II Cazul v.a. discrete

Fie X o v.a. discretă cu repartiția:

$$X : \begin{pmatrix} x_1 & x_2 & x_3 & \dots \\ p_1 & p_2 & p_3 & \dots \end{pmatrix} \qquad p_i > 0 \ \forall i, \quad \sum_{i=1}^{\infty} p_i = 1$$

Pentru a simula valori din v.a. X folosind metoda inversă procedăm astfel:

$$\boxed{\begin{array}{l} 1)\ \text{Generăm } U \sim Unif(0,1) \\ 2)\ X = \begin{cases} x_1, & \text{dacă } U < p_1 \\ x_2, & \text{dacă } p_1 \le U < p_1 + p_2 \\ \dots & \dots \\ x_j, & \text{dacă } \sum_{i=1}^{j-1} p_i \le U < \sum_{i=1}^{j} p_i \end{cases} \end{array}} \quad (*)$$

Exemplu

$$X : \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ \frac{1}{3} & \frac{1}{30} & \frac{2}{15} & \frac{7}{30} & \frac{4}{15} \end{pmatrix}$$

- Generez $U_1 = 0.015$

  Cum $0.015 < \frac{1}{3} \Rightarrow \underline{\underline{X_1 = 1}}$

- Generez $U_2 = 0.45$

  Cum $0.45 \not< \frac{1}{3} \Rightarrow X_2 \neq 1$

  Cum $0.45 \not< \frac{1}{3} + \frac{1}{30} \Rightarrow X_2 \neq 2$

  Cum $0.45 < \frac{1}{3} + \frac{1}{30} + \frac{2}{15} \Rightarrow \underline{\underline{X_2 = 3}}$

$\boxed{\text{OBS}}$: O eficientizare a procedeului s-ar produce dacă am reordona valorile lui X a.î. probabilitățile să fie descrescatoare. De ce?

demonstrație

$$\mathbb{P}(X = x_j) = \mathbb{P}\left(\sum_{i=1}^{j-1} p_i \le U < \sum_{i=1}^{j} p_i\right) =$$

$$\overset{(1)}{=} F_U\Big(\underbrace{\sum_{i=1}^{j} p_i}_{\ge 0}\Big) - F_U\Big(\underbrace{\sum_{i=1}^{j-1} p_i}_{\ge 0}\Big)$$

$$\overset{(2)}{=} \sum_{i=1}^{j} p_i - \sum_{i=1}^{j-1} p_i = p_j,$$ ceea ce arată că X are repartiția dorită

Reamintim:

(1) $F \to$ funcție de repartiție

$$\boxed{\mathbb{P}(a < X \le b) = F(b) - F(a)}$$

$\boxed{\text{OBS}}$: În cazul v.a. continue

$$\mathbb{P}(a \le X \le b) = \underline{\mathbb{P}(a \le X < b)} =$$

$$= \underline{\mathbb{P}(a < X \le b)} = \mathbb{P}(a < X < b)$$

(2) $U \sim Unif(a,b)$

$$F_U(x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & a \le x < b \\ 1, & x \ge b \end{cases}$$

Pentru $U \sim Unif(0,1)$

$$F_U(x) = \begin{cases} 0, & x < 0 \\ x, & 0 \le x < 1 \\ 1, & x \ge 1 \end{cases}$$

Comentariu

Dacă F este funcția de repartiție a v.a. X atunci algoritmul (*) devine:

$$\boxed{X = x_j \text{ pentru } F(x_{j-1}) \leq U < F(x_j)}$$

(presupunem că valorile v.a. sunt ordonate crescător)

Relația de mai sus arată că algoritmul (*) se reduce la a identifica intervalul $[F(x_{j-1}), F(x_j))$ în care se găsește U, ceea ce este echivalent cu a inversa F.

aici se folosește noțiunea de inversă generalizată pentru că F nu e bijectivă

Caz particular

Dacă $X \sim Unif(\{1, 2, \dots n\})$ atunci: (uniforma pe caz discret)

$X = j$ pentru $\frac{j-1}{n} \leq U < \frac{j}{n}$, ceea ce implică $\boxed{X = [n \cdot U] + 1}$ parte întreagă

! Temă: Justificați relația de mai sus, apoi creați o funcție în R care implementează relația de mai sus. Generați 10^6 valori din v.a. X (alegeți voi un n particular, $n \geq 10$) și faceți histograma comparativă a acestor valori cu cele generate de funcția sample din R.

Aplicarea metodei inverse în cazul unor repartiții de v.a. discrete ce iau un număr infinit de valori

① Repartiția geometrică $X \sim Geom(p)$

$$X : \begin{pmatrix} 0 & 1 & 2 & \dots & j & \dots \\ p & p\cdot q & p\cdot q^2 & \dots & p\cdot q^{j} & \dots \end{pmatrix}$$ → dacă număr câte eșecuri am până la primul succes

$\boxed{q = 1-p}$

sau

⊙ $$X : \begin{pmatrix} 1 & 2 & 3 & \dots & j & \dots \\ p & p\cdot q & p\cdot q^2 & \dots & p\cdot q^{j-1} & \dots \end{pmatrix}$$ → dacă număr câte încercări am până la primul succes

! Vom folosi în continuare reprezentarea ⊙.

$\mathbb{P}(X=j) = p\cdot q^{j-1} \quad \forall j \geq 1$

probabilitatea evenimentului complementar

Cum $\sum_{i=1}^{j-1} \mathbb{P}(X=i) = 1 - \mathbb{P}(X > j-1)$

$= 1 - \mathbb{P}(\text{„primele } j-1 \text{ încercări sunt eșecuri"})$

$= 1 - q^{j-1} \quad , \forall j \geq 1$

Deci, folosind metoda inversă $X = j$ pentru $\boxed{1-q^{j-1} \leq U < 1-q^{j}}$, ceea ce este echivalent cu $q^{j} < 1-U \leq q^{j-1}$

Putem reformula astfel:

$$X = Min\{\, j \mid q^{j} < 1-U \,\}$$

$q^{j} < 1-U \Leftrightarrow \ln q^{j} < \ln(1-U) \Leftrightarrow$

$j \cdot \underbrace{\ln q}_{<0} < \ln(1-U) \Leftrightarrow j > \dfrac{\ln(1-U)}{\ln q}$

Deci, $X = Min\left\{ j \mid j > \dfrac{\ln(1-U)}{\ln q} \right\}$,

ceea ce se poate rescrie:

$$\boxed{X = \left[\frac{\log(1-U)}{\log(q)} \right] + 1}$$

parte întreagă

② Repartiția Poisson $X \sim Pois(\lambda)$

$$X : \begin{pmatrix} 0 & 1 & 2 & \dots & j & \dots \\ \dots & & & & \frac{e^{-\lambda} \cdot \lambda^j}{j!} & \dots \end{pmatrix}, \ j \geq 0$$

Ideea centrală în folosirea metodei inverse este legată de folosirea următoarei relații de recurență:

$$\boxed{p_{j+1} = \frac{\lambda}{j+1} \cdot p_j, \ j \geq 0}$$

TEMĂ: de demonstrat

Algoritm

Pas 1: Generez U ($\sim Unif(0,1)$)

contor, prob. curentă, valoarea funcției de repartiție în punctul curent

Pas 2: Inițializez $\underline{i=0}$, $\underline{p = e^{-\lambda}}$, $F = p$

Pas 3: Dacă $U < F$ atunci $X = i$ $\boxed{STOP}$

Pas 4: $p = \frac{\lambda p}{i+1}$, $F = F+p$, $i = i+1$

Pas 5: Merg la Pas 3.

③ Repartiția binomială $X \sim Binom(n, p)$

$$X : \begin{pmatrix} 0 & 1 & \dots & k & \dots & n \\ \dots & & & C_n^k p^k q^{n-k} & \dots & \end{pmatrix} \quad \begin{matrix} k \geq 0 \\ k = \overline{0,n} \end{matrix}$$

Similar, ideea centrală în folosirea metodei inverse este legată de o relație de recurență:

$$\boxed{p_{j+1} = \frac{n-j}{j+1} \cdot \frac{p}{1-p} \cdot p_j}$$

TEMĂ: de demonstrat

Algoritm

Pas 1: Generez U ($\sim Unif(0,1)$)

Pas 2: Inițializez $c = \frac{p}{1-p}$, $i=0$, $prob = (1-p)^n$, $F = prob$

Pas 3: Dacă $U < F$ atunci $X = i$ $\boxed{STOP}$

Pas 4: $prob = \frac{c(n-i)}{i+1} \cdot prob$, $F = F + prob$, $i = i+$

Pas 5: Merg la Pas 3.