Metody Obliczeniowe w Nauce i Technice

Liczby losowe (random numbers)

Marian Bubak, Katarzyna Rycerz

Department of Computer Science
AGH University of Science and Technology
Krakow, Poland
kzajac@agh.edu.pl
dice.cyfronet.pl

Contributors
Dawid Prokopek
Paweł Matejko
Kamil Doległo

Plan wykładu

Wstęp

Komputer - urządzenie precyzyjne, deterministyczne. Czy może służyć do produkowania liczb losowych?

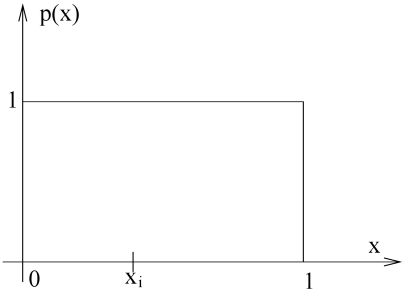
Historyczne rozróżnienie:
random (losowe) - uzyskiwane z procesów istotnie losowych (generatory fizyczne):
- detektor Geigera-Müllera,
- szum lamp elektronowych,
- ruletka...

quasirandom (pseudolosowe) - sekwencje generowane przez komputery (generatory programowe).

UWAGA: w symulacjach komputerowych potrzebujemy odtwarzalności, błędem jest rezygnacja z niej!

Liczby losowe o rozkładzie równomiernym (uniform deviate)

Podstawowy typ generatora liczb to ten dla liczb o rozkładzie równomiernym. $P(x \in [a, b]) = \int_a^b p(x)dx$



Wykres funkcji gęstości prawdopodobieństwa p(x)
$P(x_i \in [a, b]) = \int_a^b p(x)dx$

$\{x_i\}-$ ciąg liczb z przedziału $(0,1)$ –równomierny na $(0,1)$, gdy:

$$\forall(a,\ b)\ :\ 0 \leqslant a \leqslant b \leqslant 1\ \lim_{N\to\infty} \frac{1}{N}\sum_{i/1}^{N} \eta_{a,b}(x_i) = b - a$$

gdzie: $\eta_{a,b} = \begin{cases} 1, & a < x < b \\ 0, & \text{pozostale.} \end{cases}$

W większości bibliotek programów procedura *ranf*.

$$x = ranf(iseed)$$

iseed –dowolna, zadawana przy pierwszym wywołaniu;
- ta sama początkowa wartość $\Rightarrow$ ta sama sekwencja liczb pseudolosowych.

Generatory liczb równomiernych

Ogólnie

$$x_{n+1} = f(\underbrace{x_n, x_{n-1}, \ldots, x_{n-k+1}}_{k \text{ stalych poczatkowych}})(mod \quad M)$$

Założenie:

$$Z_M = \{0,\ 1,\ \ldots,\ M-1\}$$

$$\text{Dziedzina } D(f) = Z_M^{\otimes k}$$

$$\text{Przeciwdziedzina } D^{-1}(f) = Z_M$$

Takie generatory są *okresowymi*:

$$\exists N, r : \forall n \geqslant N \ x_n = x_{n+jr} \ , j = 1, 2, .$$

r - okres ciągu

$$\underbrace{x_0, \ldots, x_N, x_{N+1}, \ldots, x_{N+r-1}}_{\text{okres aperiodyczności ciągu}}, x_{N+r}, x_{N+1+r} \ldots, x_{N+2r}, x_{N+1+2r}$$

Generator Fibonacciego

$$x_{n+1} = (x_n + x_{n-1})(mod M)$$

- okres $\leqslant M^2$,
- prosty,
- wada: korelacje w ciągach generowanych liczb.

Generatory liniowe kongruentne

Większość generatorów to generatory *liniowe kongruentne*:

$$I_{j+1} = (aI_j + c) mod\ m$$

gdzie:

$$\left.\begin{aligned} a &- multiplier \\ c &- increment \\ m &- modulus \end{aligned}\right\} \text{liczby całkowite } \in [0, m]$$

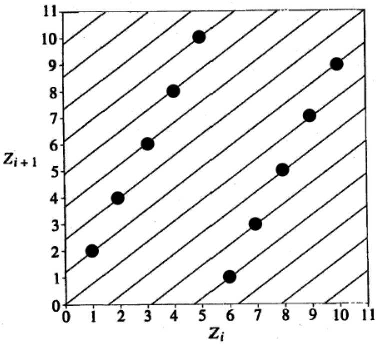
Liczby zmiennoprzecinkowe: $\dfrac{I_{j+1}}{m} \in [0,\ 1)$

sekwencja: $I_1, I_2, I_3, \ldots ; 0 \leqslant I_i \leqslant m - 1$
W końcu jakaś liczba musi się powtórzyć, a wtedy cały ciąg będzie się powtarzać
okres $\leqslant m$; zależy od wyboru a oraz c,
$c \neq 0 \;\longrightarrow$ generatory mieszane,
$c = 0 \;\longrightarrow$ generatory multiplikatywne.
Zaleta:
a) *szybkość generacji*

Wady:
a) *korelacje sekwencji*
- k -liczb losowych $\rightarrow$punkt w przestrzeni $k - D$,
- punkty nie zapełniają równomiernie przestrzeni lecz układają się na $(k-1) - D$ hiperpłaszczyznach.

Korelacje sekwencji

- przestrzen 2D,
- 2 podprzestrzenie (hiperpłaszczyzny) 1D
- punkty(I_i, I_{i+1})
- $I_{i+1} = (2 \cdot I_i) \mod 11$, $I_0 = 1$ (seed)



Kiedyś IBM wsławił się zdaniem: "*gwarantuje tylko losowość każdej liczby indywidualnie*"

b) *niższe bity są "mniej losowe"* niż *wyższe*:
- nie wykorzystywać liczb losowych w *kawałkach*,
- np. do generowania liczb losowych $\in [1, 10]$

używać:

$J = 1 + INT(10.0*\mathrm{RANF}\ (\mathrm{iseed})\ )$

a nie:

$J = 1 + MOD(INT(100000.0 * RANF(\mathrm{iseed})),\ 10)$

Zasady doboru stałych generatora liniowego kongruentnego

Pojawiające sie w literaturze wnioski:

I_0:

- nie ma większego znaczenia

a:

- $a(mod 8) = 5$
- $\frac{m}{100} < a < m - \sqrt{m}$
- brak powtarzającego się wzorca w zapisie w systemie dwójkowym

c:

- nieparzyste
- spełniające $\frac{c}{m} \approx \frac{1}{2} - \frac{\sqrt{3}}{6}$

m:

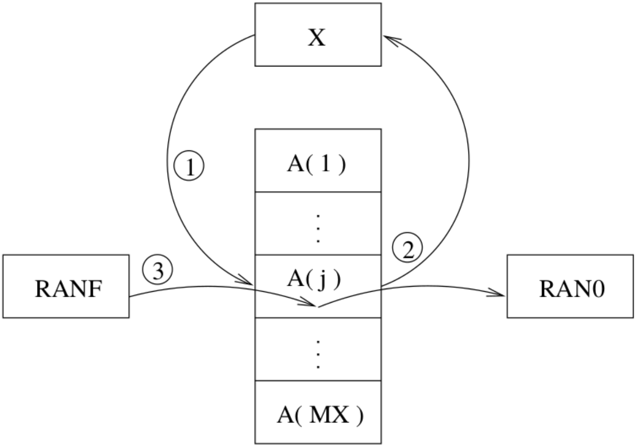
$m = 2^t$, t-liczba bitów przeznaczonych na 1 liczbę całkowitą

Zmniejszenie korelacji w sekwencji liczb losowych

- procedura "*losowego tasowania*" (random shuffling procedure)
Bays-Durham ⇒ *Knuth: "The Art of Computer Programing" vol. II.*

RANF - generator systemowy,
RANO - generator ulepszony
A - tablica pomocnicza o dlugosci wyznaczonej przez liczbę pierwszą)
Uogólnienie:
- kilka generatorów
- jeden z nich wybiera "dostarczyciela" liczb

Zmniejszenie korelacji w sekwencji liczb losowych



Rysunek 14.1: Idea ulepszonego gen. liczb losowych

Zmniejszenie korelacji w sekwencji liczb losowych

0. wypełniam tablicę A i zmienną x liczbami losowymi
1. x traktuje jak indeks j, który wskazuje na element tablicy A
2. $A(j)$ wstawiam w miejsce x oraz jednocześnie zwracam jako liczbę losową ulepszonego generatora
3. z generatora systemowego RANF losujemy brakującą liczbę w miejsce $A(j)$

Rozkład równomierny na (0,1) - dystrybuanta

Dystrybuanta jednoznacznie definiuje rozkład prawdopodobieństwa:

$$F(x) = \int_{-\infty}^{x} p(y)dy$$

Funkcja niemalejąca, określa $P(X \leqslant x)$
Dla rozkładu równomiernego na (0,1), dla $x \in (0, 1)$ $p(x) = 1$

$$F(x) = \int_{-\infty}^{x} p(y)dy = x$$

Czyli dla $x \in (0, 1)$

$$P(X \leqslant x) = x$$

Metoda odwróconej dystrybuanty

Jak uzyskać rozkład o zadanej dystrubuancie $F(x)$?
Jeśli zdefiniujemy U - zmienną losową o rozkładzie równomiernym na (0,1) to zmienna losowa

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

ma rozkład o dystrybuancie $F(x)$
Dowód:

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

Przykład rozkład wykładniczy e^{-x}

Funkcja gęstości prawdopodobieństwa: $p(x) = e^{-x}$ $x \in [0, \infty)$
Dystrybuanta:

$$F(x) = \int_{-\infty}^{x} e^{-x'} dx' = 1 - e^{-x}$$

$$y = 1 - e^{-x}$$

$$x = -ln(1 - y)$$

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

Generujemy ciąg liczb losowych o rozkładzie równomiernym U
$u_1, u_2, u_3, .., u_n \in (0, 1)$
Wtedy ciąg liczb: $y_i = -ln(1 - u_i)$ ma rozkład wykładniczy.
W ogólności odwracanie dystrybuanty sprawia często duże trudności numeryczne.

Rozkład normalny

$$p(x) = e^{-\frac{x^2}{2}} \quad ,$$

Szukanie odwrotności dystrybuanty(funkcja nieelementarna!):

$$F(x) = \int_{-\infty}^{x} e^{\frac{-y^2}{2}} dy' = erf(x)$$

jest kosztowne.
Zwykle stosuje się metodę Boxa-Mullera

Metoda Box-Muller

Biorę dwa niezależne rozkłady normalne i liczę prawdopodobieńswo łączne (iloczynu zdarzeń):

$$p(x_1, x_2) = e^{-\frac{x_1^2}{2}} \cdot e^{-\frac{x_2^2}{2}} = e^{-\frac{x_1^2+x_2^2}{2}}$$

$$x_1, x_2 \in (-\infty, \infty)$$

Wprowadzamy zmienne biegunowe:

$$r^2 = x_1^2 + x_2^2, r \in [0, \infty)$$

$$x_1 = r\sin(\phi), \phi \in [0, 2\pi]$$

$$x_2 = r\cos(\phi)$$

Metoda Box-Muller

Przeliczamy element prawdopodobieństwa (prawdopodobieństwo, że x i y znajdą się w małym obszarze dxdy) we współrzędnych biegunowych (r - moduł Jakobianu)

$$p(x, y)dxdy = p(r, \phi) r d\phi dr$$

$$e^{-\frac{r^2}{2}} r d\phi dr$$

Wprowadzam zmienną $z = \frac{r^2}{2}$

$$dz = rdr$$

$$e^{-z} d\phi dz$$

Metoda Box-Muller

Otrzymaliśmy rozkład wykładniczy $e^{-z} d\phi dz$.
Możemy zastosować odwrotną dystrubuantę:

$$F^{-1}(w) = -ln(1 - w)$$

oraz odwrotną funkcję do: $\frac{r^2}{2} = z$ czyli $r = \sqrt{2z}$

Dodatkowo gęstośc prawdopodobieństwa rozkładu $e^{\frac{r^2}{2}}$ nie zależy od $\phi \in [0, 2\pi]$, który losujemy zgodnie z rozkładem równomiernym na przedziale $(0, 2\pi)$

Metoda Box-Muller

Metoda Box-Muller generacji zmiennych losowych o rozkładzie normalnym

$$p(r)dr = e^{-\frac{r^2}{2}}dr = e^{-\frac{x_1^2+x_2^2}{2}}dx_1 dx_2 = e^{-\frac{x_1^2}{2}}dx_1 e^{-\frac{x_2^2}{2}}dx_2$$

transformacja:

U_1, U_2 - zm. losowe niezależne, rozkł. równomierny na $(0, 1)$

$$x_1 = rcos(\phi) = \sqrt{-2\ln(1-U_1)}\cos(2\pi U_2)$$

$$x_2 = rsin(\phi) = \sqrt{-2\ln(1-U_1)}\sin(2\pi U_2)$$

$\Rightarrow x_1 x_2$ - każda oddzielnie ma rozkład normalny (2 niezależne!)

Uproszczenie obliczeń

Zamiast:

U_1, U_2 - rozkład równomierny w jednostkowym kwadracie

bierzemy:

V_1, V_2 - współrz. punktu w jednostkowym kole:
$V_1^2 + V_2^2 < 1$,
zamiast U_1 bierzemy $R = V_1^2 + V_2^2$ (też ma rozkład równomierny)
zamiast U_2 - $\angle(V_1, V_2)$

$$\cos(2\pi U_2) = \frac{V_1}{\sqrt{R}}; \qquad \sin(2\pi U_2) = \frac{V_2}{\sqrt{R}}$$

i tak unikamy stosowania funkcji trygonometrycznych

Uproszczenie obliczeń

$$x_1 = V_1 \sqrt{\frac{-2\ln(V_1^2 + V_2^2)}{V_1^2 + V_2^2}},$$

$$x_2 = x_1 \cdot \frac{V_2}{V_1}$$

x_1, x_2 - niezależne, obie o $N(0,1)$

Kryptografia a liczby losowe

Materiały dostępne w internecie:

- Quantifying Studies of (Pseudo) Random Number Generation for Cryptography
  `http://www.tcs.hut.fi/Publications/arock/these_roeck.pdf`

Mersenne Twister

- jeden z lepszych generatorów używanych obecnie
- szybki
- dobre własności statystyczne
- wada: stosunkowo duża liczba instrukcji z których się składa

http://www.math.sci.hiroshima-u.ac.jp/m-mat/MT/ARTICLES/mt.pdf

Źródło:
N. E. Knuth, "The Art of Computer Programing vol. II, Addison-Wesley, 1969.
Wieczorkowski R., Zieliński R.: Komputerowe genaratory liczb losowych WNT 1997