AKADEMIA GÓRNICZO-HUTNICZA
IM. STANISŁAWA STASZICA
W KRAKOWIE

Metody Numeryczne Laboratoria

Sprawozdanie nr 14

Generowanie ciągu liczb pseudolosowych o rozkładzie normalnym metodą eliminacji.

Łukasz Wajda, 10.06.2021 r.

1. Wstęp teoretyczny

Metoda eliminacji pozwala wygenerować ciąg liczb pseudolosowych x_{i} o zadanej gęstości prawdopodobieństwa f w przedziale [a,b]. Wartość f jest ograniczona od góry przez stałą d w przedziale [a,b].

Aby otrzymać ciąg zmiennych losowych o rozkładzie f(x) postępujemy następująco:

- Losujemy dwie zmienne o rozkładzie równomiernym: $\mathrm{U}_1 \in [\mathrm{a}, \mathrm{b}]$ $\mathrm{U}_2 \in [0, \mathrm{d}]$
- jeżeli: $\mathrm{U}_2 \leq \mathrm{f}(\mathrm{U}_1) \Rightarrow \mathrm{X} = \mathrm{U}_1$
- jeśli powyższy warunek nie jest spełniony wówczas odrzucamy parę $\mathrm{U}_1, \mathrm{U}_2$
- wykonujemy powyższe czynności aż do uzyskania ciągu o zadanej długości

Zmienna losowa X, wygenerowana przez nas, ma rozkład prawdopodobieństwa f.

Wszystkie generatory ciągów liczb bazują na wykorzystaniu ciągów liczb losowych o rozkładzie równomiernym, więc badamy generatory liczb o takim właśnie rozkładzie.

- dla ustalonej liczby n, generujemy n kolejnych liczb startując od losowo wybranej liczby początkowej,
- obliczamy wartość statystyki testowej (T),
- obliczamy dystrybuantę statystyki T - F(T) - gdy weryfikowana hipoteza jest prawdziwa

powyższe kroki powtarzamy N razy obliczając: $\mathrm{T}_1, \mathrm{T}_2, \ldots, \mathrm{T}_\mathrm{n}$

Jeżeli weryfikowana hipoteza jest prawdziwa to: $\mathrm{F}(\mathrm{T}_1), \mathrm{F}(\mathrm{T}_2), \ldots, \mathrm{F}(\mathrm{T}_\mathrm{N})$ jest ciągiem zmiennych niezależnych o rozkładzie równomiernym. Testowanie generatora kończy się sprawdzeniem tej hipotezy.

2. Zadanie do wykonania

Rozkład jednorodny U(0, 1), gdzie startując od $X_0 = 10$ wygenerowane zostało $n = 10^n$ liczb pseudolosowych liniowym generatorem kongruentnym (mieszanym):

$$X_{i+1} = (aX_i + c) mod\ m. \tag{1}$$

Jako parametry (typu long/unsigned long) przyjęliśmy:

a) $a = 123, c = 1, m = 2^{15}$,

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

Każdy otrzymany wynik X_i został znormalizowany do rozkładu (0,1), później zapisany jako x_i. Następnym krokiem było wyznaczenie średniej

$$\mu = \frac{1}{N}\sum_{i=0}^{N-1} x_i \tag{2}$$

oraz odchylenia standardowego

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=0}^{N-1} (x_i - \mu)^2}. \tag{3}$$

Aby wykonać histogram rozkładu gęstości prawdopodobieństwa, dokonany został podział przedziału losowanych wartości $[x_{min}, x_{max}] = [0, 1]$ na k = 12 podprzedziałów, każdy o szerokości $\Delta = \frac{x_{max} - x_{min}}{k}$. W tablicy o k elementach przyjęto oznaczenia: n_j - ilość trafień w przedział o numerze j. Tablica została uzupełniona w następujący sposób:

(1) Wyznaczany został indeks podprzedziału $j = \frac{x_i - x_{min}}{\Delta}$.

(2) Zwiększany o 1 został element n_j .

Granicami j-tego przedziału były:

$[x_{j,min}, x_{j,max}] = [x_{min} + j\Delta, x_{min} + (j + 1)\Delta]$.

Rozkład normalny N(μ, σ), gdzie wykorzystując generator mieszany o parametrach (2) z poprzedniego podpunktu, wygenerowano $n = 10^n$ liczb pseudolosowych o rozkładzie normalnym z parametrami $\mu_0 = 0.2$, $\sigma_0 = 0.5$ metodą eliminacji. Liczby pseudolosowe zawierały się w przedziale $x \in [\mu - 3\sigma, \mu + 3\sigma]$. Gęstość prawdopodobieństwa dana była funkcją:

$$f(y) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right) \tag{4}$$

Wartości u1, u2 znajdujące się poniżej wykresu f(x) zostały wygenerowane z użyciem generatora o parametrach (2) z poprzedniego podpunktu, po czym zostały odpowiedni przeskalowane. Ponadto, wartość d została przyjęta jako wartość 1.

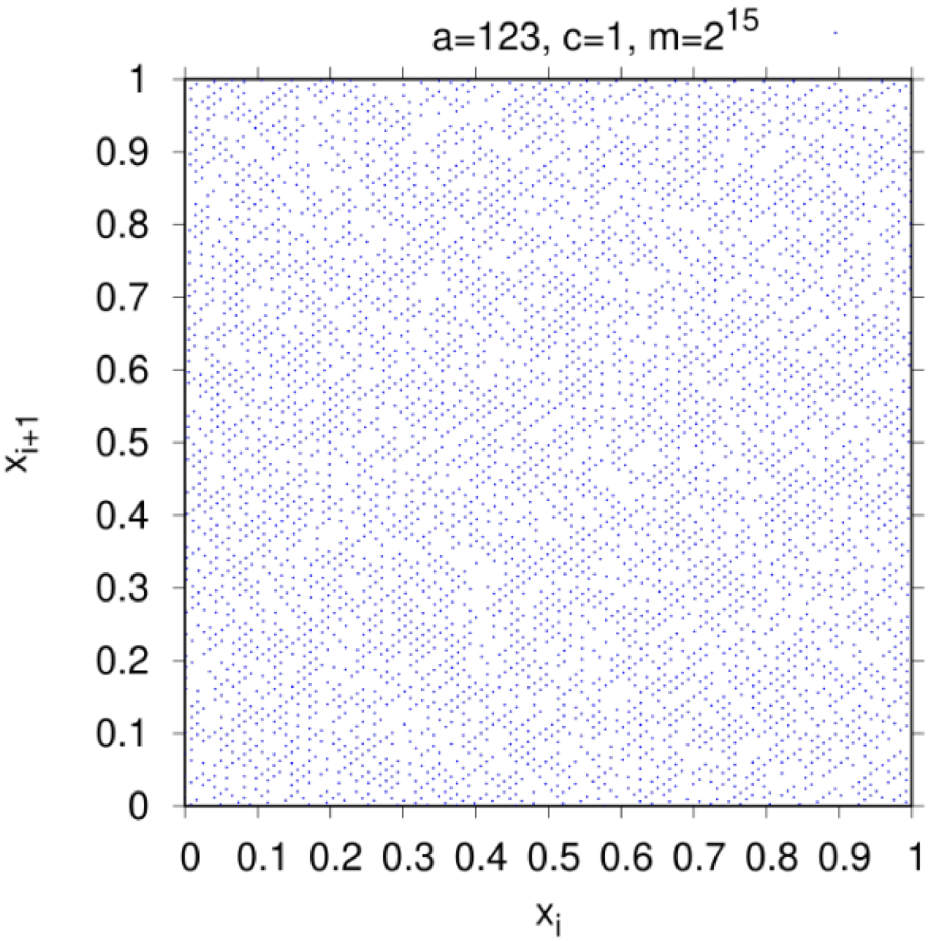
Test χ^2 generatora o rozkładzie N(μ, σ), który przebiegał następująco:

- Wyznaczana została średnia arytmetyczna μ, odchylenie standardowe oraz wariancja σ^2 rozkładu normalnego, analogicznie jak w punkcie 1. zadania z laboratorium.
- Przedział $[\mu - 3\sigma, \mu + 3\sigma]$ został dzielony na k = 12 podprzedziałów o identycznej długości oraz określona została ilość liczb pseudolosowych n_j trafiająca do podprzedziałów o indeksie j. Ten etap również przebiegł analogicznie względem tego w punkcie 1.
- Wyznaczana została wartość statystyki testowej χ^2.
- Testowana została hipoteza H0: ”wygenerowany rozkład jest rozkładem normalnym $N(\mu_0 = 0.2, \sigma_0 = 0.5)$” na poziomie istotności α. Aby jednak dokonać poprawnego testu, potrzebna była wartość z tablicy statystycznej. Aby potwierdzić hipotezę, musi zostać zdefiniowany obszar krytyczny $K = \{X : \chi^2(X) > \varepsilon\}$, przy czym $X = \{x_1, x_2, \ldots, x_n\}$ - ciąg liczb pseudolosowych a ε - poziom krytyczny danego rozkładu, gdzie została ona odczytana z Tablicy 4, gdzie etykiety kolumn oznaczały poziom istotności α a wiersze to liczba stopni swobody $\nu = k - r - 1$ dla liczby parametrów testowanego rozkładu r = 2. Gdy $\chi^2 < \varepsilon$ uznawane jest, że dla zadanego poziomu istotności hipoteza H_0 jest prawdziwa czyli rozkład jest typu $N(\mu, \sigma)$.

3. Wyniki

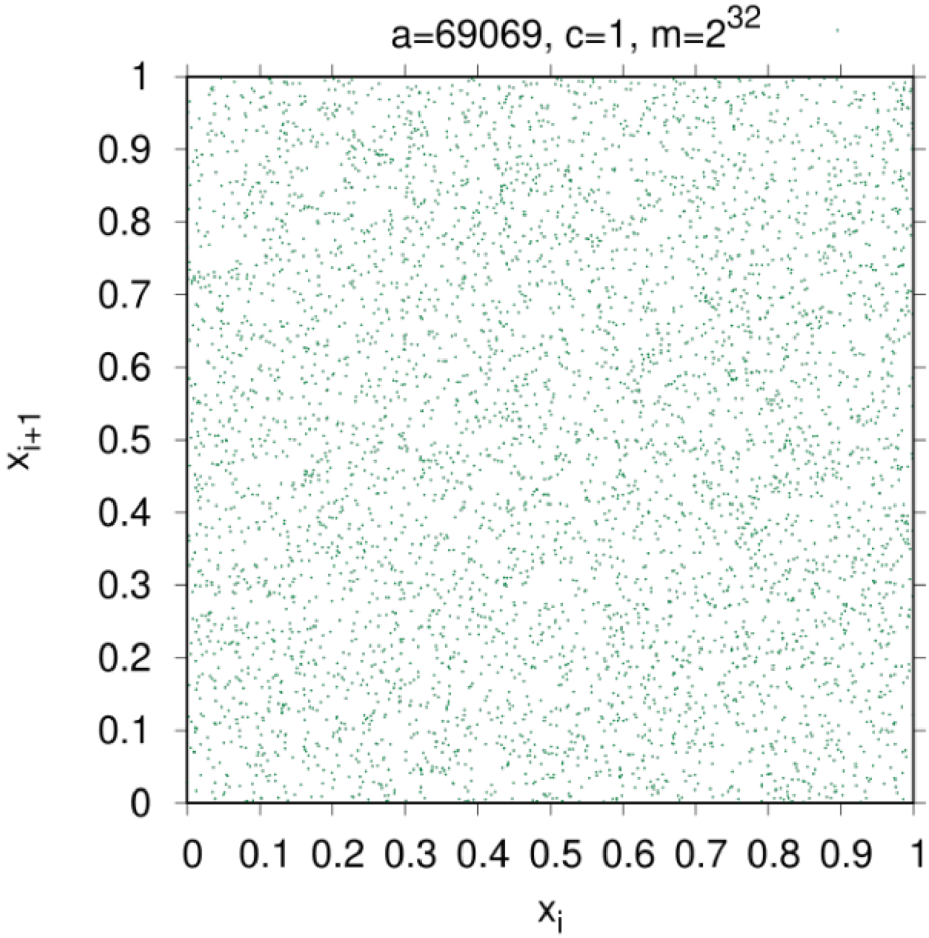
Dzięki programowi w języku C wykorzystującemu bibliotekę numeryczną Numerical Recipes, oraz przy pomocy skryptu Gnuplota wygenerowano, dla:

Rozkładu jednorodnorodnego



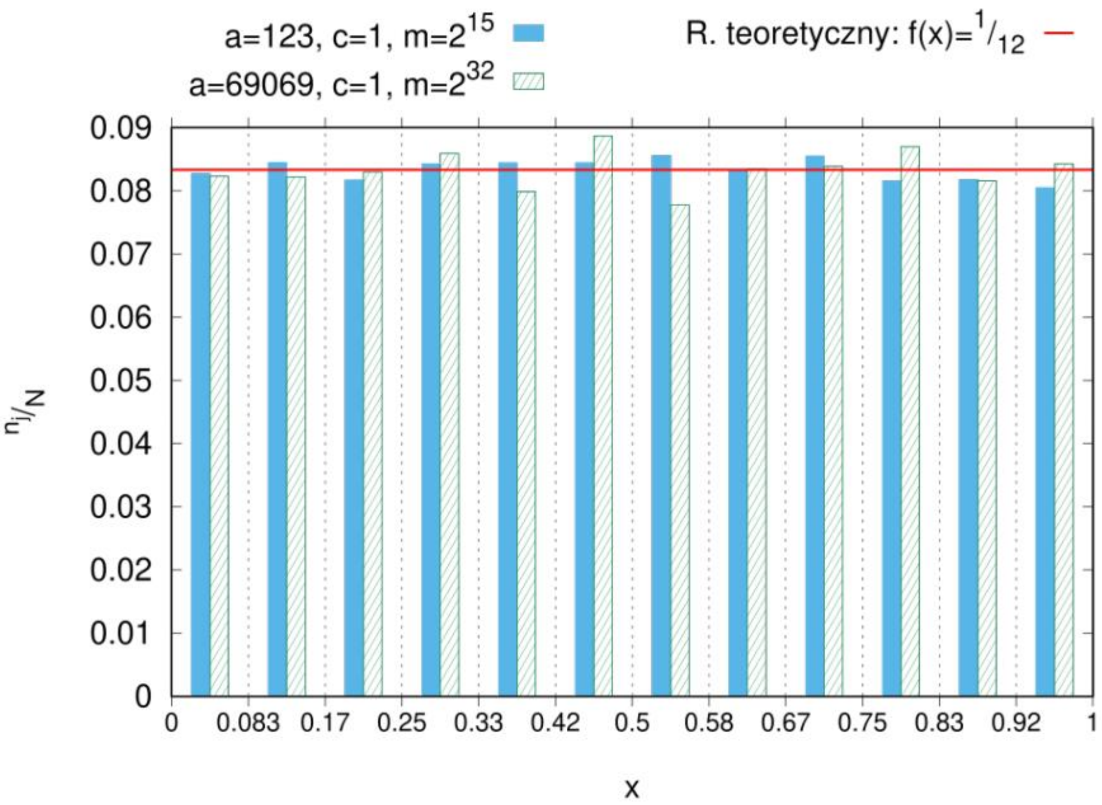
Rysunek 1.: : *Zależność $x_{i+1}(x_i)$ według generatora mieszanego o parametrach*

$a = 123, c = 1, m = 2^{15}$



Rysunek 2.: *Zależność* $x_{i+1}(x_i)$ *według generatora mieszanego o parametrach* $a = 69069,\ c = 1,\ m = 2^{32}$

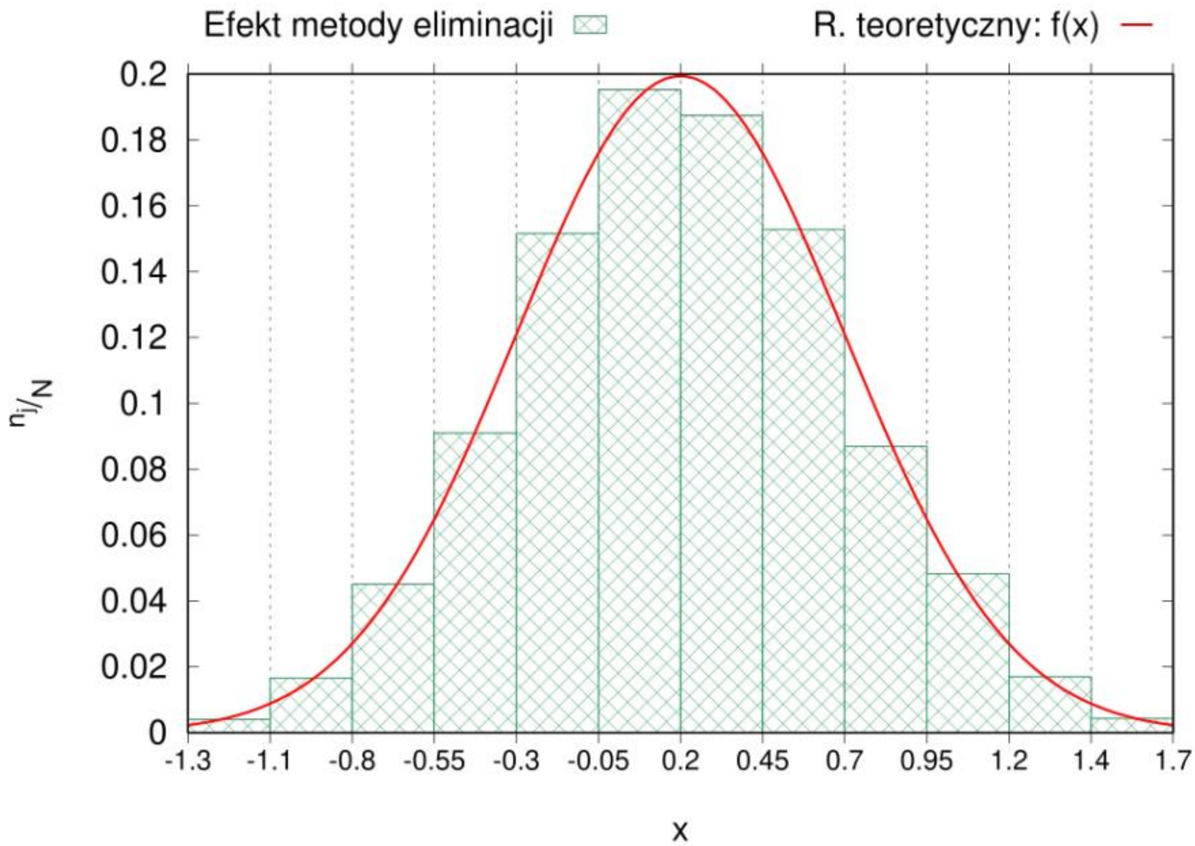
Trudno jest zauważyć jakikolwiek związek pomiędzy kolejnymi x_{i+1}, x_i , można więc stwierdzić, że generatory działają zgodnie z oczekiwaniami.



Rysunek 3.: *Rozkłady jednorodne U(0, 1) uzyskane przy użyciu generatorów mieszanych*

Obserwując słupki powyższego wykresu widzimy, że bardziej zróżnicowane są słupki generatora drugiego. Różnice między pierwszym a drugim generatorem nie są duże. Dla pierwszego generatora: średnia: $\mu = 0.498266$ odchylenie standardowe: $\sigma = 0.28712$. Dla drugiego generatora: średnia: $\mu = 0.501696$ odchylenie standardowe: $\sigma = 0.288302$ Wartości teoretyczny: średnia: $\mu = 0.5$ odchylenie standardowe: $\sigma = q\ 1\ 12 \approx 0.288675$ Bliższy wartościom teoretycznym jest generatorem numer dwa. Jednak oba nie odbiegają znacznie wartościom teoretycznym i działanie ich jest satysfakcjonujące.

Rozkładu normalnego



Rysunek 3.: *Rozkłady jednorodne U(0, 1) uzyskane przy użyciu generatorów mieszanych*

Na histogramie widzimy, że dzięki metodzie eliminacji uzyskaliśmy rozkład normalny. Uzyskane dla naszego generatora: średnia: μ = 0.19981 odchylenie standardowe: σ = 0.493137 wariancja: σ 2 = 0.243184. Wartości teoretyczne: średnia: μ = 0.2 odchylenie standardowe: σ = 0.5 Wartości obliczone dla wygenerowanego przez nas rozkładu normalnego są zbliżone do wartości teoretycznych. 5 Oznacza to, że nasz generator działa poprawnie. Statystyka testowa $\chi^2 = 12.3489$. Testowana przez nas hipoteza nie została odrzucona. Poziom ufności $P(\chi^2|\nu) = 0.805652$. Poziom istotności alfa = 0.19434

4. Wnioski

Generator kongruentny jest w stanie wygenerować ciąg liczb o rozkładzie zarówno jednorodnym U(1,0), jak również normalnym N(μ,σ). Dla obu przypadków otrzymane wartości zostały rozdzielone zgodnie z oczekiwaniami i na pierwszy rzut oka nie da się określić wzoru generującego wartości. To samo zostało zaprezentowane przez oczekiwane teoretyczne wartości odchylenia standardowego, czy średnie wartości ciągu. Trzeba jednak mieć na uwadze, że zbyt małe parametry w przepisie generującym ciąg mogą poskutkować pewnym rodzajem cyklicznie powtarzającego się schematu, co w rezultacie może nie dawać oczekiwanej losowości wartości. Oprócz tego, dla ciągu liczb o rozkładzie normalnym, został wykonany test χ^2 , który potwierdził, że otrzymywane zmienne losowe posiadały rozkład prawdopodobieństwa o oczekiwanej dystrybuancie.