Generarea variabilelor neuniforme

Curs 5

Simularea unor variabile particulare

În continuare se vor prezenta metode de generare a anumitor tipuri de variabile aleatoare. Aceste metode se bazează fie pe teoremele de simulare, fie pe proprietăţi ale acestor tipuri de variabile aleatoare.

Variabila Exponenţială

Fie X o variabilă exponenţială de parametru 1 cu densitatea de repartiţie

$$f(x) = \begin{cases} e^{-x}, \text{ dacă } x \geq 0 \\ 0, \text{ dacă } x < 0. \end{cases} \tag{1}$$

Fie Y o variabilă exponenţială de parametru λ cu densitatea de repartiţie

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, \text{ dacă } x \geq 0 \\ 0, \text{ dacă } x < 0. \end{cases} \tag{2}$$

Atunci are loc relaţia:

$$Y = \frac{X}{\lambda}$$

Prin urmare pentru a simula Y este suficient să descriem o metodă de simulare pentru X.

Metode de simulare pentru $X \sim Exp(1)$

- Metoda inversă:

$$X = -\log(U)$$

unde U este o variabilă uniformă pe $[0, 1]$.
Are dezavantajul că pentru valori ale lui U apropiate de 0 nu se poate calcula logaritmul.

- Metodă de generare cu a treia teoremă de respingere

Teorema 1. *În teorema subşirului descendent considerăm* $Z_0 = U_0$, $Z_i = U_i, i \geq 1$, *unde* $U_0, U_1, ...$ *sunt uniforme pe* $[0, 1]$. *Dacă notăm cu* N *numărul aleator de subşiruri descendente respinse până când se acceptă un subşir, atunci* $X = N + Z_0$ *este o variabilă Exp(1), unde* Z_0 *este variabila acceptătă (din ultimul subşir descendent).*

Dem:
Din exemplul de la a treia teoremă de respingere, pentru $x \in [0,1]$ avem:

$$P(Z_0 \leq x | K = nr.impar) = \frac{1}{p_a} \int_0^x e^{-x} dx = F(x) = \frac{1 - e^{-x}}{p_a}$$

cu

$$p_a = 1 - e^{-1}.$$

unde $F(x)$ este funcţia de repartiţie pentru variabila exponenţială trunchiată pe [0,1].
Deci probabilitatea de a respinge un şir descendent (de forma $Z_0 \geq Z_1 \geq ... \geq Z_{K-1} < Z_K$) este $p_r = 1 - p_a = e^{-1}$. Prin urmare

$$P(N = n) = e^{-n}(1 - e^{-1}).$$

Trebuie să mai arătăm că:

$$P(N + Z_0 \leq x) = \begin{cases} 0 \text{ dacă } x < 0; \\ 1 - e^{-x}, \text{ dacă } x \geq 0 \end{cases}$$

Fie $x > 0$ oarecare. Notăm cu $k = [x]$ şi cu $z = x - k$, $z \in [0, 1)$. Atunci avem:

$$P(N + Z_0 \leq x) = P(N + Z_0 \leq k + z) = P(N < k) + P(N = k, Z_0 \leq z) =$$

$$= \sum_{j=0}^{k-1} (1 - e^{-1}) e^{-j} + (1 - e^{-1}) \frac{e^{-k}}{1 - e^{-1}} \int_0^z e^{-u} du =$$

$$= 1 - e^{-k} + e^{-k}(1 - e^{-z}) = 1 - e^{-(k+z)} = 1 - e^{-x}$$

Algoritm Resp3-Exp
Intrare:
P1: $N = 0$;
P2: Se generează $U_0, U_1 \sim U(0,1)$ independente;
P3: $U^* = U_0$, $K = 1$;
P4: Dacă $U_0 \geq U_1$ mergi la P5, altfel mergi la P7;
P5: $K := K + 1$, $U_0 := U_1$;
P6: Se generează $U_1 \sim U(0,1)$, mergi la P4;
P7: Dacă $K \mod 2 = 1$ $X = N + U^*$, STOP. Altfel
$N := N + 1$, mergi la pasul 2.
Ieşire: Variabila aleatoare X.

Pe lângă probabilitatea de acceptare $p_a = 1 - e^{-1}$, performanţa algoritmului este caracterizată şi de numărul N^* al variabilelor uniforme $\{U_i\}_{i \geq 0}$ generate, necesare pentru a obţine o valoare de selecţie exponenţială X. Din evaluarea performanţelor algoritmului celei de-a treia metode de respingere:

$$E[N^*] = \frac{1}{p_a}\left(1 + \int_0^1 e^x dx\right) = \frac{1}{1 - e^{-1}}(1 - e^{-1}) = \frac{e^2}{e - 1} \approx 3,8$$

Variabila Gama

Fie Y o variabilă aleatoare distribuită $Gama(\alpha, \lambda, \nu)$. Atunci Y are următoarea densitate de repartiţie:

$$f(x) = \begin{cases} \frac{\lambda^\nu}{\Gamma(\nu)}(x-\alpha)^{\nu-1}e^{-\lambda(x-\alpha)}, \text{ dacă } x \geq \alpha \\ 0, \text{ altfel} \end{cases} . \tag{3}$$

Cu funcţia $\Gamma(\nu)$ definită astfel:

$$\Gamma(\nu) = \int_0^{+\infty} x^{\nu-1}e^{-x}dx.$$

Fie X o variabilă aleatoare $Gama(0, 1, \nu)$. Atunci X se mai numeşte variabilă *Gama standard* şi are densitatea de repartiţie:

$$f(x) = \begin{cases} \frac{1}{\Gamma(\nu)}x^{\nu-1}e^{-x}, \text{ dacă } x \geq 0 \\ 0, \text{ altfel} \end{cases} . \tag{4}$$

Atunci Y se poate scrie

$$Y = \alpha + \frac{X}{\lambda}$$

Prin urmare pentru a genera variabila Y este suficientă găsirea unei metode de generare a variabilei X.

Algoritmii de simulare pentru variabila X depind de intervalul în care se află parametrul ν.

1. Pentru $0 < \nu < 1$

1.1. Generarea variabilei X prin metoda respingerii cu ajutorul unei variabile Weibull (exemplu în **cursul 3**).

1.2. Generarea variabilei X printr-o metodă de compunere-respingere.

Vom scrie densitatea (4) sub forma:

$$f(x) = p_1 f_1(x) + p_2 f_2(x)$$

cu

$$f_1(x) = \begin{cases} \frac{f(x)}{p_1}, \text{ dacă } x \in [0,1] \\ 0, \text{ altfel} \end{cases} \quad , \quad f_2(x) = \begin{cases} \frac{f(x)}{p_2}, \text{ dacă } x \in (1,+\infty) \\ 0, \text{ altfel} \end{cases}$$

unde

$$p_1 = \frac{\Gamma(1;\nu)}{\Gamma(\nu)}, \quad p_2 = 1 - p_1 = \frac{\Gamma(\nu) - \Gamma(1;\nu)}{\Gamma(\nu)}.$$

Funcţia $\Gamma(1;\nu)$ este o funcţie gama incompletă:

$$\Gamma(1;\nu) = \int_0^1 x^{\nu-1} e^{-x} dx$$

Presupunem că variabila X_1 are densitatea de repartiţie f_1 şi că variabila X_2 are densitatea de repartiţie f_2. Atunci are loc următoarea teoremă.

Teorema 2. *Variabila* X_1 *se simulează folosind a treia teoremă de respingere (a subşirului descendent) cu* $Z_0 = U_0^{1/\nu}$, $Z_i = U_i$, *cu* $\{U_i\}_{i\geq 0}$ *variabile uniforme pe* $[0,1]$. *Variabila* X_2 *se simulează cu a doua teoremă de respingere, forma duală, unde densitatea* $f_2(x)$ *este de forma*

$$f_2(x) = c(1 - Q(x))r(x), \quad x > 0$$

cu

$$c = \frac{1}{e(\Gamma(\nu) - \Gamma(1;\nu))}$$

şi

$$r(x) = \begin{cases} e^{-x+1}, \ \textit{dacă}\ x \geq 1 \\ 0\ \textit{dacă}\ x < 1 \end{cases}, \quad Q(x) = \begin{cases} 1 - x^{\nu-1}, \ \textit{dacă}\ x \geq 1 \\ 0\ \textit{dacă}\ x < 1 \end{cases}$$

Dem:
Fie $x \in [0,1]$. Atunci aplicând a treia teoremă de respingere avem:

$$G_0(x) = P(U_0^{1/\nu} < x) = P(U_0 < x^\nu) = x^\nu$$

$$H(x) = P(U_0 < x | K = nr.impar) = \frac{1}{p_a}\int_0^x e^{-t} dG_0(t) = \frac{1}{p_a}\int_0^x e^{-t}\nu t^{\nu-1}dt$$

cu

$$p_a = \int_0^1 e^{-t}\nu t^{\nu-1}dt = \nu\Gamma(1;\nu).$$

Dacă derivăm $H(x)$ obţinem:

$$H'(x) = h(x) = \frac{e^{-x}x^{\nu-1}}{\Gamma(1;\nu)} = f_1(x), \quad x \in [0,1]$$

ceea ce demonstrează prima parte a teoremei.

Pentru a demonstra a doua parte îl scriem pe $f_2(x)$ astfel:

$$f_2(x) = \begin{cases} \frac{x^{\nu-1}e^{-x}}{\Gamma(\nu)-\Gamma(1;\nu)}, \text{ dacă } x \geq 1 \\ 0, \text{ dacă } x < 1 \end{cases}$$

adică $f_2(x)$ are forma din enunţul teoremei.

Din această teoremă rezultă următorul algoritm de generare a variabilei X.

Algoritm Gama2

Intrare: p_1, p_2 $g = \Gamma(\nu)$, $g_1 = \Gamma(1;\nu)$, $p_1 = \frac{g_1}{g}$, $p_2 = \frac{g-g_1}{g}$, $c = \frac{1}{e(\Gamma(\nu)-\Gamma(1;\nu))}$, $a = \frac{1}{\nu}$, $b = -\frac{1}{1-\nu}$

P1: Se generează $U \sim U(0,1)$;

P2: Dacă $U \leq p_1$ mergi la P3, altfel mergi la P4;

P3: Se generează $X_1 \sim f_1(x)$, $X := X_1$.

P4: Se generează $X_2 \sim f_2(x)$, $X := X_2$.

Ieşire: Variabila aleatoare X.

Variabilele X_1 şi X_2 se generează cu următorii algoritmi:

Algoritm Gama2-X_1
Intrare:
P1: Se generează $U \sim U(0,1)$;
P2: $Z_0 := U^a$;
P3: Se generează $Z_1 \sim U(0,1)$;
P4: $K := 1$, $Z^* := Z_0$;
P5: Dacă $Z_0 \geq Z_1$ mergi la P6, altfel, mergi la P7;
P6: $Z_0 := Z_1$, se generează $Z_1 \sim U(0,1)$, $K := K+1$, mergi la P5;
P7: Dacă $K \mod 2 = 1$ $X_1 = Z^*$. Altfel, mergi la P1.
Ieşire: Variabila aleatoare X_1.

Pentru algoritmul Gama2-X_1 avem:

$$p_a = \nu\Gamma(1;\nu); \quad E(N_1^*) = \frac{1}{p_a}\left(1 + \int_0^1 \nu t^{\nu-1} e^t dt\right)$$

Algoritm Gama2-X_2
Intrare:
`P1:` Se generează $U \sim U(0,1)$;
`P2:` $Z := U^b$;
`P3:` Se generează $X_0 \sim Exp(1)$;
`P4:` $Y := X_0 + 1$ `P5:` Dacă $Y > Z$ mergi la P1, altfel mergi la P6;
`P6:` $X_2 := Y$;
Ieşire: Variabila aleatoare X_2.

Pentru algoritmul Gama-X_2 avem:

$$p_2 = 1 - p_1; \quad E(N_2^*) = \frac{2}{p_2}$$

şi prin urmare numărul mediu de variabile necesare pentru a genera în final un X este:

$$E(N*) = p_1 E(N_1^*) + p_2 E(N_2^*) = p_1 E(N_1^*) + 2$$

2. Pentru $\nu > 1$

Doi algoritmi de respingere bazaţi pe metoda înfăşurătoarei (prima teoremă de respingere).

2.1 Primul algoritm de respingere pentru variabila $X \sim Gama(0, 1, \nu)$ **cu** $\nu > 1$**.**

Considerăm ca înfăşurătoare densitatea $h(x) \sim Exp\left(\frac{1}{\nu}\right)$, adică:

$$h(x) = \begin{cases} \frac{1}{\nu} e^{-\frac{x}{\nu}} \text{ dacă } x \geq 0 \\ 0 \text{ dacă } x < 0 \end{cases} . \quad (5)$$

Notăm cu $r(x)$ raportul:

$$r(x) = \frac{f(x)}{h(x)} = \frac{\nu x^{\nu-1} e^{-x}}{\Gamma(\nu) e^{-\frac{x}{\nu}}},\ \nu > 1$$

Atunci funcţia $r(x)$ are un punct de maxim pentru $x = \nu$ şi constanta α din teorema înfăşurătoarei este:

$$\alpha = r(\nu) = \frac{\nu^{\nu} e^{1-\nu}}{\Gamma(\nu)}$$

Algoritmul de respingere va avea probabilitatea de acceptare:

$$p_a = \frac{\Gamma(\nu)}{\nu^{\nu} e^{1-\nu}} \approx \sqrt{\frac{e^2 2\pi}{\sqrt{\nu - 1}}}$$

ultima relaţie rezultând din aproximarea lui Stirling pentru $\nu \to \infty$:

$$\Gamma(\nu) \approx (\nu - 1)^{\nu - 1} e^{-(\nu - 1)} \sqrt{2\pi(\nu - 1)}$$

2.2 Al doilea algoritm de respingere pentru variabila $X \sim Gama(0, 1, \nu)$ cu $\nu > 1$.

Algoritmul de mai sus este lent pentru ν foarte mare, de aceea vom prezenta un alt algoritm de respingere, de data aceasta bazat pe o înfăşurătoare dată de o densitate Cauchy nestandard trunchiată pe $[0, \infty)$:

$$h(x) = \frac{k}{1 + \frac{(x-(\nu-1))^2}{c}}, \quad x \geq 0 \tag{6}$$

unde k este o constantă de normare. Atunci are loc următoarea teoremă:

Teorema 3. *Dacă se înfăşoară densitatea* $Gama(0, 1, \nu)$, $\nu > 1$ *cu densitatea* $h(x)$ *dată de (6), atunci pentru* $c \geq 2\nu - 1$ *avem:*

$$r(x) = \frac{f(x)}{h(x)} \leq \alpha = \frac{1}{k\Gamma(\nu)}(\nu - 1)^{\nu-1}e^{-(\nu-1)} \tag{7}$$

Dem:
Avem

$$r(x) = \frac{1}{k\Gamma(\nu)}\varphi(x)$$

unde

$$\varphi(x) = x^{\nu-1}e^{-x}\left[1 + \frac{(x-(\nu-1))^2}{c}\right].$$

Atunci:

$$\varphi'(x) = -\frac{e^{-x}x^{\nu-1}}{c}[x-(\nu-1)][(x-\nu)^2 + c - (2\nu-1)]$$

de unde rezultă că ecuaţia $\varphi'(x) = 0$ are soluţia $x_0 = \nu - 1 > 0$ iar dacă $c \geq 2\nu - 1$ atunci x_0 este punct de maxim. Dacă luăm $c = 2\nu - 1$ atunci avem:

$$\alpha = \frac{1}{k\Gamma(\nu)}(\nu-1)^{\nu-1}e^{-(\nu-1)}$$

Din această teoremă rezultă următorul algoritm de generare a unei variabile $Gama(0, 1, \nu)$ cu $\nu > 1$:

Algoritm Gama3

Intrare: ν, $b = \nu - 1$, $c = \nu + b$, $s = \sqrt{2\nu - 1}$

P1: Se generează $U \sim U(0, 1)$, $T := s \cdot tg[\pi(U - 0.5)]$ (T este o variabilă Cauchy standard generată cu metoda inversă);

P2: $Y = b + T$ (Y este o variabilă Cauchy nestandard);

P3: Dacă $Y > 0$ mergi la P4, altfel mergi la P1;

P4: Se generează $U_1 \sim U(0, 1)$;

P5: Dacă $U_1 \leq e^{b \log(Y/b) - T + \log(1+T^2/c)}$ mergi la P6, altfel mergi la P1;

P6: $X = Y$.

Ieşire: Variabila aleatoare X.

Constanta de normare k nu intervine în construcţia algoritmului, dar ea este necesară pentru a calcula probabilitatea de acceptare p_a. Se verifică uşor că:

$$k = \left[\frac{\pi}{2} + arctg \left(-\frac{\nu - 1}{\sqrt{2\nu - 1}} \right) \right]^{-1}$$

Un **alt algoritm de generare** pentru o variabilă $Gama(0, 1, \nu)$ cu $\nu > 1$ se poate obţine folosind următorul raţionament: fie

$$\nu = k + p$$

unde $k = [\nu]$ este partea întreagă şi $p = \nu - k \in [0, 1)$. Atunci

$$X = E_k + Y$$

unde E_k este variabila Erlang de parametru k, Y este o variabilă $Gama(0, 1, \nu)$ cu $\nu < 1$, iar E_k şi Y sunt independente.