

Generowanie dowolnych rozkładów

Dariusz Borkowski

Metoda odwracania dystrybuanty

Założmy, że U jest zmienną losową o rozkładzie równomiernym na odcinku $(0, 1)$. Niech F będzie ciągłą i ściśle rosnącą dystrybuantą pewnego rozkładu prawdopodobieństwa. Zdefiniujmy nową zmienną losową

$$X = F^{-1}(U).$$

Ponieważ

$$P\{X \leq x\} = P\{F^{-1}(U) \leq x\} =$$

$$P\{U \leq F(x)\} = F(x)$$

więc zmienna losowa X ma rozkład prawdopodobieństwa o dystrybuancie F .

Opisane rozumowanie można uogólnić na przypadek dowolnych, niekoniecznie ciągłych i ściśle rosnących dystrybuant F . W tym celu wystarczy przyjąć

$$X = F^{-1}(U),$$

gdzie

$$F^{-1}(t) = \inf\{x : t \leq F(x)\}.$$

Przykład 1. *Poniższa funkcja jest "wystarczającym" przybliżeniem odwrotności dystrybuanty rozkładu normalnego $N(0, 1)$.*

$$\Phi^{-1}(u) = \begin{cases} g(u) & 10^{-20} < u < 0.5 \\ -g(1-u) & 0.5 \leq u < 1 - 10^{-20} \end{cases}$$

gdzie

$$g(u) = t - \frac{L(t)}{M(t)} \quad t$$

$$L(t) = 0.322232431088 + t + 0.342242088547t^2 + \\ 0.0204231210245t^3 + 0.0000453642210148t^4$$

$$M(t) = 0.0993484626060 + 0.588581570495t \\ + 0.531103462366t^2 + 0.103537752850t^3 \\ + 0.0038560700634t^4.$$

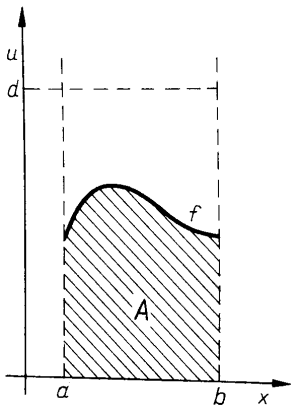
W ogólności odwracanie dystrybucyj jest zabiegiem skomplikowanym numerycznie i dlatego tą metodę rzadko się stosuje.

Metoda eliminacji

Niech f będzie gęstością prawdopodobieństwa interesującego nas rozkładu, dodatnią na pewnym ograniczonym przedziale (a, b) , równą zero poza tym przedziałem i ograniczoną przez pewną stałą $d > 0$. Następujący algorytm generuje liczby losowe o rozkładzie z gęstością $f(x)$:

1. Wygenerować dwie niezależne zmienne losowe U_1 i U_2 o rozkładach równomiernych $U(a, b)$ i $U(0, d)$.
2. Jeżeli $U_2 \leq f(U_1)$, to przyjąć $X = U_1$; w przeciwnym przypadku parę (U_1, U_2) wyeliminować i powtórzyć obliczenia od wygenerowania nowej pary zgodnie z punktem 1.

Wadą metody eliminacji jest fakt, że musimy wygenerować dwie kolejne niezależne liczby z rozkła-



du równomiernego. Taki generator dość często jest obarczony niedoskonałością i takie liczby mogą nie być niezależne, a więc użycie kilku kolejnych liczb

może deformować wynikowy rozkład prawdopodobieństwa. Obserwuje się to szczególnie wtedy, gdy używa się równomiernych generatorów wbudowanych w kompilatory typowych języków programowania.