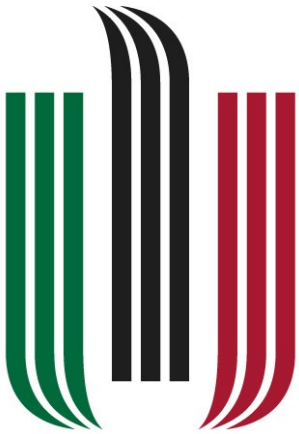


AGH

AKADEMIA GÓRNICZO-HUTNICZA
IM. STANISŁAWA STASZICA
W KRAKOWIE

Sprawozdanie - laboratorium nr 14

Generowanie ciągu liczb pseudolosowych o rozkładzie jednorodnym w kuli 3D

Klaudia Fil, 10.05.2019 r.

1. Wstęp teoretyczny

Generatory liniowe jest to jeden z rodzajów generatorów liczb losowych, które tworzą ciąg liczb według schematu:

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

gdzie $a_1, a_2, ..., a_k, c, m$ są parametrami generatora, z góry ustalonymi.

Operację:

$$r=a \bmod n \quad , \quad a, n, r \in Z \tag{2}$$

nazywa się dzieleniem modulo. Po jej przeprowadzeniu otrzymuję się resztę z dzielenia r liczb a przez n, które muszą być całkowite.

Generatory korzystające z dzielenia modulo nazywa się kongruentne lub kongruencyjne.

Aby wygenerować ciąg liczb pseudolosowych wymagana jest definicja parametrów generatora. Liczby $X_0, X_1, ... X_k$ nazywane są ziarnem generatora.

Jedną z odmian generatorów liniowych jest generator multiplikatywny, który istnieje gdy $c=0$, który możemy zdefiniować:

$$X_{i+1}=a X_{i-1} \bmod m \quad . \tag{3}$$

Wprowadzając oznaczenie:

$$k_i=\lfloor \frac{a X_{i-1}}{m} \rfloor, i \geq 1 \quad , \tag{4}$$

możemy rozwinąć:

$$\begin{aligned} X_1 &= aX_0 - mk_1 \\ X_2 &= a^2 X_0 - m k_2 - mk_1 a \\ &\vdots \\ X_n &= a^n X_0 - m(k_n + k_{n-1} a + ... + k_1 a^{n-1}) \end{aligned} \tag{5}$$

Ostatnie równanie można zapisać w krótszej postaci:

$$X_n = a^n X_0 \bmod m \quad , \tag{6}$$

wynika z niego, że X_0 determinuje wszystkie liczby w generowanym ciągu.

Okres generatora multiplikatywnego określa wzór:

$$T = min\{i : X_i = X_0, i > 0\} \quad . \tag{7}$$

Maksymalny okres takiego generatora uzyskamy, gdy:

$$a^{(m-1)/p} \neq 1 \bmod m \quad , \tag{8}$$

dla m – liczba pierwsza, p – czynnik pierwszy liczby m-1.

W praktyce wykorzystuje się liczby Mersenne'a $m=(2^p-1)$, które zazwyczaj okazują się pierwszymi.

Dla generatora o rozkładzie równomiernym w $x(0,1) \rightarrow U(0,1)$, przy założeniu, że liczby są od siebie niezależne, wartość oczekiwana powinna wynosić:

$$\mu = \int_0^1 x\,dx = \frac{1}{2} \quad , \tag{9}$$

$$\bar{\mu} = \frac{1}{N}\sum_{i=1}^{N} x_i \quad , \tag{10}$$

wariancja:

$$\sigma^2 = \int_0^1 (x-\mu)^2\,dx = \frac{1}{12} \quad , \tag{11}$$

$$\bar{\sigma} = \frac{1}{N}\sum_{i=1}^{N} (x_i-\bar{\mu})^2 \quad . \tag{12}$$

Wykorzystując powyższe wzory możemy opisać funkcję autokorelacji, która opisuje zależność elementów od poprzednich wyrazów:

$$R_r = \frac{1}{(N-r)\,\sigma^2}\sum_{i=0}^{N-r} (X_i-\mu)(X_{i+r}-\mu) \quad , \tag{13}$$

$$\bar{R}_r = \frac{E[(X_t-\mu)(X_s-\mu)]}{\sigma^2} \quad . \tag{14}$$

Metoda Boxa-Mullera wykorzystywana jest, gdy szukanie funkcji błędu:

$$F(x) = \int_{-\infty}^{x} e^{-x'^2}\,dx' = erf(x) \tag{15}$$

jest kosztowne.

Definiujemy fpg w 2D jako funkcję gaussowską:

$$f(x,y) = f(x)\cdot f(y) = e^{-\frac{x^2+y^2}{2}} \quad , \quad x,y \in (-\infty,\infty) \tag{16}$$

Ponieważ docelowo chcemy policzyć prawdopodobieństwo $p(x,y)=f(x,y)dxdy$, czyli że wylosowana liczba znajdzie się w obszarze dxdy, wprowadzamy nowe zmienne:

$$\begin{matrix} x=r\cos\theta \\ y=r\sin\theta \end{matrix} \quad , \quad \begin{matrix} r\geq 0 \\ \theta\in[0,2\pi] \end{matrix} \quad . \tag{17}$$

Dzięki radialnym zmiennym możemy określić nowy wzór na prawdopodobieństwo:

$$p=f(x,y)dxdy=f(r,\theta)r\,dr\,d\theta \quad , \tag{18}$$

Po kolejnych przekształceniach i separacji zmiennych:

$$p(r,\theta)=r\cdot e^{\frac{-r^2}{2}}\,dr\,d\theta \tag{19}$$

Wprowadzając kolejną zmienną:

$$z=\frac{r^2}{2}\rightarrow dz=rdr\,,z\leq 0 \quad , \tag{20}$$

która ułatwia nam całkowanie:

$$p(z,\theta)=e^{-z}\,dz\,d\theta=f(z)dz\,d\theta \quad . \tag{21}$$

Otrzymany rozkłada jest wykładniczy, a więc bez problemu możemy go całkować:

$$f(z)=e^{-z}\rightarrow z=-\ln(1-U_1) \quad , \quad U_1\in(0,1) \quad . \tag{22}$$

Podstawiając $r=\sqrt{2z}$ oraz wiedząc, że kąt ma rozkład jednorodny, dla pary (U_1,U_2) dostajemy rozwiązanie dla rozkładu $N(0,1)$:

$$\begin{matrix} x=r\cos\theta=\sqrt{-2\ln(1-U_1)}\cos(2\pi U_2) \\ y=r\sin\theta=\sqrt{-2\ln(1-U_1)}\sin(2\pi U_2) \end{matrix} \quad . \tag{23}$$

2. Zadanie do wykonania

2.1. Opis problemu

Zadanie było podzielone na dwie części, w pierwszej z nich należało stworzyć 3 generatory multiplikatywne liczb pseudolosowych zadanych wzorami:

$$\begin{matrix} U_1(0,1)=(17\cdot X_{i-1})\,mod\,(2^{13}-1) \\ U_2(0,1)=(85\cdot X_{i-1})\,mod\,(2^{13}-1) \\ U_3(0,1)=(1176\cdot X_{i-1}+1476\,X_{i-2}+1776\,X_{i-3})\,mod\,(2^{32}-5) \end{matrix} \quad , \tag{24}$$

za parametry początkowe dla U_1, U_2 należało przyjąć $X_0 = 10$, dla U_3 - $X_0 = X_{-1} = X_{-2} = 10$. Wylosować N = 2000 liczb oraz stworzyć wykresy dla każdego z nich w funkcjach : $x_{i+1}(x_i)$, $x_{i+2}(x_i)$, $x_{i+3}(x_i)$.

W kolejnej części wykorzystując metodę Boxa-Mullera należało utworzyć N trójwymiarowych wektorów $\vec{r}_i = [x_i, y_i, z_i]$ o rozkładzie normalnym, stosując wzory:

$$\begin{aligned} &u_1 \in U_3(0,1) \\ &u_2 \in U_3(0,1) \\ &x_i = \sqrt{-2\ln(1-u_1)}\cos(2\pi u_2) \\ &y_i = \sqrt{-2\ln(1-u_1)}\sin(2\pi u_2) \quad , \\ &u_3 \in U_3(0,1) \\ &u_4 \in U_3(0,1) \\ &x_i = \sqrt{-2\ln(1-u_3)}\cos(2\pi u_4) \end{aligned} \tag{25}$$

znormalizować stosując normę euklidesową i sporządzić ich wykres.

Dysponując poprzednim układem jednorodnym na sferze, dla każdego punktu generujemy nową zmienną o rozkładzie: $h_d(s) = d \cdot s^{d+1}$, gdzie d to liczba wymiarów, więc w naszym przypadku d = 3, a $s \in (0,1)$. Wykorzystujemy do tego algorytm:

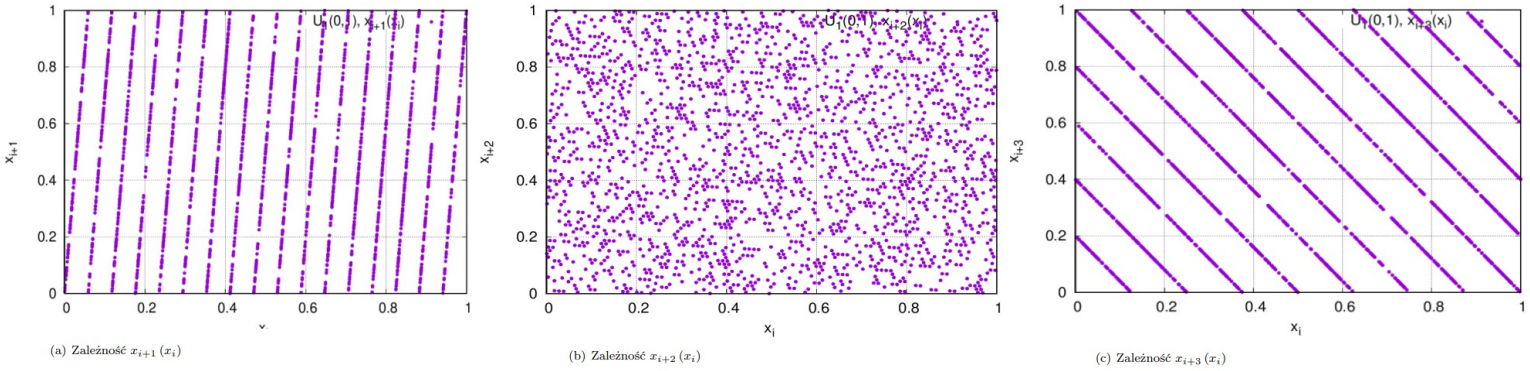
$$\begin{aligned} &u_i \in U_3(0,1) \\ &s_i = (u_i)^{\frac{1}{d}} \quad , \\ &\vec{r}_i = [s_i \cdot x_i, s_i \cdot y_i, z_i \cdot s_i] \end{aligned} \tag{26}$$

oraz ponownie sporządzamy wykres.

Na koniec sprawdzamy, czy rozkład punktów w kuli jest jednorodny tj. czy gęstość losowanych punktów jest stała w obszarze kuli. Dzielimy promień na 10 podprzedziałów, a następnie dla każdego określamy przynależność. Określamy również jego gęstość, czyli ilość liczb wpadająca w podprzedział w stosunku do jego objętości i generujemy histogram dla trzech przypadków $N = 2000, 10^4, 10^7$.

2.2. Wyniki

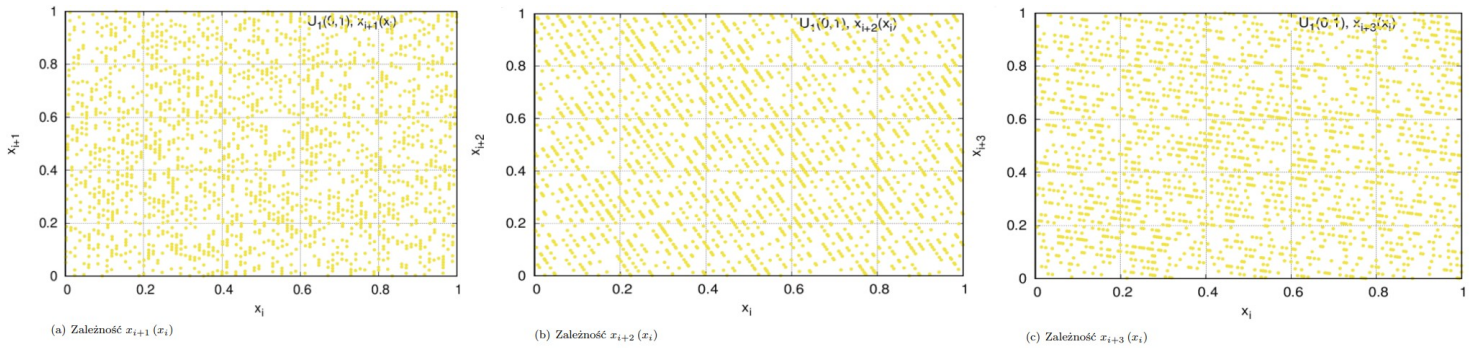
Dzięki programowi w języku C wykorzystującemu bibliotekę numeryczną Numerical Recipes, oraz przy pomocy skryptu Gnuplota wygenerowano:



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

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

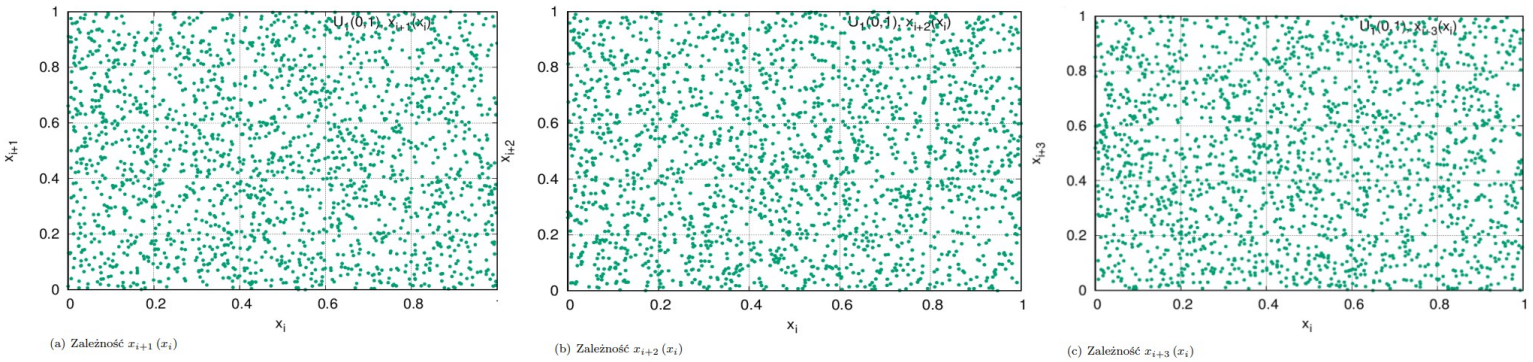
Jak widać na powyższym wykresie, generator $U_1(0,1)$ tworzy pseudolosowe liczby, jednak ponieważ jego parametry były stosunkowo niskie, co wpływa na duża autokorelację względem poprzedniego punktu, punkty na rys 1. układają się w siatki, co oznacza, że dość łatwo jest je przewidzieć.



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

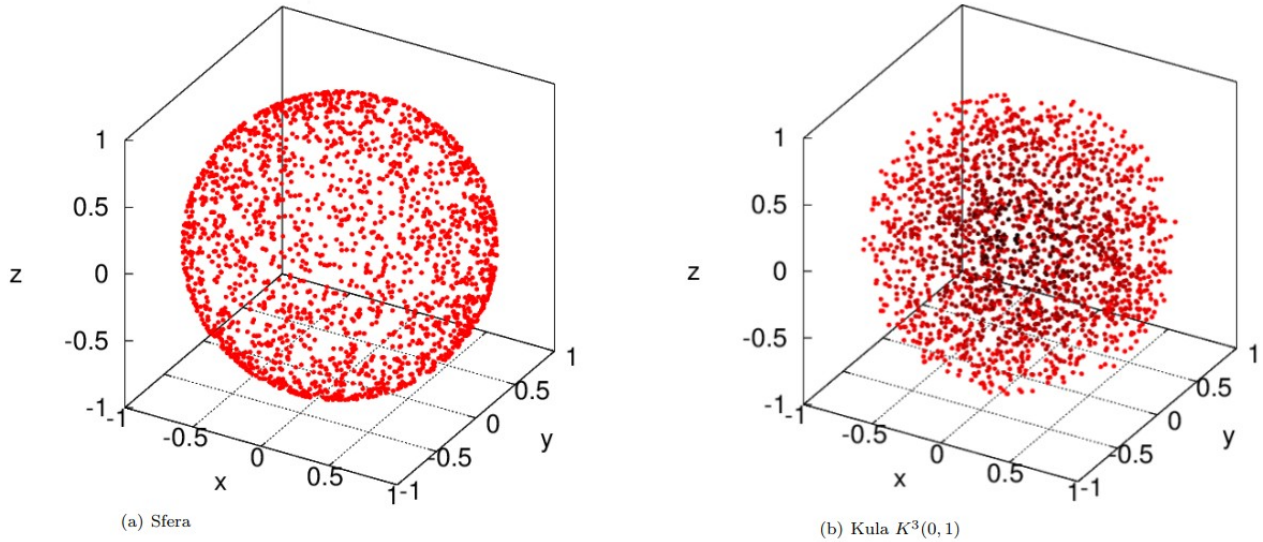
Rys. 2: Zależność par kolejnych liczb pseudolosowych dla rozkładu jednorodnego $U_2(0,1)$

Po zwiększeniu *a* możemy zauważyć, że punkty pokrywają znacznie większą część płaszczyzny co jest bardziej pożądane, jednak nadal pozostają okresowe. Współczynnik korelacji wciąż nie został zminimalizowany.



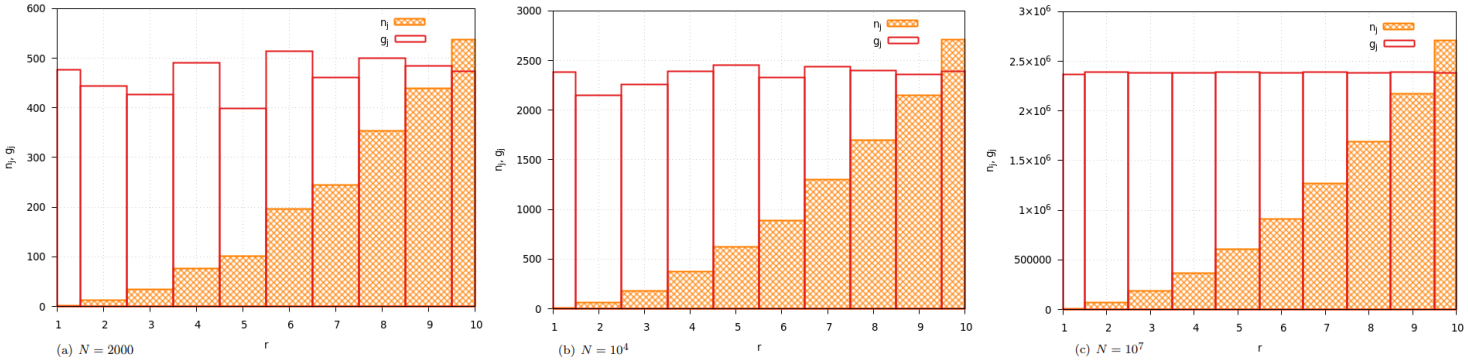
Rys. 3: Zależność par kolejnych liczb pseudolosowych dla rozkładu jednorodnego $U_3(0,1)$

Jak można było się spodziewać, generator trzeci tworzy najlepsze wyniki. Dopiero wprowadzenie zależności od trzech poprzednich elementów oraz ustawienie wysokich a_1, a_2, a_3 pozwoliło na pozbycie się widocznych na pierwszy rzut oka siatek, a więc minimalizację współczynnika korelacji.



Rys. 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).

Kolejnym wynikiem była wygenerowana na rys. 4.a) sfera, którą tworzy 2000 punktów, którą następnie przekształcono zgodnie z wzorem (26), dzięki czemu punkty rozłożono równomiernie wzdłuż promienia i przedstawiono wizualizację na rys. 4.b).



Rys. 5: Histogramy dla rozkładu jednorodnego w trójwymiarowej kuli K3 (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.

Kolejnym krokiem było sprawdzenie, czy rzeczywiście punkty zostały równomiernie rozłożone w kuli. Do sprawdzenia podzielono promień na 10 równych odcinków i obliczono ilość w każdej warstwie, oraz gęstość i zilustrowano to na histogramie rys. 5. dla trzech przypadków. Jak można było się spodziewać, dla wykresu ilości punktów na przedział powinniśmy zaobserwować wzsort dla każdej kolejnej wartości, spowodowane jest to większą objętością powłok licząc od środka, ale za to wykres gęstości powinien być stały, co w każdym przypadku w przybliżeniu udało się osiągnąć. Dla najmniejszego N widać największe rozbieżności, chociaż wartości oscylują wokół właściwej. Wraz ze wzrostem N sytuacja ulega poprawie. Możemy wysnuć wniosek, że dla tego generatora wymagana jest większa ilość liczb losowych w celu jego ujednorodnienia.

3. Wnioski

Generatory liczb pseudolosowych o rozkładzie jednorodnym są bardzo łatwe w implementacji, działają relatywnie szybko, chociaż dla ostatniej sytuacji, gdy $N=10^7$, trzeba było odczekać parę sekund dłużej. Należy jednak pamiętać o odpowienim doborze parametrów tj. uzależnieniem nowej zmiennej od kilku poprzednich, stosunkowo dużych liczbach *a*, oraz *m*, które powinny być liczbami pierwszymi.