3.
Generacja liczb losowych o różnych rozkładach

1. Jak uzyskać liczby pseudolosowe za pomocą komputera? [Zieliński]

nieliniowe sprzężenie zwrotne

$$x_k = \mathcal{F}(x_{k-1}, x_{k-2}, ..., x_{k-q})$$

Postulaty dotyczące $\mathcal{F}$:

1) powinna gwarantować jak największy okres ciągu $\{x_k\}$ (często opiera się ona na operacji wyznaczania reszty z dzielenia)

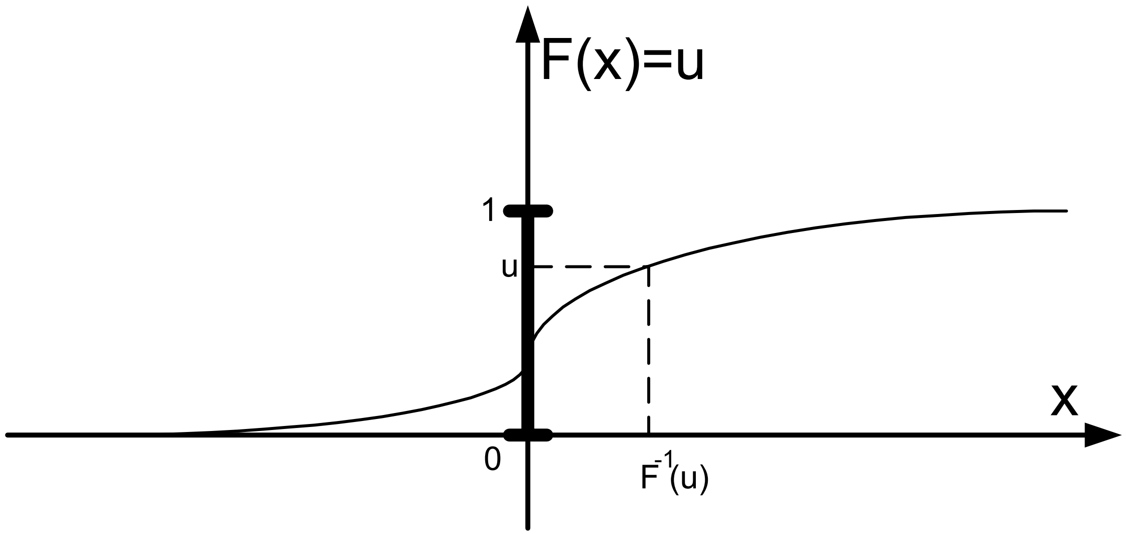
2) powinna gwarantować w miarę regularny rozkład wartości x_k

2. Metoda odwracania dystrybuanty (tzw. inwersyjna)

Założenie i cel:

– dysponyjemy dobrym generatorem realizacji zmiennej lowosej o rozkładzie $\mathcal{U}[0, 1]$

– chcemy zaprojektować generator realizacji zmiennej losowej X, o funkcji gęstości prawdopodobieństwa $f(x)$, której dystrybuanta $F(x) = P(X \leqslant x) = \int_{-\infty}^{x} f(x)dx$ jest ściśle monotoniczna (tj. ściśle rosnąca)



Wtedy $F^{-1}()$ istnieje i jest dobrze określona – odwzorowanie wzajemnie jednoznaczne

$$\begin{aligned} u &= F(x), \qquad u \in [0,1] \\ x &= F^{-1}(u), \qquad x \in R \end{aligned}$$

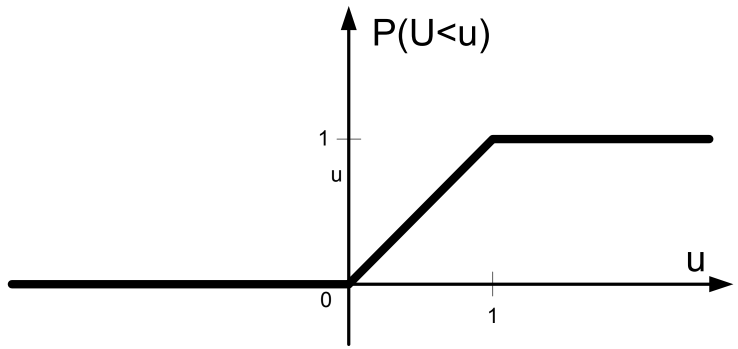
odcinek $[0,1]$ jest uprzywilejowany ze względu na definicję dystrybuanty

Lemat 1

Jeżeli $U\tilde{}\mathcal{U}[0,1]$, to zmienna losowa $X = F^{-1}(U)$ ma dystrybuantę $F(X)$ (czyli gęstość $f(x)$).

Dowód

$P\{X \leqslant x\} = P\left\{F^{-1}(U) \leqslant x\right\} = P\{U \leqslant F(x)\} = (...)$



$(...) = F(x)$

Lemat 2

Jeżeli $X\tilde{}F(x)$, to zmienna losowa $U = F(X)\tilde{}\mathcal{U}[0,1]$.

Dowód

$P(U \leqslant u) = P(F(X) \leqslant u) = P(X \leqslant F^{-1}(u)) = F(F^{-1}(u)) = u$

Schemat metody

1) weź dobry generator $\mathcal{U}[0,1]$

2) dla funkcji gęstości prawdopodobieństwa (konstruowanego generatora) $f(x)$ wyznacz dystrybuantę $F()$ i wzór na jej odwrotność $F^{-1}()$

3) wylosuj u~$\mathcal{U}[0,1]$ i oblicz $F^{-1}(u)$

Przykład

zaprojektować generator rozkładu wykładniczego

$$\begin{aligned}
f(x) &= \begin{cases} \alpha e^{-\alpha x}, \text{ gdy } x \geqslant 0 \\ 0, \text{ gdy } x < 0 \end{cases} \\
F(x) &= \int_0^x \alpha e^{-\alpha x} dx = [-e^{-\alpha x}]_0^x = 1 - e^{-\alpha x} = u \\
e^{-\alpha x} &= 1 - u \\
-\alpha x &= \ln(1-u) \\
x &= -\frac{\ln(1-u)}{\alpha}
\end{aligned}$$

spostrzeżenie

$$\text{generator } -\frac{\ln u}{\alpha} \text{ "ma ten sam rozkład"}$$

[przykład w Statistica]

Generatory do samodzielnego zaprojektowania

o rozkładzie trójkątnym

o rozkładzie Laplace'a

o rozkładzie Cauchy'ego

Przybliżony generator rozkładu normalnego $\mathcal{N}$(0,1)

$$
\begin{aligned}
& \Phi(x) \text{ – nie ma jawnej postaci analitycznej} \\
\text{dla } x \geqslant\ & 0,\ \Phi_+(x) \approx F_+(x) = \frac{1}{1+e^{-kx_+}},\ \text{gdzie } k = \sqrt{\frac{8}{\pi}} \\
& F_+(x) \text{ – ma rozkład równomierny na odcinku } [\frac{1}{2}, 1] \\
u =\ & 2(F_+(x) - \frac{1}{2}) \text{ – ma rozkład równomierny na odcinku } [0,1] \\
u =\ & 2F_+(x) - 1 = \frac{2}{1+e^{-kx_+}} - 1 \\
1 + e^{-kx_+} =\ & \frac{2}{u+1} \to x_+ = \frac{1}{k} \ln \frac{1+u}{1-u} \to x = \eta x_+,\ \text{gdzie } P(\eta = -1) = P(\eta = 1) = \frac{1}{2}
\end{aligned}
$$

Generacja $\mathcal{N}$(0,1) z zastosowaniem CTG – tzw. prawo tuzina

$$
\begin{aligned}
& \text{mamy ciąg } \{u_i\}_{i=1}^N \text{ typu i.i.d., } u_i \tilde{\ } \mathcal{U}[0,1] \\
\text{ponieważ } \mathbf{E}u_i =\ & \frac{1}{2} \text{ i } var u_i = \frac{1}{12} \text{ to} \\
\xi_N =\ & \frac{\sum_{i=1}^N u_i - \frac{N}{2}}{\sqrt{\frac{N}{12}}} \to \mathcal{N}(0,1) \text{ według rozkładu, gdy } N \to \infty
\end{aligned}
$$

dla $N = 12$

$$\xi_N = \sum_{i=1}^{12} u_i - 6$$

Jak na podstawie generatora $\mathcal{N}$(0,1) zbudować generator $\mathcal{N}(\boldsymbol{\mu},\boldsymbol{\sigma}^2)$?

$$x \tilde{\ } \mathcal{N}(0,1) \Rightarrow \sigma x + m \tilde{\ } \mathcal{N}(\mu, \sigma^2)$$

3. Metoda odrzucania

Podstawa teoretyczna

Lemat 1. Jeżeli $X \tilde{} f(x)$ i $U \tilde{} \mathcal{U}[0,1]$ to para (X,Y), gdzie $Y = cf(X)U$, zać c jest dowolną stałą, czyli para

$$(X, cf(X)U)$$

ma rozkład jednostajny na zbiorze

$$A = \{(x,y) : x \in R, y \in [0, cf(x)]\}$$

Dowód.

obliczamy dystrybuantę warunkową (przy ustalonym X)

$$F_{y|x}(\alpha) = P(Y \leqslant \alpha | X = x) = P(cf(X)U \leqslant \alpha | X = x) = P(cf(x)U \leqslant \alpha) = P(U \leqslant \frac{\alpha}{cf(x)}) = \frac{\alpha}{cf(x)}$$

gęstość warunkowa

$$f_{y|x}(\alpha) = \frac{\partial F_{y|x}(\alpha)}{\partial \alpha} = \frac{1}{cf(x)} \text{ – rozkład warunkowy jest jednostajny}$$

weźmy dowolny zbiór $B \subseteq A$

$$P((X,Y) \in B) = P(B) = \iint_B f(x,y)dxdy = [\text{tw. Bayesa}] = \iint_B f_{y|x}(\alpha) f(x) dx d\alpha = \frac{1}{c} \iint_B dxdy = \frac{1}{c}\mu(B)$$

w szczególności dla $B = A$

$$P((X,Y) \in A) = 1 = \frac{1}{c}\mu(A), \text{ stąd } c = \mu(A)$$

wniosek

$$P(B) = \frac{\mu(B)}{\mu(A)} \text{ – zatem rozkład na całym zbiorze } A \text{ jest równomierny}$$

Lemat 2. Jeżeli pewna para zmiennych losowych (X, Y) ma rozkład równomierny na zbiorze

$$A = \{(x, y) : x \in R, y \in [0, cf(x)]\}$$

to zmienna X ma funkcję gęstości prawdopodobieństwa równą $f(x)$.

Dowód.

wprowadzamy oznaczenie D – dowolny podzbiór zbioru liczb rzeczywistych

$$D \subseteq R$$

wprowadzamy oznaczenie B_D – zbiór par (x, y), takich że $x \in D$

$$B_D = \{(x, y) : x \in D, y \in [0, cf(x)]\} \subseteq A$$

wyznaczamy rozkład X

$$P(X \in D) = P(D) = P((X, Y) \in B_D) = P(B_D) = [\text{r. równ.}] = \frac{\mu(B_D)}{\mu(A)} = \frac{\int_D cf(x)dx}{\int_R cf(x)dx} = \int_D f(x)dx, \text{ c.n.u.}$$

Schemat metody

$f(x)$ – gęstość prawdopodobieństwa zmiennej X, której realizacje chcemy wygenerować

$g(x)$ – pomocnicza gęstość innej (łatwej w generacji) zmiennej losowej oraz isnieje $c > 0$, takie że

$$f(x) \leqslant cg(x) \text{ dla każdego } x \in R$$

Etap 1. Generacja

– wygenerować realizacje x zm. los. X o gęstości $g(x)$ (np. metodą inwersyjną)

– wygenerować realizacje u zm. los. U z rozkładu równomiernego $\mathcal{U}[0, 1]$

– utworzyć pary $(x, y) = (x, cg(x)u)$, o rozkładzie równomiernym na zbiorze $A_g = \{(x, y) : x \in R, y \in [0, cg(x)]\}$

Etap 1. Odrzucanie

– odrzucić te pary (x, y) z *Etapu 1*, które nie należą do zbioru $A_f = \{(x, y) : x \in R, y \in [0, f(x)]\}$, tzn. pozostają pary spełniające warunek

$$cg(x)u \leqslant f(x)$$

i mają one rozkład równomierny na zbiorze A_f

Wniosek – generator jest dokłany (żadnego przybliżenia).

Warunki stosowalności metody

1) dla funkcji $f(x)$ istnieje odpowiednie $g(x)$

2) $g(x)$ jest łatwe w generacji

3) $g(x)$ musi "ciasno" opasać $f(x)$, dla zagwarantowania dużego prawdopodobieństwa przyjęcia

$$P(\text{przyjęcia } (x,y)) \;=\; \frac{\mu(A_f)}{\mu(A_g)} = \frac{\int_{-\infty}^{\infty} f(x)dx}{c\int_{-\infty}^{\infty} g(x)dx} = \frac{1}{c}$$

$$\text{chcemy, aby } \frac{1}{c} \;\to\; \max\text{, ale } c \geqslant 1$$

Dokładny generator rozkładu normalnego

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \qquad\qquad g(x) = \frac{1}{2} e^{-|x|} \text{ – rozkład Laplace'a}$$

można pokazać, że

$$f(x) \leqslant cg(x),\text{ gdzie } c = \sqrt{\frac{2e}{\pi}}$$

określenie warunku selekcji

$$\sqrt{\frac{2e}{\pi}}\frac{1}{2}e^{-|x|}u \;\leqslant\; \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$$
$$\cdots$$
$$(|x|-1)^2 \;\leqslant\; -2\ln u$$

przed odrzuceniem realizacje x mają rozkład Laplace'a, natomiast po odrzuceniu – $x\tilde{\ }\mathcal{N}(0,1)$