Sprawozdanie - laboratorium nr 14
Generowanie ciągu liczb pseudolosowych o rozkładzie jednorodnym w kuli 3D.

Damian Płóciennik

5 czerwca 2019

1 Wstęp teoretyczny

1.1 Generatory liniowe

Generatory liniowe tworzą ciąg liczb według schematu:

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

gdzie $a_1, \dots, a_k, c$ nazywami parametrami generatora (są to ustalone liczby).
Operację

$$r = (a \mod n) \tag{2}$$

nazywamy dzieleniem modulo, której wynikiem jest reszta z dzielenia liczb całkowitych a i n.

Do poprawnego działania generatora konieczne jest zdefiniowanie parametrów, między innymi tak zwanego ziarna:

$$X_0, X_1, X_2, \dots, X_k, \tag{3}$$

które daje początek generatorowi, a otrzymane może być na przykład korzystając z innego generatora lub zegara systemowego.

$$X_0, X_1, X_2, \dots, X_k \tag{4}$$

Najprostsze generatory liniowe można podzielić na multiplikatywne dla $c = 0$ lub mieszane dla $c \neq 0$.

1.2 Generator multiplikatywny

Generatorem multiplikatywnym nazywa się generator liniowy dla $c = 0$, taki że:

$$X_{i+1} = aX_{i-1} \mod m, \tag{5}$$

$$k_i = \left\lfloor \frac{aX_{i-1}}{m} \right\rfloor, \quad i \geqslant 1, \tag{6}$$

$$\begin{aligned} X_1 &= aX_0 - mk_1, \\ X_2 &= a^2 X_0 - mk_2 - mk_1 a, \\ X_3 &= a^3 X_0 - mk_3 - mk_2 a - mk_1 a^2, \\ &\dots \\ X_n &= a^n X_0 - m(k_n + k_{n-1} a + \dots + k_a^{n-1}), \end{aligned} \tag{7}$$

co można zapisać w postaci:

$$X_n = a^n X_0 \mod m, \tag{8}$$

skąd wynika, że wybór X_0 determinuje wszystkie liczby w generowanym ciągu (a i m są ustalone) – uzyskany ciąg liczb jest deterministyczny.

Okres generatora multiplikatywnego określa wzór:

$$T = \min\{i : X_i = X_0, \quad i > 0\}. \tag{9}$$

Maksymalny możliwy okres takiego generatora można uzyskać dla:

$$a^{(m-1)/p} \neq 1 \mod m, \tag{10}$$

gdzie m jest liczbą pierwszą, p czynnikiem pierwszym liczby $(m-1)$. W praktyce wykorzystujemy więc często liczby Mersenne'a, które bardzo często okazują się być pierwsze ($m = 2^p - 1$).

Wadą takich generatorów jest nierównomierne pokrycie d-wymiarowej kostki, ponieważ generowane liczby lokalizują sie na hiperpłaszczyznach, których położenie uzależnione jest od parametrów generatora.

1.3 Metoda Boxa-Mullera

Zdefiniowano fgp w 2D jako funkcję gaussowską:

$$f(x, y) = f(x) \cdot f(y) = e^{-\frac{x^2+y^2}{2}}, \quad x, y \in (-\infty, \infty). \tag{11}$$

Celem jest policzenie prawdopodobieństwa:

$$p(x, y) = f(x, y)\mathrm{d}x\mathrm{d}y. \tag{12}$$

W tym celu wprowadzano nowe zmienne:

$$\begin{aligned}
r^2 &= x^2 + y^2, \\
x &= r\cos(\theta) \quad r \in [0, \infty), \\
y &= r\sin(\theta) \quad \theta \in [0, 2\pi], \\
p &= f(x, y)\mathrm{d}x\mathrm{d}y = f(r, \theta)r\mathrm{d}r\mathrm{d}\theta, \\
p(r, \theta) &= r \cdot e^{-r^2/2}\mathrm{d}r\mathrm{d}\theta, \\
z &= \frac{r^2}{2} \rightarrow \mathrm{d}z = r\mathrm{d}r \quad z \in [0, \infty), \\
p(z, \theta) &= e^{-z}\mathrm{d}z \cdot \mathrm{d}\theta.
\end{aligned} \tag{13}$$

Otrzymano rozkład wykładniczy:

$$\begin{aligned}
f(z) &= e^{-z}, \\
z &= -\ln(1 - U_1), \quad U_1 \in (0, 1), \\
r &= \sqrt{2z}.
\end{aligned} \tag{14}$$

Kąt θ ma rozkład jednorodny, więc użyto generatora o rozkładzie jednorodnym:

$$\theta = U_2 \cdot 2\pi, \quad U_2 \in (0, 1). \tag{15}$$

Dla pary (U_1, U_2) dostajemy (x, y) z rozkładu $N(0, 1)$:

$$x = r\cos(\theta) = \sqrt{-2\ln(1 - U_1)}\cos(2\pi U_2), \tag{16}$$

$$y = r\sin(\theta) = \sqrt{-2\ln(1 - U_1)}\sin(2\pi U_2). \tag{17}$$

2 Zadanie do wykonania

2.1 Opis problemu

Na początku przeanalizowano działanie trzech multiplikatywnych generatorów liczb pseudolosowych o rozkładzie jednorodnym. Dla dwóch pierwszych przypadków przyjęto:

$$X_i = a\, X_{i-1} \mod m, \tag{18}$$

których wyjście było normowane:

$$x_i = \frac{X_i}{m + 1.0}. \tag{19}$$

Przyjęto:

- $U_1(0,1) : a = 17,\ m = 2^{13} - 1,\ X_0 = 10,$
- $U_2(0,1) : a = 85,\ m = 2^{13} - 1,\ X_0 = 10.$

W ostatnim przypadku skorzystano z generatora multiplikatywnego opisanego wzorem:

$$X_i = (1176 \cdot X_{i-1} + 1476 \cdot X_{i-2} + 1776 \cdot X_{i-3}) \mod (2^{32} - 5) \tag{20}$$

dla przyjętego parametru startowego $X_0 = X_{-1} = X_{-2} = 10$.

Dla każdego z przypadków wylosowano 2000 liczb pseudolosowych i sporządzono dwuwymiarowe wykresy kolejnych par: $(x_i,\, x_{i+1})$, $(x_i,\, x_{i+2})$,$(x_i,\, x_{i+3})$.

Następnie przy pomocy metody Boxa-Mullera utworzono $N = 2000$ wektorów o rozkładzie normalnym, które znajdowały się na powierzchni sfery:

$$x_i = \sqrt{-2\ln(1 - u_1)}\cos(2\pi u_2), \tag{21}$$
$$y_i = \sqrt{-2\ln(1 - u_1)}\sin(2\pi u_2), \tag{22}$$
$$z_i = \sqrt{-2\ln(1 - u_3)}\cos(2\pi u_4), \tag{23}$$
$$u_1, u_2, u_3, u_4 \in U_3(0,1), \tag{24}$$

a następnie każdy z nich znormalizowano do jedynki z zastosowaniem normy euklidesowej:

$$\vec{r_i} \leftarrow \frac{\vec{r_i}}{||\vec{r_i}||_2}. \tag{25}$$

Z wykorzystaniem zmiennej o rozkładzie wielomianowym rozłożono punkty równomiernie w kuli, stosując:

$$u_i \in U_3(0,1), \tag{26}$$
$$s_i = (u_i)^{\frac{1}{d}}, \tag{27}$$
$$\vec{r_i} = [s_i \cdot x_i, s_i \cdot y_i, s_i \cdot z_i]. \tag{28}$$

Dla obu przypadków sporządzono wykresy 3D.

Na koniec sprawdzono, czy rozkład punktów w kuli jest jednorodny, tj. czy gęstość losowanych punktówjest stała w obszarze kul .W tym celu podzielono promień kuli na $K = 10$ podprzedziałów o równej długości, a następnie dla każdego punktu określono jego przynależność do konkretnego przedziału:

$$j = (int)\left(\frac{||\vec{r_i}||_2}{\Delta}\right) + 1, \quad j = 1, 2, \ldots, K \quad \Delta = \frac{1}{K}. \tag{29}$$

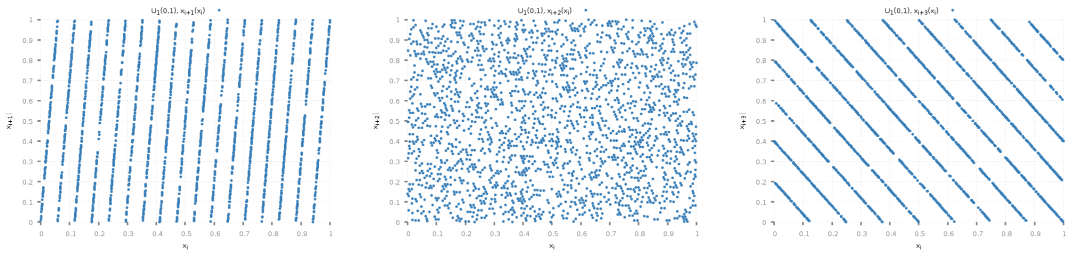
Objętość obliczano zgodnie ze wzorem:

$$R_j = \Delta \cdot j, \tag{30}$$
$$R_{j-1} = \Delta \cdot (j-1), \tag{31}$$
$$V_j = \frac{4}{3}\pi R_j^3, \tag{32}$$
$$V_{j-1} = \frac{4}{3}\pi R_{j-1}^3, \tag{33}$$
$$g_j = \frac{n_j}{V_j - V_{j-1}}. \tag{34}$$

Obliczenia wykonano kolejno dla $N = 2000, 10^4$ oraz 10^7 punktów.

2.2 Wyniki

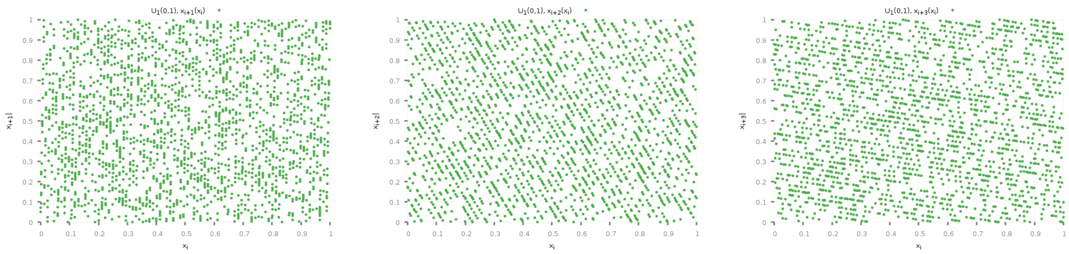
W celu wykonania zadania napisano program w języku C.



(a) Zależność $x_{i+1}(x_i)$ (b) Zależność $x_{i+2}(x_i)$ (c) Zależność $x_{i+3}(x_i)$

Rysunek 1: Zależność par kolejnych liczb pseudolosowych dla rozkładu jednorodnego $U_1(0,1)$

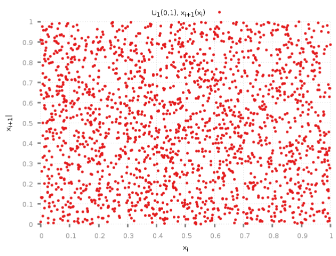
Jak wydać na powyższych wykresach pierwszy generator pozwolił na uzyskanie liczb pseudolosowych, jednak dość małe parametry m i a spowodowały dość dużą zależność par kolejnych liczb pseudolosowych, co widać szczególnie na wykresach (a) i (c). Łatwo dostrzec również występującą okresowość generatora. Punkty układające się w siatki mogą świadczyć o tym, że nie jest to dobry generator.



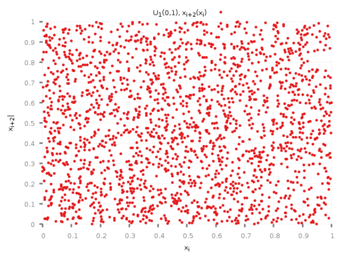
(a) Zależność $x_{i+1}(x_i)$ (b) Zależność $x_{i+2}(x_i)$ (c) Zależność $x_{i+3}(x_i)$

Rysunek 2: Zależność par kolejnych liczb pseudolosowych dla rozkładu jednorodnego $U_2(0,1)$

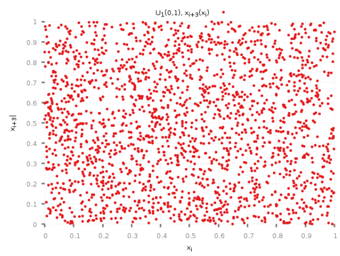
Generator, którego wyniki działania przedstawiono na powyższych wykresach, różni od poprzedniego wyłącznie zwiększonym parametrem a. Łatwo zauważyć, tak jak w poprzednim przypadku, silną korelację elementów oraz okresowość otrzymanych ciągów. W wyniku zmiany owego parametru punkty pokrywają teraz większą część płaszczyzny, jednak nadal takiego generatora nie można nazwać dobrym.



(a) Zależność $x_{i+1}(x_i)$



(b) Zależność $x_{i+2}(x_i)$

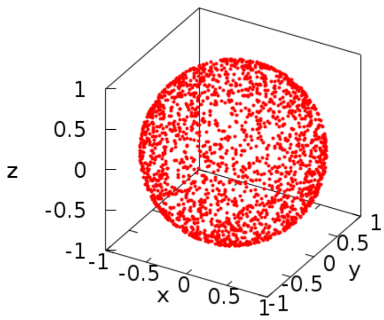


(c) Zależność $x_{i+3}(x_i)$

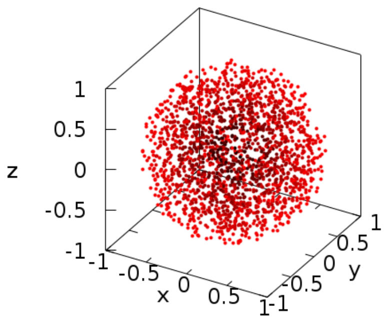
Rysunek 3: Zależność par kolejnych liczb pseudolosowych dla rozkładu jednorodnego $U_3(0,1)$

Ostatni z badanych generatorów jako jedyny bazował na zmienionym ziarnie, składającym się z trzech elementów. Ustawiono również dość duże wartości parametrów a_1, a_2, a_3 oraz zwiększono zdecydowanie czynnik związany z dzieleniem modulo. W tym przypadku na wykresach ciężko dostrzec jakąkolwiek korelację pomiędzy elementami, co świadczy o tym, że taki generator można uznać za dobry.

Kolejno przy pomocy metody Boxa-Mullera wylosowano 2000 punktów tworzących sferę wpisaną w sześcian, którego każda współrzędna mieściła się między -1 a 1, a następnie przy pomocy zmiennej o rozkładzie wielomianowym punkty rozłożono równomiernie wewnątrz kuli.



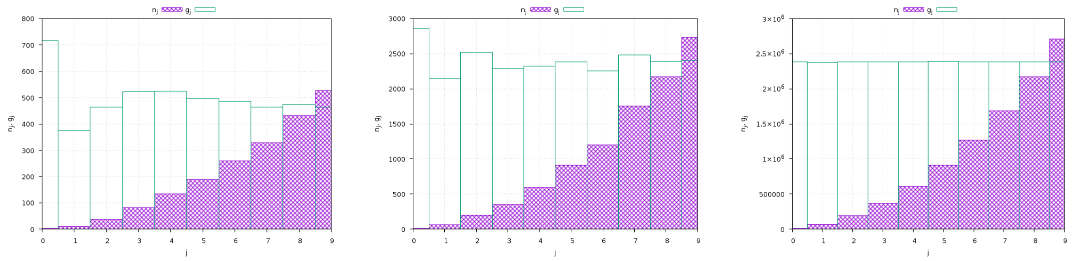
(a) Sfera



(b) Kula $K^3(0,1)$

Rysunek 4: Rozkład wylosowanych punktów w trzech wymiarach dla (b) kuli o promieniu $r = 1$ oraz (a) sfery wokół niej. Na wykresie (b): im ciemniejszy kolor punktu, tym bliżej środka układu współrzędnych $(0,0,0)$.

Następnie w celu stwierdzenia, czy rozkład punktów w kuli jest jednorodny (tj. czy gęstość losowanych punktów jest stała w obszarze kuli), podzielono promień kuli na $K = 10$ podprzedziałów o równej długości i dla każdego punktu określono jego przynależność do konkretnego przedziału. Efekty zostały zaprezentowane na poniższych histogramach.



(a) $N = 2000$ (b) $N = 10^4$ (c) $N = 10^7$

Rysunek 5: Histogramy dla rozkładu jednorodnego w trójwymiarowej kuli $K^3(0,1)$; n_j - liczba wylosowanych punktów znajdujących się w j-tym podzbiorze (warstwie kuli utworzonej przez równy podział względem promienia),g_j - gęstość wylosowanych punktów, tj. n_j dzielone przez objętość j-tego podzbioru kuli. Wykresy różnią się liczbą wygenerowanych punktów N.

Łatwo zauważyć, że dla dużej ilości wylosowanych punktów gęstość jest praktycznie stała. Dla $N = 2000$ widzimy już wyraźne odchylenia, zwłaszcza znacznie większą gęstość w pierwszym przedziale. Można zatem stwierdzić, że w przypadku dużej liczby losowań rozkład punktów w kuli jest jednorodny.

3 Wnioski

Zaletą generatorów liniowych jest przede wszystkim prostota, ale również szybkość jego działania i implementacji. Przy odpowiednim doborze parametrów można uzyskać rozkład, który faktycznie wygląda na losowy. Należy jednak nie zapominać, że liczby uzyskane w ten sposób nigdy nie będę w pełni losowe, a jedynie pseudolosowe. Ciągi generowanych w tym ćwiczeniu liczb układają się na hiperpłaszczyznach, aby tego uniknąć należałoby zastosować generatory nieliniowe, które mogą wykorzystywać na przykład odwrotność modulo.

4 Bibliografia

- Chwiej Tomasz: Generatory liczb pseudolosowych. [online]. [dostęp: 10.06.2019]. Dostęp w Internecie: `http://galaxy.agh.edu.pl/~chwiej/mn/generatory_1819.pdf`