

# Generarea variabilelor neuniforme

## Curs 6

December 4, 2023

# Repartiția Beta

Fie  $X$  o variabilă aleatoare cu densitatea de repartiție:

$$f(x) = \begin{cases} \frac{1}{B(a,b)} x^{a-1} (1-x)^{b-1}, & \text{dacă } x \in [0, 1] \\ 0, & \text{altfel} \end{cases} \quad (1)$$

unde

$$B(a, b) = \int_0^1 x^{a-1} (1-x)^{b-1} dx, \quad B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

este funcția beta.

Variabila  $X$  are o distribuție  $Beta(a, b)$ . Media și dispersia variabilei  $X$  sunt:

$$E[X] = \frac{a}{a+b}, \quad Var[X] = \frac{ab}{(a+b)^2(a+b+1)}. \quad (2)$$

Variabila  $X$  se poate genera folosind următoarea teoremă:

### Teoremă

*Dacă  $X_1 \sim \text{Gama}(0, 1, a)$ ,  $X_2 \sim \text{Gama}(0, 1, b)$  sunt două variabile independente, atunci variabila:*

$$X = \frac{X_1}{X_1 + X_2} \quad (3)$$

*este o variabilă Beta( $a, b$ ).*

### Dem:

Densitatea comună de repartiție a variabilelor  $X_1, X_2$  este:

$$f(x_1, x_2) = \frac{1}{\Gamma(a)\Gamma(b)} x_1^{a-1} x_2^{b-1} e^{-(x_1+x_2)}.$$

Făcând transformarea de variabile:

$$U = \frac{X_1}{X_1 + X_2}, \quad V = X_2,$$

densitatea comună a variabilelor  $U$  și  $V$  este

$$g(u, v) = f(x_1(u, v), x_2(u, v)) \cdot J$$

cu

$$J = \det \left( \frac{\partial(x_1, x_2)}{\partial u \partial v} \right) = \frac{v}{(1-u)^2}, \quad 0 < u < 1.$$

Avem

$$g(u, v) = \frac{1}{\Gamma(a)\Gamma(b)} \frac{u^{a-1} v^{a+b-1}}{(1-u)^{a+1}} e^{-\frac{v}{1-u}}$$

cu  $0 < v < \infty$ . Densitatea de repartiție a variabilei  $U$  este

$$h(u) = \int_0^\infty g(u, v) dv = \frac{1}{\Gamma(a)\Gamma(b)} \int_0^\infty \frac{u^{a-1} v^{a+b-1}}{(1-u)^{a+1}} e^{-\frac{v}{1-u}} dv$$

adică

$$h(u) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} u^{a-1} (1-u)^{b-1},$$

ceea ce demonstrează teorema.

Prin urmare, un algoritm de generare a unei variabile *Beta* este dat de relația (3). Acest algoritm ar presupune generarea a două variabile *Gama* și acest lucru implică o complexitate mare. De aceea în cazuri particulare se aplică algoritmi rezultați din următoarele teoreme.

### Teoremă

*Fie  $a, b \in \mathbb{N}_+$ ,  $n = a + b - 1$  și fie  $U_1, U_2, \dots, U_n$  variabile aleatoare uniforme pe  $[0, 1]$ , independente. Fie  $U_{(1)} < U_{(2)} < \dots < U_{(n)}$  statisticile de ordine obținute prin ordonarea valorilor  $U_1, U_2, \dots, U_n$ . Atunci  $U_{(a)} \sim \text{Beta}(a, b)$ .*

### Teoremă

*Fie  $0 < a < 1$ ,  $0 < b < 1$  și  $U_1, U_2$  variabile aleatoare uniforme pe  $[0, 1]$  independente. Dacă  $V = U_1^{\frac{1}{a}}$ ,  $T = U_2^{\frac{1}{b}}$ , atunci repartiția variabilei  $X = \frac{V}{V+T}$  condiționată de  $V + T < 1$  este  $\text{Beta}(a, b)$ .*

Algoritmul rezultat este următorul:

### Algoritm Beta3

**Intrare:**  $0 < a, b < 1$

P1: Se generează  $U_1, U_2 \sim U(0, 1)$  independente;

P2:  $V = U_1^{\frac{1}{a}}, T = U_2^{\frac{1}{b}};$

P3: Dacă  $V + T < 1$  mergi la P4, altfel mergi la P1;

P4:  $X := \frac{V}{V+T};$

**Ieșire:** Variabila aleatoare  $X$ .

Probabilitatea de acceptare a acestui algoritm de respingere este:

$$p_a = P(V + T < 1) = \frac{ab}{a+b} B(a, b).$$

## Teoremă

Fie  $0 < a < 1$ ,  $b > 1$  și  $U_1, U_2$  variabile aleatoare uniforme pe  $[0, 1]$  independente. Dacă  $V = U_1^{\frac{1}{a}}$ ,  $T = U_2^{\frac{1}{b-1}}$ , atunci repartiția variabilei  $V$  condiționată de  $V + T < 1$  este Beta( $a, b$ ).

### Dem:

Observăm că pentru  $x \in [0, 1]$  avem:

$$F(x) = P(V < x) = P(U_1^{\frac{1}{a}} < x) = P(U < x^a) = x^a.$$

De unde rezultă că densitatea de repartiție a lui  $V$  este

$$f(x) = ax^{a-1}, \quad x \in [0, 1].$$

Asemănător, densitatea de repartiție a lui  $T$  este:

$$h(y) = (b-1)y^{b-2}, \quad y \in [0, 1].$$

Prin urmare, densitatea comună de repartiție a variabilelor  $V, T$ , independente este:

$$g(x, y) = a(b-1)x^{a-1}y^{b-2}.$$

De unde rezultă că:

$$P(V + T < 1) = a(b-1) \int_0^1 \left( \int_0^{1-x} y^{b-2} dy \right) x^{a-1} dx = aB(a, b).$$

Deci densitatea comună a variabilelor  $V$ ,  $T$  condiționată de  $V + T < 1$  este:

$$p(x, y) = \frac{b-1}{B(a, b)} x^{a-1} y^{b-2}, \quad x \in [0, 1], y \in [0, 1].$$

Atunci densitatea lui  $V$  condiționată de  $V + T < 1$  este

$$q(x) = \int_0^{1-x} p(x, y) dy = \frac{1}{B(a, b)} x^{a-1} (1-x)^{b-1},$$

ceea ce demonstrează teorema.

Algoritmul de generare rezultat este asemănător cu algoritmul de generare rezultat din teorema precedentă și are probabilitatea de acceptare:

$$p_a = P(V + T < 1) = aB(a, b).$$



## Repartiția normală

Fie  $X$  o variabilă normală  $N(0, 1)$ . Atunci  $X$  are densitatea de repartiție:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}, \quad x \in \mathbb{R}.$$

Dacă  $Y$  este o variabilă normală  $N(m, \sigma)$ , atunci are densitatea de repartiție:

$$g(y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(y-m)^2}, \quad x \in \mathbb{R}.$$

Media și dispersia variabilei  $Y$  sunt:

$$E[Y] = m, \quad \text{Var}[Y] = \sigma^2. \quad (4)$$

Relația dintre  $Y$  și  $X$  este:

$$Y = m + \sigma X.$$

Prin urmare, pentru a-l genera pe  $Y$  este suficient să știm să-l generăm pe  $X$ .

**1. Metoda bazată pe teorema limită centrală**  
(a fost prezentată).

**2. O metodă de compunere-respingere**

Fie  $X_1$  variabila aleatoare cu densitatea:

$$f_1(x) = \begin{cases} \sqrt{\frac{2}{\pi}} e^{-\frac{x^2}{2}}, & \text{dacă } x \geq 0 \\ 0, & \text{dacă } x < 0 \end{cases}$$

Fie  $X_2 = -X_1$ , atunci densitatea de repartiție a variabilei  $X_2$  este:

$$f_2(x) = \begin{cases} \sqrt{\frac{2}{\pi}} e^{-\frac{x^2}{2}}, & \text{dacă } x < 0 \\ 0, & \text{dacă } x \geq 0 \end{cases}$$

Prin urmare densitatea variabilei aleatoare  $X \sim N(0, 1)$  se poate scrie:

$$f(x) = \frac{1}{2}f_1(x) + \frac{1}{2}f_2(x)$$

adică  $f(x)$  este o compunere discretă a densităților  $f_1(x)$  și  $f_2(x)$ .

Pentru generarea variabilei aleatoare  $X_1$  folosim următoarea teoremă:

### Teoremă

*Fie  $h(x)$  densitatea de repartiție a unei variabile aleatoare  $\text{Exp}(1)$ . Atunci dacă înfășurăm  $f_1(x)$  cu  $h(x)$  avem*

$$\frac{f_1(x)}{h(x)} \leq \alpha = \sqrt{\frac{2e}{\pi}}.$$

**Dem:**

Observăm că:

$$r(x) = \frac{f_1(x)}{h(x)} = \sqrt{\frac{2}{\pi}} e^{-\frac{x^2}{2} + x},$$

iar ecuația  $r'(x) = 0$  are soluția  $x_0 = 1$  care este un punct de maxim pentru  $r(x)$ , adică

$$r(x) \leq r(x_0) = \sqrt{\frac{2e}{\pi}}$$

ceea ce demonstrează teorema.

Deci un algoritm pentru generarea variabilei  $X \sim N(0, 1)$  este următorul:

## Algoritm Norm2

### Intrare:

P1: Se generează  $U \sim U(0,1)$ ;

P2: Se generează  $Y \sim \text{Exp}(1)$ ;

P3: Dacă  $U \leq e^{-\frac{Y^2}{2} + Y - 0.5}$  mergi la P4, altfel mergi la P1;

P4:  $X_1 := Y$ ;

P5: Se generează  $U \sim U(0,1)$ ;

P6: Dacă  $U \leq 0.5$  atunci  $s = 1$ , altfel  $s := -1$ ;

P7:  $X := sX_1$ .

**Ieșire:** Variabila aleatoare  $X$ .

Se observă că probabilitatea de acceptare este:

$$p_a = \sqrt{\frac{\pi}{2e}} \approx 0.72$$

adică în medie, din patru perechi  $(U, Y)$  trei sunt acceptate pentru a genera un  $X_1$ .

### 3. Metoda polară

#### Teoremă

*Dacă variabilele  $U_1, U_2$  sunt uniforme pe  $[0, 1]$  și independente, atunci variabilele aleatoare*

$$Z_1 = V_1 \sqrt{\frac{-2 \log S}{S}}, \quad Z_2 = V_2 \sqrt{\frac{-2 \log S}{S}} \quad (5)$$

*cu*

$$V_1 = 2U_1 - 1, \quad V_2 = 2U_2 - 1, \quad S = V_1^2 + V_2^2, \quad S < 1$$

*sunt variabile  $N(0, 1)$  independente.*

## Dem:

Trebuie să arătăm că repartiția bidimensională comună a variabilelor  $Z_1$  și  $Z_2$  condiționată de  $\{S < 1\}$  este repartiția comună a două variabile normale independente.

Observăm că  $(V_1, V_2)$  este un vector aleator uniform pe suprafața mărginită de pătratul  $I^2 = [-1, 1] \times [-1, 1]$  iar  $V_1, V_2$  sunt uniforme pe  $[-1, 1]$  și independente. Condiția  $S < 1$  face ca repartiția vectorului  $(V_1, V_2)$  condiționată de  $\{S < 1\}$  să fie vector uniform pe suprafața mărginită de cercul unitate. De aceea putem să-i scriem pe  $V_1$  și  $V_2$  în funcție de coordonatele polare:

$$V_1 = R \cos \theta, \quad V_2 = R \sin \theta$$

cu  $R$  și  $\theta$  variabile aleatoare  $0 \leq R \leq 1$ ,  $0 \leq \theta \leq 2\pi$ . Identificând aceste relații cu (5) rezultă că:

$$S = R^2, \quad Z_1 = \sqrt{-2 \log S} \cos \theta, \quad Z_2 = \sqrt{-2 \log S} \sin \theta \quad (6)$$

Dar și  $Z_1, Z_2$  se pot exprima în coordonate polare:

$$Z_1 = R' \cos \theta', \quad Z_2 = R' \sin \theta' \quad (7)$$

cu  $R', \theta'$  variabile aleatoare,  $R' \in \mathbb{R}_+, 0 \leq \theta' \leq 2\pi$ . Din (6) și (7) rezultă că:

$$\theta' = \theta, \quad R' = \sqrt{-2 \log S}$$

și pentru că  $V_1, V_2$  sunt independente pe  $I^2$ , atunci și perechile  $R, \theta$  și  $R', \theta'$  sunt independente.

Deoarece  $(V_1, V_2)$  are o repartiție uniformă pe cercul unitate rezultă că  $\theta = \theta'$  are o repartiție uniformă pe  $[0, 2\pi]$ , adică are densitatea:

$$\varphi(\theta) = \begin{cases} \frac{1}{2\pi} & \text{dacă } \theta \in [0, 2\pi] \\ 0, & \text{altfel.} \end{cases}$$



Repartiția lui  $R'$  este:

$$F(r) = P(R' \leq r) = P(\sqrt{-2 \log S} \leq r).$$

Dar  $S = R^2$  este uniformă pe  $[0, 1]$ . Atunci rezultă că:

$$F(r) = P(S \geq e^{-\frac{r^2}{2}}) = 1 - e^{-\frac{r^2}{2}}.$$

Prin urmare densitatea de repartiție a variabilei  $R'$  este:

$$\psi(r) = re^{-\frac{r^2}{2}}, \quad r \in [0, 1].$$

Vom determina acum funcția comună de repartiție a variabilelor  $Z_1$ ,  $Z_2$  condiționată de  $S < 1$ . Pentru aceasta considerăm domeniile:

$$D_{(r,\theta)} = \{(r, \theta) | r \cos \theta \leq z_1, r \sin \theta \leq z_2\}$$

$$D_{(x,y)} = \{(x, y) | x \leq z_1, y \leq z_2\}.$$

Atunci

$$F(z_1, z_2) = P(Z_1 \leq z_1, Z_2 \leq z_2) = \int \int_{D(r, \theta)} \frac{1}{2\pi} e^{-\frac{r^2}{2}} r dr d\theta.$$

Facem schimbările de variabile:

$$\theta = \arctg\left(\frac{x}{y}\right), \quad r = \sqrt{x^2 + y^2}$$

și rezultă că:

$$\begin{aligned} F(z_1, z_2) &= \frac{1}{2\pi} \int \int_{D(x, y)} e^{-\frac{x^2 + y^2}{2}} dx dy = \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x_1} e^{-\frac{x^2}{2}} dx \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x_2} e^{-\frac{y^2}{2}} dy, \end{aligned}$$

ceea ce demonstrează teorema.

Algoritmul de generare se deduce ușor din teoremă și el produce simultan două valori de selecție ale unor variabile  $N(0, 1)$  independente.

Observăm că se resping valorile pentru care  $S \geq 1$  iar probabilitatea de acceptare este:

$$p_a = \frac{\text{aria cercului } C(0, 1)}{\text{aria pătratului } [-1, 1] \times [-1, 1]} = \frac{\pi}{4}$$

Din demonstrația teoremei rezultă că variabilele  $Z_1, Z_2$  pot fi simulate și cu formulele:

$$Z_1 = \sqrt{-2 \log U_1} \cos(2\pi U_2), \quad Z_2 = \sqrt{-2 \log U_1} \sin(2\pi U_2) \quad (8)$$

În acest caz nu se fac respingeri, dar complexitatea algoritmului poate fi mai mare decât în cazul (5) din cauza funcțiilor trigonometrice și a funcției logaritm.