Sprawozdanie

Metody numeryczne

Laboratorium 14.

Generowanie ciągu liczb pseudolosowych o rozkładzie jednorodnym i trójkątnym.

18.06.2020 r.

Aleksandra Rolka

1. Wstęp teoretyczny

Generatory liniowe

Generatory liniowe tworzą ciąg liczb według schematu:

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

gdzie:

$a_1, a_2, \ldots, a_k, c, m$ –parametry generatora(ustalone liczby)

Operację

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

nazywa się dzieleniem modulo, a jej wynikiem jest reszta z dzielenia liczb całkowitych a i n.

Generator wykorzystujący operacje dzielenia modulo to tzw. generator kongruentny lub kongruencyjny.

Aby wygenerować ciąg liczb pseudolosowych należy zdefiniować jego parametry.
Liczby

$$X_0, X_{1,}, X_2, \ldots, X_k \tag{3}$$

nazywa się ziarnem generatora(z ang. *seed*).

Dla bardziej rozbudowanych generatorów liczby te otrzymuje się z innego generatora lub np. używając zegara systemowego (X_0).

Najprostszy generator liniowy ma dwie odmiany

- generator multiplikatywny, gdy $c = 0$
- generator mieszany, dla $c \neq 0$

Rozkład trójkątny

Funkcję gęstości prawdopodobieństwa dla rozkładu trójkątnego $T(\mu, \Delta)$ (rys.1) definiuje się następująco:

$$f(x\colon \mu, \Delta) = -\frac{|x-\mu|}{\Delta^2} + \frac{1}{\Delta} \tag{4}$$

gdzie: μ to środek rozkładu, a Δ to jego szerokość.

Dystrybuanta rozkładu trójkątnego:

$$F(a) = P(x < a) = \int_{\mu-\Delta}^{a} f(x\colon \mu, \Delta) dx = \begin{cases} -\frac{1}{\Delta^2}\left(-\frac{x^2}{2} + \mu x\right) + \frac{x}{\Delta}, & x \leq \mu \\ -\frac{1}{\Delta^2}\left(\frac{x^2}{2} - \mu x + \mu^2\right) + \frac{x}{\Delta}, & x > \mu \end{cases} \tag{5}$$

Jeśli $\xi_1 \in U(0,1)$ i $\xi_2 \in U(0,1)$ to zmienną o rozkładzie trójkątnym oraz parametrach μ i Δ generuje się stosując formułę:

$$x = \mu + (\xi_1 + \xi_1 - 1) \cdot \Delta \tag{6}$$

2. Zadanie do wykonania

2.1 Opis problemu

I) **rozkład jednorodny**

Startując od $x_0 = 10$ należało wygenerować $n = 10^4$ liczb pseudolosowych przy użyciu generatora mieszanego:

$$x_{n+1} = (ax_n + c) mod\ m \tag{7}$$

o odpowiednich parametrach dla każdej wersji:

a) $a = 123,\ c = 1,\ m = \ 2^{15}$,
$b)$ $a = 69069,\ c = 1,\ m = \ 2^{32}$.

Dla każdego przypadku należało sporządzić rysunek$X_{i+1} \ = \ f(X_i)$ $(X_i = x_i/(m \ + \ 1.0)$, z warunku normalizacji do rozkładu $U(0,1)$ -> całkowite pole pod funkcją gęstości prawdopodobieństwa wynosi 1.

Należało również na samym początku napisać funkcję, w której zmienna (static) $x = 10$, będzie inicjalizowana tylko podczas pierwszego wywołania, a jej aktualna wartość będzie zachowywana w pamięci po zakończeniu działania funkcji. Szablon, wzorzec funkcji do utworzenia, z którego można było skorzystać, jednak lepszą alternatywą była forma funkcji z przekazywanymi parametrami *a, c ,m* :

```
double gen_1() {
    static long int x=10;
    int a=...;
    int c=...;
    long int m=...;
    x=(a*x+c) % m;
     return x/(m+1.0);
}
```

II) **rozkład trójkątny**

Następnie należało wygenerować $n = 10^3$ liczb o rozkładzie trójkątnym, zgodnie ze wzorem (6) o parametrach $\mu = 4,\ \Delta = 3$.

Kolejno podzielić przedział $[\mu - \Delta, \mu + \Delta]$ na K=1- podprzedziałów i zliczyć ile liczb wpada do każdego z nich.

Dla rozkładu trójkątnego przeprowadzić test X^2, tj. określić wartość statystyki testowej:

$$X^2 = \sum_{i=1}^{K} \frac{(n_i - n \cdot p_i)^2}{n \cdot p_i} \tag{8}$$

gdzie:

n_i – ilość liczb znajdujących się w podprzedziale o indeksie i,
p_i – prawdopodobieństwo teoretyczne, że zmienna losowa X znajduje się w i-tym podprzedziale

$$p_i = F(x_{i,max}) - F(x_{i,min})$$

$F(x)$ jest wartością dystrybuanty liczonej zgodnie z wzorem (5).

<u>Na koniec należało poddać testowaniu hipotezę H_0:</u> wygenerowany rozkład jest rozkładem $T(\mu, \Delta)$ wobec H_1, że nie jest to prawdą. Korzystając z odpowiednich tabel statystycznych sprawdzić czy hipoteza jest prawdziwa na poziomie istotności $\alpha = 0.05$ (α jest prawdopodobieństwem pierwszego rodzaju, czyli prawdopodobieństwem odrzucenia hipotezy H_0, gdy ta jest prawdziwa). W tym celu definiuje się obszar krytyczny testu:

$$\phi = \{X : \chi^2(X) > \epsilon\} \tag{9}$$

gdzie:

$X = \{x_1, x_2, \ldots, x_n\}$ jest ciągiem liczb pseudolosowych,
$\chi^2(X)$ wartością statystyki dla danego ciągu X,
ε jest poziomem krytycznym danego rozkładu dla określonej liczby stopni swobody i założonego poziomu istotności (odczytać z tabel statystycznych).

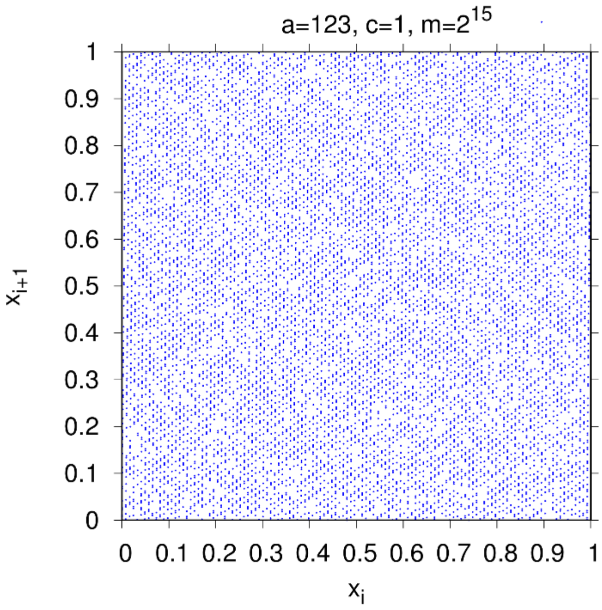
Liczbę stopni swobody określa się jako $\nu = K - r - 1$, gdzie K jest liczbą podprzedziałów, a $r = 2$ jest liczbą parametrów testowanego rozkładu (μ i Δ). Jeśli $\chi^2 < \epsilon$ to stwierdza się, że dla zadanego poziomu istotności nie ma podstaw do odrzucenia hipotezy H_0.

2.2 Wyniki

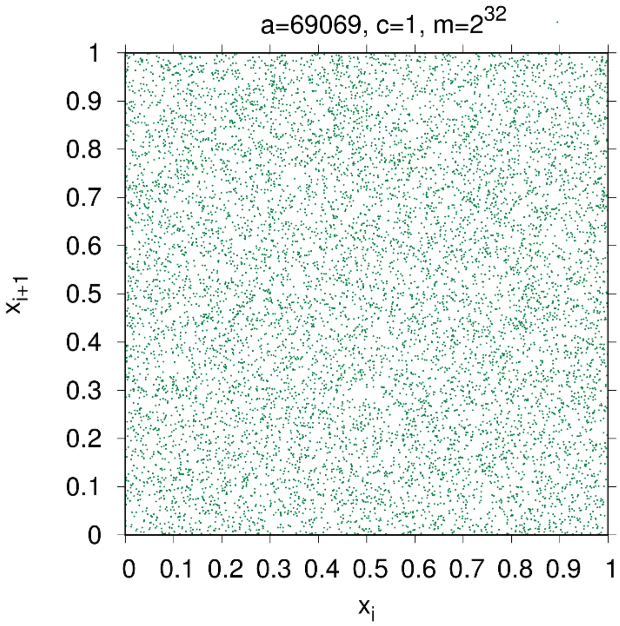
Napisany w języku C program rozwiązujący powyższy problem nie korzysta z żadnej dodatkowej biblioteki numerycznej. Odpowiednie dane zapisane zostały w plikach **.dat*:

- *U.dat* - zawiera dwie kolumny danych: x_i, x_{i+1} dla rozkladu jednorodnego *U(0,1)* w dwóch seriach dla obu wersji parametrów z podpunktów a) i b) punktu 2.1 I).
- *U_hist.dat* - zawiera dwie kolumny dla rozkładu jednorodnego: współrzędna środka *j*-tego podprzedziału, w dwóch seriach j.w.
- *T_hist.dat* - zawiera trzy kolumny dla rozkładu trójkątnego: współrzędna środka *j*-tego podprzedziału, n_j/N, p_j.

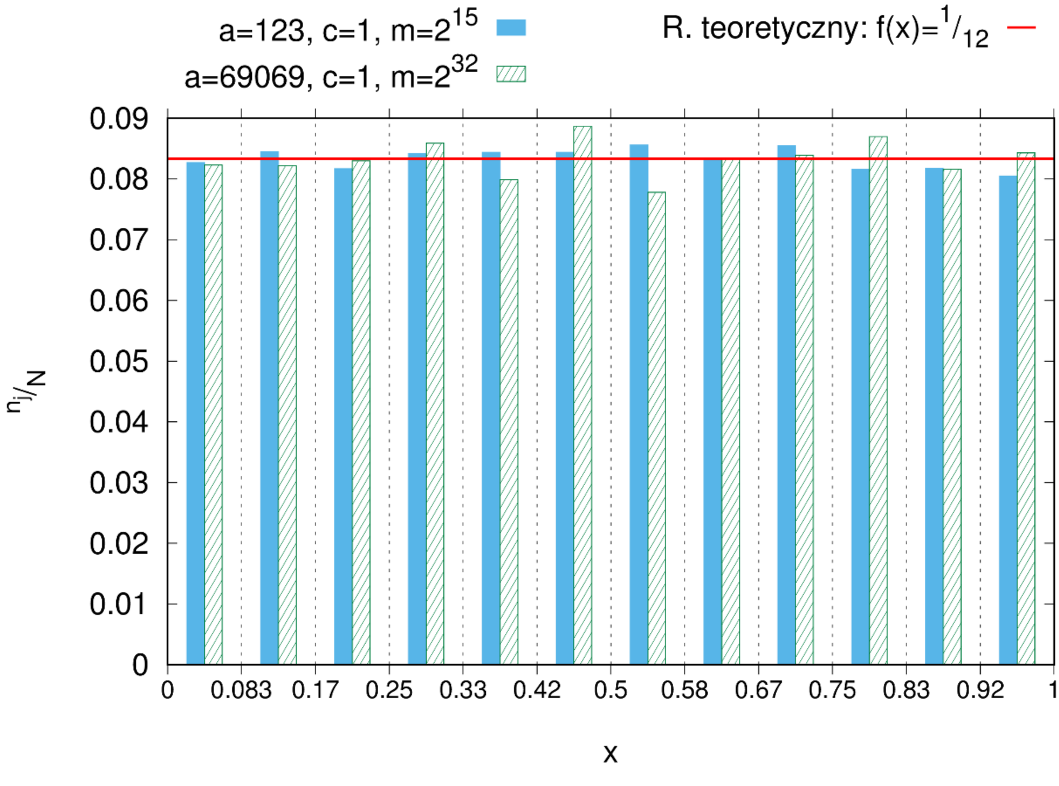
po to, aby na ich podstawie wygenerować poniższe wykresy:



(a) Zależność x_{i+1} (x_i) według generatora mieszanego o parametrach a = 123, c = 1, m = 2^{15}



(b) Zależność x_{i+1} (x_i) według generatora mieszanego o parametrach a = 69069, c = 1, m = 2^{32}

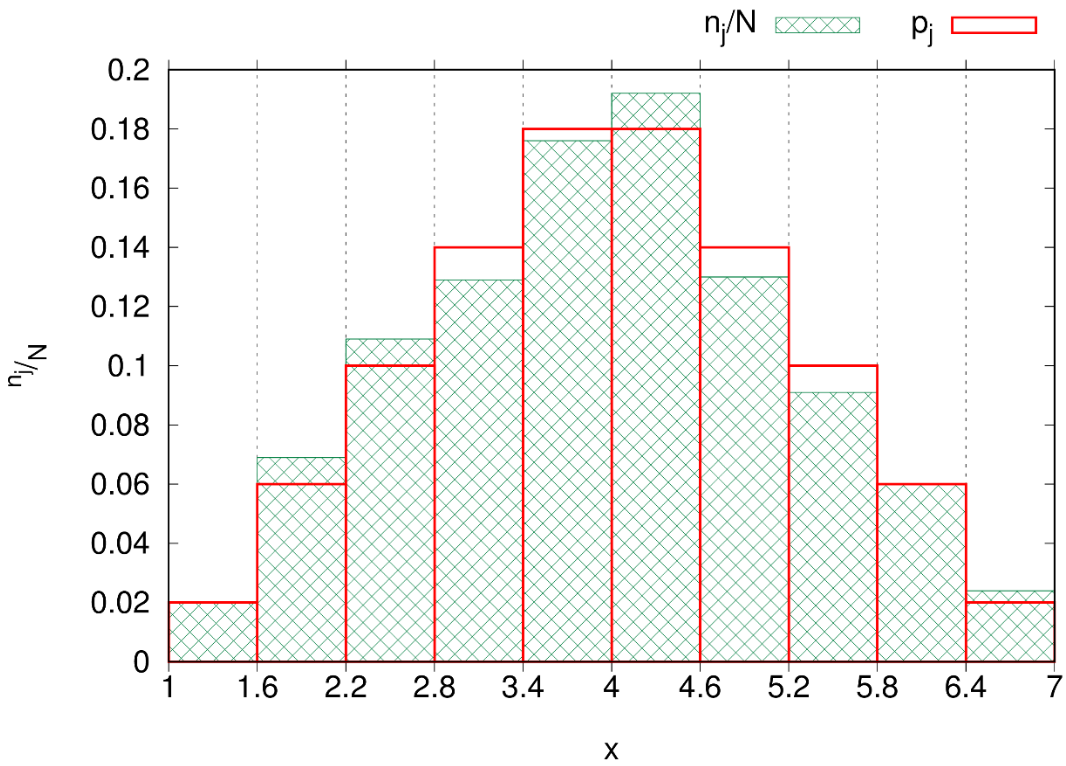


(c) Histogram dla rozkładów pochodzących z obu przypadków generatora mieszanego.

Rys. 1. Rozkłady jednorodne U(0, 1) uzyskane przy użyciu generatora mieszanego dla N = 10^4 dla dwóch przypadków parametrów a, c, m. Punkt startowy: X_0 = 10.

a	*c*	*m*	*Średnia μ*	*Odchylenie standardowe σ*
123	1	2^{15}	0.498215	0.287120
69069	1	2^{32}	0.501622	0.288302

Tabela 1. Średnia μ i odchylenie standardowe σ obliczone dla rozkładów jednorodnych U(0, 1) uzyskanych przy użyciu generatora mieszanego dla n = 10^4 liczb. Wartości teoretyczne: $\mu_0 = 0.5$, $\sigma_0 = \sqrt{\frac{1}{12}}$.



Rys. 2. Histogram dla wygenerowanego rozkładu trójkątnego T(μ = 4, Δ = 3). n_j/n – rozkład wylosowanych liczb w j-tym podprzedziale; p_j – teoretyczne prawdopodobieństwo wylosowania liczby w j-tym podprzedziale.

Statystyka testowa χ^2	6.237460

Tabela 2. Statystyka testowa χ^2obliczona dla rozkładu trójkątnego T(μ = 4, Δ = 3).

Odpowiednia wartość ϵ odczytana z tablic statystycznych dla testu χ^2:

$$\epsilon = 14{,}06.$$

3. Wnioski

Rysunek 1. pokazał, że dla utworzonego generatora mieszanego przy mniejszych wartościach parametrów widać pewne korelacje między elementami (Rys. 1a). Przy zwiększeniu tych wartości otrzymana została lepsza losowość, ale kosztem tego, że histogram rozkładu ma większe odchylenia w stosunku do teoretycznego.
Również wartości odchylenia standardowego i średniej (Tabela 1.) okazały się dokładniejsze dla 2. wersji.

W przypadku generatora o rozkładzie trójkątnym, na histogramie (Rys. 2.) widać, że udało się uzyskać w uogólnieniu żądany rozkład, jednak również nie pokrywa się on z teoretycznym.
Aby sprawdzić postawioną hipotezę, że jest to rozkład trójkątny przeprowadzono test chi-kwadrat. Ponieważ obliczona statystyka testowa χ^2 wyniosła 6.237460, a odczytana odpowiednia wartość ϵ z tablic statystycznych wynosi 14.06, hipoteza nie została odrzucona na poziomie istotności $\alpha = 0.05$.

Stwierdzono, że utworzone generatory dają zadowalające wyniki, ale trzeba pamiętać, że nie otrzymuje się z ich pomocą rzeczywiście losowych liczb, a jedynie pseudolosowe. Jednak dużą zaletą generatorów liniowych jest prostota implementacji i szybkość działania.

Źródła:

- Dr hab inż. Tomasz Chwiej – Notatki do wykładu *„Generatory liczb pseudolosowych”*. Dostępne w Internecie: http://home.agh.edu.pl/~chwiej/mn/generatory_1819.pdf