Generarea variabilelor neuniforme

Curs 4

November 11, 2016

Algoritmul 3 de respingere

Analiza performanței

Probabilitatea de acceptare p_a dă informații asupra vitezei algoritmului. O probabilitate p_a mare înseamnă acceptarea mai rapidă a lui Z_0 (când K este impar). Dar probabilitatea p_a nu este suficientă pentru a caracteriza pe deplin performanța algoritmului. Trebuie verificat câte Z_i sunt necesare pentru acceptarea unui Z_0. Nr. mediu de șiruri (care au $K+1$ variabile) generate până la acceptarea unui șir (și implicit al unei variabie Z_0) este $\frac{1}{p_a}$. Fie N^* variabila aleatoare care reprezintă numărul total de variabile $\{Z_i\}_{i\geq 0}$ generate până la acceptarea unui Z_0. Atunci:

$$E[N^*] = \frac{1}{p_a} E[K+1].$$

Observăm că:

$$E[K+1] = E[K] + 1$$

$$E[K] = \sum_{k=1}^{\infty} k \int_{-\infty}^{+\infty} \left[\frac{(G(x))^{k-1}}{(k-1)!} - \frac{(G(x))^{k}}{k!} \right] dG_0(x) =$$

$$= \int_{-\infty}^{+\infty} \left\{ \sum_{k=1}^{\infty} k \left[\frac{(G(x))^{k-1}}{(k-1)!} - \frac{(G(x))^{k}}{k!} \right] \right\} dG_0(x) =$$

$$= \int_{-\infty}^{+\infty} \left\{ 1 + \sum_{k=2}^{\infty} k \frac{(G(x))^{k-1}}{(k-1)!} - \sum_{k=1}^{\infty} k \frac{(G(x))^{k}}{k!} \right\} dG_0(x) =$$

$$= \int_{-\infty}^{+\infty} \left\{ 1 + \sum_{k=1}^{\infty} (k+1) \frac{(G(x))^{k}}{k!} - \sum_{k=1}^{\infty} k \frac{(G(x))^{k}}{k!} \right\} dG_0(x) =$$

$$= \int_{-\infty}^{+\infty} \left\{ 1 + \sum_{k=1}^{\infty} \frac{(G(x))^{k}}{k!} \right\} dG_0(x) = \int_{-\infty}^{+\infty} e^{G(x)} dG_0(x)$$

În concluzie

$$E[N^*] = \frac{1}{p_a}\left(1 + \int_{-\infty}^{+\infty} e^{G(x)} dG_0(x)\right). \tag{1}$$

Exemplu

Presupunem că variabilele Z_i, $i \geq 0$ *sunt uniforme pe intervalul* $[0,1]$, $Z_i = U_i$, $i \geq 0$. *Atunci conform celei de-a treia teoreme de respingere avem:*

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

cu

$$p_a = \int_0^1 e^{-x} dx = 1 - e^{-1}.$$

Cu alte cuvinte U_0 *acceptat are funcția de repartiție*

$$F(x) = \begin{cases} 0, \ \text{dacă } x < 0 \\ \frac{1-e^{-x}}{1-e^{-1}}, \ \text{dacă } 0 \leq x \leq 1 \\ 1, \ \text{dacă } x > 1 \end{cases}$$

care reprezintă funcția de repartiție a unei variabile $Exp(1)$ *trunchiată pe intervalul* $[0,1]$.
Numărul mediu $E[N^*]$ *de variabile* $\{Z_i\}_{i\geq 0}$ *care trebuie generate conform (1) este:*

$$E[N^*] = \frac{1}{1-e^{-1}}\left(1 + \int_0^1 e^x dx\right) = \frac{e}{1-e^{-1}} = \frac{e^2}{e-1} \approx 4.3.$$

Alte metode de generare

În această secțiune vom prezenta câteva metode de generare ale unor variabile aleatoare, metode care nu se înscriu în cazurile teoremelor de mai sus.

Variabila modul

Variabila aleatoare X are distribuția modul dacă densitatea ei de repartiție este:

$$f(x) = \begin{cases} 1 - |x|, \text{ dacă } x \in [-1, 1] \\ 0, \text{ altfel.} \end{cases}$$

Atunci funcția de repartiție a variabilei X este:

$$F(x) = \begin{cases} 0, \text{ dacă } x < -1 \\ x - |x|\frac{x}{2} + \frac{1}{2} \text{ dacă } x \in [-1, 1] \\ 1 \text{ dacă } x > 1. \end{cases}$$

Fie U_1, U_2 două variabile aleatoare uniforme pe $[0,1]$, independente. Atunci variabila aleatoare $Y = U_1 - U_2$ are aceeași distribuție ca și X.
Arătăm că funcția de repartiție a lui Y este F. Fie $y \in [-1, 0]$, atunci funcția de repartiție a lui Y, calculată în y este:

$$P(Y < y) = P(U_1 - U_2 < y) = \int_D du_1 du_2 = y + \frac{y^2}{2} + \frac{1}{2}.$$

Fie $y \in [0,1]$, atunci funcția de repartiție a lui Y, calculată în y este:

$$P(Y < y) = P(U_1 - U_2 < y) = \int_D du_1 du_2 = y - \frac{y^2}{2} + \frac{1}{2}.$$

Repartiția maximului

O variabilă aleatoare X are repartiția maximului dacă are densitatea de repartiție:

$$f(x) = \begin{cases} nx^{n-1}, \text{ dacă } x \in [0,1] \\ 0, \text{ altfel.} \end{cases}$$

Fie $U_1, U_2, ..., U_n$ variabile aleatoare uniforme pe $[0,1]$, independente. Atunci variabila aleatoare $Y = \max\{U_1, U_2, ..., U_n\}$ are aceeași repartiție ca și variabila aleatoare X.

Funcția de repartiție a variabilei X este:

$$F(x) = \begin{cases} 0 \text{ dacă } x < 0 \\ x^n \text{ dacă } x \in [0,1] \\ 1 \text{ dacă } x > 1. \end{cases}$$

Calculăm funcția de repartiție a variabilei Y în punctul $y \in [0,1]$:

$$P(Y < y) = P(\max\{U_1, U_2, ..., U_n\} < y) =$$

$$= P(U_1 < y, U_2 < y, ..., U_n < y) = \prod_{i=1}^{n} P(U_i < y) = y^n$$

Repartiția minimului

O variabilă aleatoare X are repartiția minimului dacă are densitatea de repartiție:

$$f(x) = \begin{cases} n(1-x)^{n-1}, \text{ dacă } x \in [0,1] \\ 0, \text{ altfel.} \end{cases}$$

Fie $U_1, U_2, ..., U_n$ variabile aleatoare uniforme pe $[0, 1]$, independente. Atunci variabila aleatoare $Y = \min\{U_1, U_2, ..., U_n\}$ are aceeași repartiție ca și variabila aleatoare X.
Funcția de repartiție a variabilei X este:

$$F(x) = \begin{cases} 0 \text{ dacă } x < 0 \\ 1 - (1 - x)^n \text{ dacă } x \in [0, 1] \\ 1 \text{ dacă } x > 1. \end{cases}$$

Calculăm funcția de repartiție a variabilei Y în punctul $y \in [0, 1]$:

$$P(Y < y) = P(\min\{U_1, U_2, ..., U_n\} < y) =$$

$$= 1 - P(\min\{U_1, U_2, ..., U_n\} \geq y) = 1 - P(U_1 \geq y, U_2 \geq y, ..., U_n \geq y)$$

$$= 1 - \prod_{i=1}^{n} P(U_i \geq y) =$$

$$= 1 - \prod_{i=1}^{n} (1 - P(U_i < y)) = 1 - (1 - y)^n$$

Repartiția Erlang

Fie X o variabilă aleatoare $Erlang(k)$, $k \in \mathbb{N}^*$, cu densitatea de repartiție:

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

Fie $Z_1, Z_2, ..., Z_k$ variabile distribuite $Exp(1)$ independente. Atunci variabila $Y = \sum_{j=1}^{k} Z_j$ are aceeași distribuție ca și variabila X.

Variabilele aleatoare Z_j pot fi generate cu metoda inversă și prin urmare putem să scriem:

$$X = -\ln\left\{\prod_{j=1}^{k} U_j\right\}.$$

Repartiția Normală

O variabilă X are repartiția normală standard, $N(0,1)$ dacă are densitatea de repartiție:

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

Teorema limită centrală (formă simplificată): Dacă $\{V_n\}_{n\in\mathbb{N}}$ este un șir de variabile aleatoare independente și identic distribuite care au momente de ordinul 1 și 2 și dacă $S_n = \sum_{i=1}^{n} V_i$, atunci

$$\lim_{n\to\infty} P\left(\frac{S_n - E[S_n]}{\sqrt{Var[S_n]}} < x\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt. \qquad (2)$$

În membrul al doilea este funcția de repartiție normală N(0,1). Dacă considerăm ca variabile V_i variabilele uniforme U_i, atunci:

$$E[U_i] = \frac{1}{2} \text{ și } Var[U_i] = \frac{1}{12}.$$

Viteza de convergență depinde de viteza de generare a variabilelor V_i. În cazul $V_i = U_i$ membrul stâng din (2) se apropie suficient de mult de funcția de repartiție normală $N(0,1)$ din membrul drept, dacă $n \geq 10$. Pentru $n = 12$ se obține:

$$E[S_n] = 6, \quad Var[S_n] = 1 \Rightarrow \sum_{i=1}^{12} U_i - 6$$

care are (aproximativ) repartiția normală $N(0,1)$.

Simularea unor repartiții înrudite cu repartiția normală
Repartiția χ^2
Fie $Z_1, Z_2, ..., Z_f$ variabile normale $N(0,1)$, independente. Atunci:

$$\chi_f^2 = \sum_{i=1}^{f} Z_i^2 \tag{3}$$

se numește variabilă χ^2 centrată, cu f grade de libertate.

Dacă $Z_i \sim N(m_i, 1)$ atunci variabila (3) se notează cu $\chi_{f,\delta}^2$ și se numește variabilă χ^2 necentrată, cu f grade de libertate și cu parametrul de excentricitate δ, unde

$$\delta^2 = \sum_{i=1}^{f} m_i^2.$$

Se poate arăta că χ_f^2 centrată este o variabilă Gamma$(0, \frac{1}{2}, \frac{f}{2})$.
Din formula (3) variabilele χ_f^2 și $\chi_{f,\delta}^2$ se pot simula direct folosind definiția lor.

Variabila t Student

Dacă $Z \sim N(0,1)$ este o variabilă independentă de variabila χ_f^2, atunci variabila

$$t_f = \frac{Z}{\sqrt{\frac{\chi_f^2}{f}}} \tag{4}$$

se numește variabila *t Student* cu f grade de libertate. Dacă în (4) în loc de χ_f se folosește $\chi_{f,\delta}^2$, atunci se obține o variabilă $t_{f,\delta}$ numită t Student necentrată, cu f grade de libertate și cu parametrul de excentricitate δ.
Variabilele t Student se simulează folosind (4).

Variabila F Snedecor

Dacă $\chi^2_{f_1}$, $\chi^2_{f_2}$ sunt independente, atunci variabila:

$$F_{f_1,f_2} = \frac{f_2\chi^2_{f_1}}{f_1\chi^2_{f_2}} \tag{5}$$

se numește variabila F a lui *Snedecor* centrată, cu f_1, f_2 grade de libertate.
Dacă în (5) se folosește câte una din $\chi^2_{f_1,\delta_1}$, $\chi^2_{f_2,\delta_2}$, sau ambele, atunci se obțin variabilele F simplu necentrate $F_{f_1,f_2,\delta_1,0}$, $F_{f_1,f_2,0,\delta_2}$ cu parametrii corespunzători de excentricitate, sau variabila F dublu necentrată $F_{f_1,f_2,\delta_1,\delta_2}$.
Variabilele F se pot simula direct din (5).

Variabila log-normală

Variabila aleatoare Y se numește log-normală $LN(\mu,\sigma)$ de parametrii μ și σ dacă variabila $X = \log(Y)$ este normală $N(\mu,\sigma)$.

Prin urmare, dacă densitatea variabilei X este:

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

atunci densitatea variabilei Y este:

$$g(y) = \frac{1}{\sqrt{2\pi} y\sigma} e^{-\frac{(\log(y)-\mu)^2}{2\sigma^2}} \quad y \in \mathbb{R}_+ \tag{6}$$

cu media m și dispresia s^2 date de:

$$m = E[Y] = e^{\mu+\sigma^2/2}, \quad s^2 = Var[Y] = m^2(e^{\sigma^2} - 1).$$

Dacă se cunosc m și s^2 atunci parametrii μ și σ ai variabilei log-normale se pot determina astfel:

$$\mu = log(m) - \frac{1}{2}\log\left[\frac{s^2}{m^2}+1\right], \quad \sigma^2 = \log\left[\frac{s^2}{m^2}+1\right]$$

Prin urmare simularea unei variabile aleatoare Y log-normale de medie m și dispersie s^2 se face prin următorul algoritm:

Algoritm Lnorm

Intrare: `m, s^2. Se calculează μ, σ.`

`P1: Se generează` $Z \sim N(0,1)$`;`

`P2: Calculează` $X = \mu + \sigma Z$ `(`X `este o variabilă repartizată` $N(\mu,\sigma)$`);`

`P3:` $Y = e^X$`;`

Ieșire: `Variabila aleatoare` Y `cu densitatea (6).`