# Generowanie liczb (pseudo)losowych

Seminarium: Algorytmy numeryczne i graficzne

Dawid Żywczak

06 kwietnia 2020

# Wprowadzenie

- RNG w życiu codziennym
- Zastosowania
- Co będziemy tak właściwie robić?

Pierwszy zaproponowany generator liniowy (algorytm kwadratowy von Neumanna)

$F(X_n) = X_{n+1}$ gdzie funkcja F oblicza $Z = X_n^2$, jeżeli trzeba uzpełnia liczbę Z wiodącymi zerami, tak aby miała $2 \cdot N$ cyfr, a następnie wycina z liczby Z N środkowych cyfr.

Ogólna postać generatora liniowego

$$X_{n+1} = (a_1 X_n + a_2 X_{n-1} + ... + a_k X_{n-k+1} + c) \text{ mod m}$$

Dodatkowo zdefiniujmy pojęcie okresu

Niech $P = \min\{i : X_i = X_0, i > 0\}$ oraz $v \in \mathbb{N}$ wtedy jeśli $\forall i$, $i \geq v$ zachodzi $X_i = X_{i+j \cdot P}$, $j = 1, 2, ...$ to fragment ciągu $X_0, X_1, ..., X_{v+P-1}$ nazywamy okresem aperiodyczności ciągu, v parametrem aperiodyczności, a liczbę P okresem ciągu.

Dla generatora multiplikatywnego najczęściej wykorzystywany jest punkt 4 twierdzenia 3. Czyli aby otrzymać maksymalny okres, powinniśmy dobierać $m = 2^e$ $e \geq 4$ oraz $a = 3 \mod 8$ lub $a = 5 \mod 8$.
Inna możliwość to m pierwsze. (Dokładnie - twierdzenia 1, 2, 3)

- Okresowość ostatnich bitów
- Struktura przestrzenna

Ogólna postać generatorów opartych na rejestrach przesuwnych

$$b_i = (a_1 b_{i-1} + \ldots + a_k b_{i-k}) \bmod 2, \; i = k+1, k+2, \ldots$$

Korzystamy z faktu, że łatwo zaimplementować na komputerze. Mając wzór na poszczególne bity, możemy generować

$$U_i = \sum_{j=1}^{L} 2^{-j} b_{is+j} = 0.b_{is+1} b_{is+2} \ldots b_{is+L} \text{ gdzie s jest ustaloną liczbą naturalną}$$

Generator Tauswortha:

- $B = ((A << q) \text{ xor } A) << (L - p)$
- $B = ((A << s) \text{ xor } A) >> (L - s)$
- Return A

Generatory Fibonacciego:

- Ogólna forma $X_n = X_{n-1} + X_{n-2} \mod m$
- Uogólnienie
- Zmiana działania

Generatory nieliniowe:

- Eichenauera-Lehna $X_{n+1} = (aX_n^{-1} + b) \bmod m$
- Eichenauera-Hermanna $X_{n+1} = (a(n + n_0) + b)^{-1} \bmod m$

Załóżmy, że mamy zmienne losowe X i Y określone na zbiorze S={1, 2,..., n} o rozkładach prawdopodobieństwa równych:

$$P(X = i) = p_i,\ P(Y = i) = q_i, \text{ i=1, 2,..., n}$$

Zdefiniujmy teraz normę wektora t=$(t_1, t_2, ..., t_n)$ dla ustalonego p=1, 2,...,$\infty$ jako:

$$\|t\| = (\sum_{i=1}^{n} t_i^p)^{\frac{1}{p}}$$

Teraz dla zdefiniowanej wyżej zmiennej X wprowadźmy miarę podobieństwa do rozkładu jednostajnego na zbiorze S:

$$\delta(X) = \|(p_1, p_2, ..., p_n) - (\tfrac{1}{n}, \tfrac{1}{n}, ..., \tfrac{1}{n})\|$$

Okazuje się, że dla "dobrych"działań $\circ$ zachodzi:

$$\delta(X \circ Y) \leq min\{\delta(X), \delta(Y)\}$$

Czyli ciąg $(X_1 \circ Y_1, X_2 \circ Y_2, ...)$ powinien być bardziej równomiernie rozłożony niż każdy z ciągów składowych.

Dla rozkładów ciągłych:

- Metoda odwracania dystrybuanty
- Metoda eliminacji

Metody dla rozkładów dyskretnych są podobne i ich opis znajduje się w notatce.

Metoda odwracania dystrybuanty korzysta z faktu, że dla zmiennej losowej U o rozkładzie jednostanym U(0,1) oraz ściśle rosnącej i ciągłej dystrybuanty F, zmienna X=$F^{-1}(U)$ ma rozkład prawdopodobieństa o dystrybuancie F, ponieważ

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

Przykład metody odwracania dystrybuanty dla rozkładu wykładniczego

$$F(x) = 1 - e^{-x}$$
$$x = 1 - e^{-y}$$
$$1 - x = e^{-y}$$
$$-ln(1 - x) = y = F^{-1}(x)$$

Zauważmy, że $1 - x$ zachowuje się jak zmienna losowa o rozkładzie jednostajnym $U(0, 1)$ zatem $F^{-1}(x) = -lnU$.

Metoda eliminacji wymaga wprowadzenia najpierw kilku oznaczeń oraz twierdzeń. Omówimy działanie metody eliminacji dla dwuwymiarowego punktu losowego oraz podamy schemat dla uogólnionej na k wymiarów wersji. Uogólnienie to jest dokładnie omówione w notatce.

Ogólny schemat można przedstawić w dwóch krokach:
Niech f jest gęstością oczekiwanego rozkładu ppb, dodatnią na pewnym przedziale (a,b) i ograniczoną przez stałą $d > 0$. Wtedy liczba X generowana według poniższego schematu ma rozkład o gęstości f(x)

- Generuj dwie niezależne zmienne losowe $U_1$ $U(a, b)$ i $U_2$ $U(0, d)$
- Jeśli $U_2 \leq f(U_1)$ to $X = U_1$ wpp. powtórzyć generowanie.

Niech $A = \{(x, u) : a \leq x \leq b,\ 0 \leq u \leq f(x)\}$.

**Twierdzenie**

*Niech $(X_1, U_1), (X_2, U_2), \ldots$ będzie ciągiem punktów losowych o rozkładzie równomiernym na prostokącie $(a, b) \times (0, d)$ i niech $(X,U)$ będzie pierwszym punktem tego ciągu, który wpada do zbioru $A$. Wtedy punkt losowy $(X,U)$ ma rozkład jednostajny na zbiorze $A$.*

## Dowód.

Rozważmy pozdbiór B zbioru A i niech $l_2(B)$ oznacza jego pole powierzchni. Chcemy udowodnić, że $P((X,U)\in B) = l_2(B)/l_2(A)$.

$$P((X,U)\in B) = \sum_{i=1}^{\infty} P((X_1,U_1)\notin A, ..., (X_{i-1},U_{i-1})\notin A, (X_i,U_i)\in B) = \sum_{i=1}^{\infty}(1-\frac{l_2(A)}{(b-a)d})^{i-1}\frac{l_2(B)}{(b-a)d} = \text{(z sumy szeregu geometrycznego)}\frac{l_2(B)}{l_2(A)}$$

□

**Twierdzenie**

*a) Jeżeli U ma rozkład jednostajny U(0,1), X ma rozkład o gęstości f(x) oraz X i U są niezależne, to punkt losowy (X, Uf(X)) ma rozkład jednostajny na zbiorze A.*
*b) Jeżeli punkt losowy (X, U) ma rozkład jednostajny na zbiorze A, to zmienna losowa X ma rozkład o gęstości f(x).*

## Dowód.

a) Jeżeli U ma rozkład jednostajny U(0,1), to dla każdego ustalonego x zmienna losowa $V = Uf(x)$ ma rozkład jednostajny $U(0, f(x))$. Dla ustalonego x i dla danego zbioru $B \subset A$ oznaczamy $B_x$=$\{u : (x, u) \in B\}$

$$P((X, Uf(X)) \in B) = \int (\int_{B_x} \frac{dv}{f(x)}) f(x) dx = \int \int_B dvdx = l_2(B) = \frac{l_2(B)}{l_2(A)}$$

czyli (X, Uf(X)) ma rozkład jednostajny na zbiorze A.
b) Oznaczmy $A_t = \{(x, u) : a \leq x \leq t, 0 \leq u \leq f(x)\}$. Wtedy

$$P(X \leq t) = P((X, U) \in A_t) = \int_a^t \int_0^{f(x)} \frac{dudx}{l_2(A)} = \int_a^t f(x) dx$$

czyli f(x) jest gęstością zmiennej losowej X. □

1. Wybierz gęstośc g, żeby generowanie liczb losowych o tej gęstości było łatwe i szybkie oraz wyznacz stałą $c > 0$, taką żeby $f(x) \leq cg(x)$ dla wszystkich x.
Ze względu na ten warunek gęstość g, będziemy nazywać dominującą. Za obszar $\Omega$ przyjąć
$\Omega = \{(x, u) : x \in \mathbb{R}^k, 0 \leq u \leq cg(x)\}$.
2. Wygeneruj punkt losowy X o rozkładzie z gęstością g oraz liczbę losową U U(0,1), wtedy punkt losowy (X, cUg(X)) ma rozkład jednostajny na zbiorze $\Omega$.
3. Powtarzać generowanie według p. 2, dopóki kolejno wygenerowany punkt nie wpadnie do zbioru
$A = \{(x, u) : x \in \mathbb{R}^k, 0 \leq u \leq f(x)\}$, tzn. dopóki nie zostanie spełniony warunek akceptacji

$$U \leq \frac{f(X)}{cg(X)}$$

Zastanówmy się jak wyznaczyć stałą c, tak aby warunek akceptacji był jak najszybciej osiągany. Możemy to osiągnąć przez dobranie wartości c takiej, żeby prawdopodobieństwo spełnienia warunku akceptacji było jak największe tzn.

$$P(Ucg(X) \leq f(X)) = \int_{\mathbb{R}^k} g(x)dx \int_0^{f(x)/cg(x)} du = \frac{1}{c}$$

Optymalną wartością powyższego warunku jest $c = \sup_x \frac{f(x)}{g(x)}$.

# Generatory wielowymiarowe

- Wielowymiarowy rozkład jednostajny
- Wielowymiarowy rozkład normalny

Niech

$$A= \begin{bmatrix} \sigma_{1,1} & \dots & \sigma_{1,n} \\ \dots & \dots & \dots \\ \sigma_{n,1} & \dots & \sigma_{n,n} \end{bmatrix}$$

Jesli $Z = (Z_1, ..., Z_n)$ oraz każda składowa Z jest niezależna i ma taki sam rozkład N(0,1), to zmienna losowa CZ, gdzie C jest pewną nieosobliwą macierzą, ma rozkład normalny z macierzą kowariancji $CC^T$. Potrzebujemy więc macierz C taką, że $CC^T = A$

$$c_{i,1} = \frac{\sigma_{i,1}}{\sqrt{\sigma_{1,1}}}$$
$$c_{i,i} = (\sigma_{i,i} - \sum_{r=1}^{i-1} c_{i,r}^2)^{1/2}$$
$$c_{i,j} = \sigma_{j,j}^{-1}(\sigma_{i,j} - \sum_{r=1}^{j-1} c_{i,r}c_{j,r}) \text{ gdy } i > j$$
$$c_{i,j} = 0 \text{ gdy i<j}$$

## Testowanie poprawności generatorów

- Opis metodologii
- Test $\chi^2$
- Test Kołomogorowa
- Test pokerowy

- Cel testu
- Statystyka - $\phi = \sum_{i=1}^{k} \frac{(n_i - np_i)^2}{np_i}$

- Cel testu
- Dystrybuanta empiryczna - $F_n(x) = \frac{1}{n}\sum_{j=1}^{n} 1_{(-\infty,x]}(X_j)$ gdzie $1_{(a,b)}$ to funckja charakterystyczna zbioru (a,b)
- Statystyka - $D_n = \sup_{-\infty<x<+\infty} |F_n(x) - F(x)|$

- Cel testu
- Opis procesu testowania:
  $X_1, X_2, ..., X_n \rightarrow Y_j = i$ jeśli $X_j \in (a_i, a_{i+1}) \rightarrow$ pięcioelementowe krotki $\rightarrow$ sprawdzenie rozkładu