care urmează *legea uniformă discretă* $Unif(n)$ ($n \in \mathbb{N}^*$ parametru dat), folosind metoda inversei.

$$X \sim Unif(n) \ \Leftrightarrow \ X \sim \begin{pmatrix} 1 & 2 & \ldots & n \\ \frac{1}{n} & \frac{1}{n} & \ldots & \frac{1}{n} \end{pmatrix}$$

Fie $U \sim \mathcal{U}[0,1]$

$$X = k \ (k = 1, 2..., n) \Leftrightarrow k \text{ este cel mai mic indice a.î. } U \leq \sum_{j=0}^{k} \frac{1}{n} = \frac{k}{n}$$

$$X = k \ (k = 1, 2..., n) \Leftrightarrow k - 1 < nU \leq k$$

Simularea de la metoda inversei revine la calculul valorilor $X = \left\lceil nU \right\rceil$,
adică $U \sim \mathcal{U}[0,1] \;\Rightarrow\; X = ceil(nU) \sim Unif(n)$.
Aplicaţie: Se ia o variabilă aleatoare $Y \sim \mathcal{U}[0,1]$. Ce distribuţie are variabila aleatoare $\lfloor 10 \cdot Y \rfloor$ ($\lfloor x \rfloor = floor(x)$ partea întreagă a numărului pozitiv x)?

P6. Realizaţi un program care generează N valori aleatoare care urmează *legea binomială* $Binomial(n,p)$, folosind metoda inversei.

$$X \sim Binomial(n,p) \;\Leftrightarrow\; X \sim \begin{pmatrix} 0 & 1 & \ldots & k & \ldots & n \\ (1-p)^n & np(1-p)^{n-1} & \ldots & C_n^k p^k (1-p)^{n-k} & \ldots & p^n \end{pmatrix}.$$

Aplicaţie: Într-un test de tip grilă (multiple-choice) sunt formulate 10 întrebări; la fiecare întrebare sunt date trei răspunsuri posibile, dintre care doar unul este corect. Un elev nu s-a pregătit pentru test şi s-a decis că va răspunde aleator la întrebări. Cu ce probabilitate va răspunde corect la
a) exact 6 întrebări?
b) cel puţin 6 întrebări?