Generarea funcțiilor de repartiție

Prima problemă în acest domeniu este obținerea unei prime funcții de repartiție de la care plecând să se construiască altele. Majoritatea limbajelor moderne de programare au incorporată repartiția uniformă pe intervalul [0,1), pe care noi o vom nota cu U.

Fie $f_{\boldsymbol{x}}(x)$ densitatea de repartiție a variabilei X. Cu ea putem construi o nouă variabilă aleatoare Y definită prin funcția $y = g(x)$ (continuă în general). Se pune problema care este densitatea de repartiție $f_{\boldsymbol{y}}(x)$ pentru variabila Y contruită astfel.

Pentru determinarea acesteia se cunoaşte următoarea teoremă [PAP 65]

Dacă ecuația $y = g(x)$ are un număr de n rădăcini reale $\{x_i\}_{i=1,\ldots,n}$ pentru un y dat atunci

$$f_{\boldsymbol{y}}(y) = \sum_{k=1}^{n} \frac{f_{\boldsymbol{x}}(x_k)}{|g'(x_k)|}$$

unde $g'(x) = \dfrac{dg(x)}{dx}$

Această teoremă este des aplicată când funcția g este monotonă.

Problema construirii unei variabile aleatoare cu o densitate de repartiție dată este o problemă inversă. Adică se presupun cunoscute $f_{\boldsymbol{x}}(x)$ şi $f_{\boldsymbol{y}}(x)$ şi se cere determinarea funcției $y = g(x)$. Această problemă nu are întotdeauna soluție, însă atunci când se poate rezolva ea determină aşa numita metodă a funcției inverse.

Metoda funcției inverse

Dacă dorim o funcție g strict monotonă teorema de mai înainte ne dă imediat o ecuație diferențială pentru determinarea funcției g

$$f_{\boldsymbol{y}}(y)g'(x) = f_{\boldsymbol{x}}(x)$$

sau sub formă integrală

$$F_{\boldsymbol{y}}(g) = \int_0^g f_{\boldsymbol{y}}(y)dy = \int_0^x f_{\boldsymbol{x}}(\xi)d\xi = F_{\boldsymbol{x}}(x)$$

de unde rezultă imediat

$$g(x) = F_{\boldsymbol{y}}^{-1}(F_{\boldsymbol{x}}(x))$$

care explică numele metodei.

Construirea unei noi variabile aleatoare pornind de la o alta se poate generaliza astfel:

Fie un set de variabile aleatorii $\{X_k\}_{k=1,\ldots,n}$. Pe baza acestui set se poate construi variabila aleatoare Z prin definiția

$$z = g(x_1, x_2, \ldots, x_n)$$

fiecare din variabile având propria funcție densitate de repartiție $f_{\boldsymbol{x}_k}(x_k)$.

Unul dintre modurile cele mai simple de alegere a funcției g este combinația liniară

$$z = \sum_{k=1}^{n} \alpha_k x_k$$

Un caz particular des întâlnit este $z = x + y$ când cele două variabile *X, Y* sunt independente. Se poate arăta pentru acest caz că funcția de densitate a sumei se poate calcula cu

$$f_z(z) = \int_{-\infty}^{\infty} f_x(x) f_y(z-x) dx = \int_{-\infty}^{\infty} f_x(z-y) f_y(y) dy$$

care nu este altceva decât convoluția celor două funcții densitate.

Observăm că dacă cele două variabile pot lua numai valori pozitive formula de mai sus devine

$$f_z(z) = \int_{0}^{z} f_x(z-y) f_y(y) dy$$

Prezentăm în tabelul următor funcțiile densitate utilizate în această lucrare deduse prin metoda funcției inverse

Nume	Funcția densitate	Funcția generatoare	Codomeniul
exponențială	$f(t,\Gamma) = \Gamma e^{-\Gamma t}$	$g(x;\Gamma) = -\frac{1}{\Gamma}\ln(1-x)$	$[0,\infty)$
sinus	$f(x) = \frac{1}{2}\sin x$	$g(x) = \arccos(1-2x)$	$[0,\pi)$
Lorenz	$f(x;\Gamma) = \frac{\Gamma}{\pi}\frac{1}{x^2+\Gamma^2}$	$g(x;\Gamma) = \Gamma \tan\left(\pi x - \frac{\pi}{2}\right)$	$(-\infty,\infty)$

Pentru toate exemplele din tabel, funcția densitate inițială este funcția uniformă în [0,1) iar ultima coloană precizează codomeniul funcției *g*.

Metoda polară folosită la deducerea distribuției normale

Funcția de distribuție Gauss sau normală este definită de

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\bar{x})^2}{2\sigma^2}}$$

unde σ abaterea standard, $\sigma^2 = \overline{(x-\bar{x})^2}$ este varianța sau dispersia. Aici vom nota clasa variabilelor de acest tip cu $N(\bar{x},\sigma)$

Metoda folosită de noi pentru generarea variabilei normale rezultă din următoarea teoremă a lui Box şi Müller [VAD 77]

Dacă U_1 şi U_2 sunt două variabile independente uniforme pe (0,1) atunci variabilele aleatoare: $Z_1 = V_1\sqrt{-2\frac{\ln S}{S}}$, $Z_2 = V_2\sqrt{-2\frac{\ln S}{S}}$ sunt variabile aleatoare normale dacă $S < 1$.

Notațiile folosite sunt: $V_1 = 2U_1 - 1$, $V_2 = 2U_2 - 1$ şi $S = V_1^2 + V_2^2$.

Metoda folosită la generarea modulului vectorului viteză conform unei distribuții maxwelliene

Funcția de distribuție maxwelliană satisface

$$f(v_x, v_x, v_x)dv_x dv_y dv_z = \left(\sqrt{\frac{\beta}{\pi}}\right)^3 e^{-\beta v^2} v^2 \sin\theta d\theta d\varphi$$

unde $\beta = \frac{m}{2kT}$ iar integrala de tip Poisson satisface $\int_{-\infty}^{\infty} e^{-\beta x^2} = \sqrt{\frac{\pi}{\beta}}$.

Funcție de repartiție după vitezele moleculelor revine astfel la distribuțiile

$$f_v = 4\left(\sqrt{\frac{\beta^3}{\pi}}\right) e^{-\beta v^2} v^2$$

$$f_\theta = \frac{1}{2}\sin\theta d\theta$$

$$f_\varphi = \frac{1}{2\pi} d\varphi$$

Generarea repartiției f_v

Această repartiție poate fi obținută pornind de la generarea variabilei χ^2.

Variabila χ^2 (hi-pătrat cu ν grade de libertate) definită de

$$\chi_\nu^2 = \sum_{k=1}^{\nu} N_k^2$$

unde N_k sunt variabile independente normale N(0,1). Variabila χ^2 are densitatea de repartiție

$$f(x) = \frac{1}{2^{\frac{\nu}{2}}\Gamma\left(\frac{\nu}{2}\right)} x^{\frac{\nu}{2}-1} e^{-\frac{x}{2}}, \; x > 1.$$

Generarea acestei variabile se poate face direct prin formula de definiție sau generând funcția de mai sus. Dacă $\nu = 2k$, ($k \in N_+$) generarea lui χ_{2k}^2 se reduce la generarea funcției Erlang E_k adică $\chi_{2k}^2 = E_k$. Dacă însă $\nu = 2k+1$ atunci din proprietatea de stabilitate a variabilei χ^2 rezultă că $\chi_{2k=1}^2 = \chi_{2k}^2 + \chi_1^2 = E_k + N^2$ unde N este o variabilă normală N(0,1).

Funcția de repartiție Erlang este definită de

$$f_{E_k}(x) = \begin{cases} \frac{1}{\Gamma(k)} x^{k-1} e^{-x}, x > 0 \\ 0, x \le 0 \end{cases}, \; \Gamma(k) = (k-1)!$$

şi ea poate fi generată cu funcția

$$y = -\ln \prod_{n=1}^{k} U_n$$

unde U_n sunt variabile independente uniforme.

În cazul nostru cele trei variabile cu distribuție N(0,1) sunt mărimile $\frac{v_\xi}{\sqrt{kT/m}}$ cu $\xi = x, y, z$.

În concluzie pentru generarea funcției de repartiție a modulului vitezei se procedează după cum urmează:

- se generează variabila E_1 folosind funcția $y = -2\ln U$, unde U este o variabilă uniform repartizată pe [0,1).
- se generează $z = (N)^2$ unde N este de clasă N(0,1).
- se calculează $x = \sqrt{y+z}$ care este o variabilă cu funcția de repartiție

$$f(u) = \sqrt{\frac{2}{\pi}} u^2 e^{-\frac{1}{2}u^2}$$

care reprezintă repartiția mărimii $\frac{v}{\sqrt{kT/m}}$, Astfel că modulul vitezei corespunzătoare repartiției unei temperaturi date T se calculează cu $v = x\sqrt{kT/m}$.