Metody numeryczne
Sprawozdanie z ćwiczeń laboratoryjnych nr 14

Generowanie ciągu liczb pseudolosowych o rozkładzie jednorodnym i trójkątnym

Jan Zajda
Informatyka Stosowana
WFiIS
Akademia Górniczo-Hutnicza w Krakowie
6 czerwca 2020

1. Wstęp teoretyczny

Do generowania ciągu liczb pseudolosowych o rozkładzie jednorodnym możemy wykorzystać liniowy generator mieszany, będący połączeniem generatora multiplikatywnego (wykorzystujemy odpowiednio pomnożone poprzednie wartości generatora) oraz addytywnego (dodajemy stałą). Jego odmiana, którą wykorzystamy prezentuje się następująco:

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

gdzie a, c, m to parametry generatora. Wartość X_0 nazywamy ziarnem generatora. Odpowiedni wybór parametrów jest kluczowy dla jakości jego działania. Zazwyczaj wybieramy dużą wartość dla m oraz wartość a taką, aby $NWD(a, m) = 1$. Uzyskane wartości często normalizujemy, aby otrzymać liczby z przedziału $[0, 1)$:

$$x_i = \frac{X_i}{m+1}.$$

Generator o rozkładzie jednorodnym możemy wykorzystać do stworzenia generatora dowolnego rozkładu na przykład rozkładu trójkątnego zadanego wzorem:

$$f(x; \mu, \Delta) = \frac{-|x-\mu|}{\Delta^2} + \frac{1}{\Delta}.$$

Zmienną o takim rozkładzie uzyskujemy ze wzoru:

$$x = \mu + (\varepsilon_1 + \varepsilon_2 - 1) \cdot \Delta,\ \varepsilon_1, \varepsilon_2 \in U(0, 1).$$

Aby sprawdzić czy uzyskany ciąg N liczb spełnia założony rozkład to dzielimy przedział $[x_{min}, x_{max}]$ na k równych podprzedziałów. Następnie obliczamy liczbę elementów, które znalazły się w każdym z tych podprzedziałów i oznaczamy jako $n_j,\ j = 0, 1, 2, \ldots, k-1$. Następnie obliczamy teoretyczne prawdopodobieństwo, że element znajdzie się w określonym przedziale korzystając z dystrybuanty:

$$p_j = P(x_{j,min} < x \leq x_{j,max}) = F(x_{j,max}) - F(x_{j,min}),$$

$$F(a) = P(x < a) = \int_{x_{min}}^{a} f(x; \mu, \Delta).$$

Ostatecznie obliczamy statystykę testu ze wzoru:

$$\chi^2 = \sum_{j=0}^{k-1} \frac{(n_j - Np_j)^2}{Np_j}.$$

Uzyskaną liczbę porównujemy z wartościami tablicowymi dla stopnia swobody $\nu = k - r - 1$, gdzie r jest liczbą parametrów rozkładu. Jeżeli dla danego poziomu istotności α zachodzi nierówność $\chi^2 < \varepsilon$ to mówimy, że hipoteza H_0 nie została odrzucona z poziomem istotności α.

2. Zadanie do wykonania

2.1. Opis problemu

Za zadanie mamy stworzyć własny generator o rozkładzie jednorodym i utworzyć dzięki niemu $N = 10000$ znormalizowanych, pseudolosowych liczb. Porównamy wyniki dla dwóch zestawów parametrów:

$a = 123, \qquad c = 1, \qquad m = 2^{15}, \qquad X_0 = 10,$
$a = 69069, \qquad c = 1, \qquad m = 2^{32}, \qquad X_0 = 10.$

Następnie obliczymy wartość średnią i odchylenie standardowe dla obu wygenerowanych zestawów i porównamy z wartościami teoretycznymi.

W dalszej części zadania, przy pomocy wcześniej stworzonego generatora (drugi zestaw parametrów) uzyskamy ciąg $N = 1000$ liczb o rozkładzie trójkątnym $T(\mu = 4, \Delta = 3)$, a więc:

$[x_{min}, x_{max}] = [\mu - \Delta, \mu + \Delta]$.

Dla każdego ciągu liczb utworzymy histogram gęstości rozkładu prawdopodobieństwa. Dla dwóch pierwszych ciągów wybierzemy $k = 12$ podprzedziałów, a dla trzeciego $k = 10$.

Na koniec obliczymy statystykę testową χ^2 i sprawdzimy na poziomie istotności $\alpha = 0.05$ hipotezę H_0, że wygenerowany rozkład jest rozkładem $T(4, 3)$, wobec hipotezy H_1, że nie jest to prawda.

2.2. Wyniki

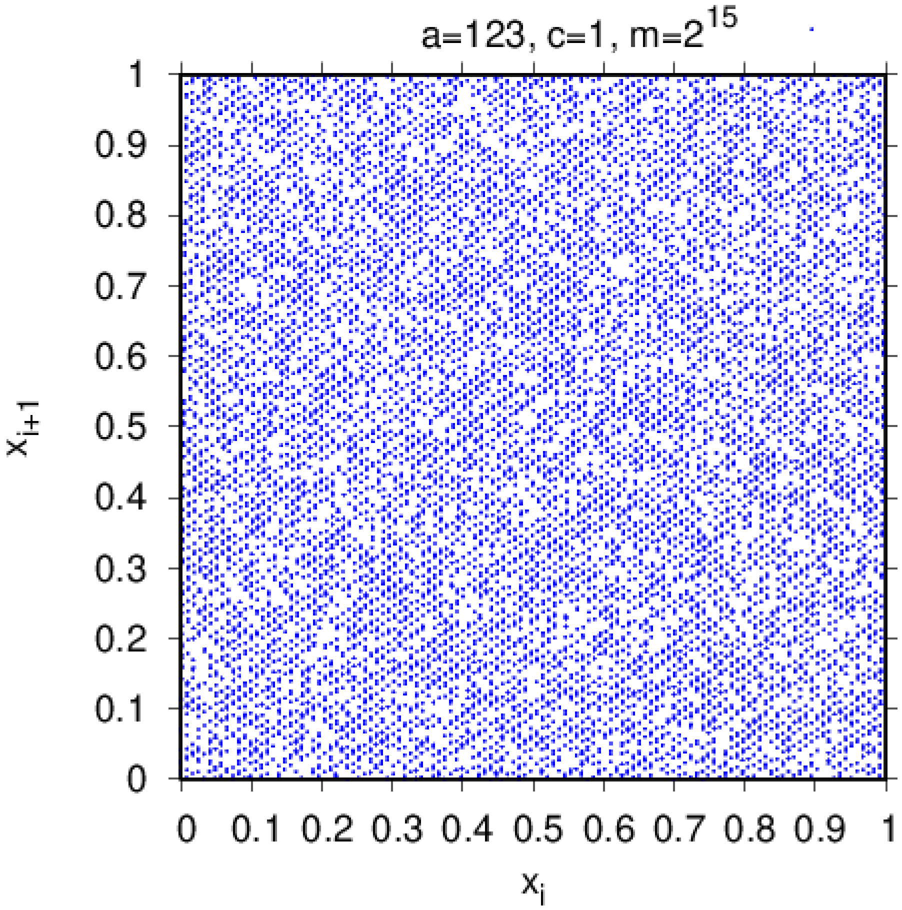
Jako wynik działania programu przedstawiono wykres przedstawiający zależność $x_{i+1}(x_i)$ dla znormalizowanych ciągów uzyskanych za pomocą generatora mieszanego.
Pokazano także dwa histogramy gęstości rozkładu prawdopodobieństwa:

- pierwszy dla wyników uzyskanych przy użyciu obu przypadków generatora mieszanego ($k = 12$),
- drugi dla wyników uzyskanych dla rozkładu trójkątnego ($k = 10$).

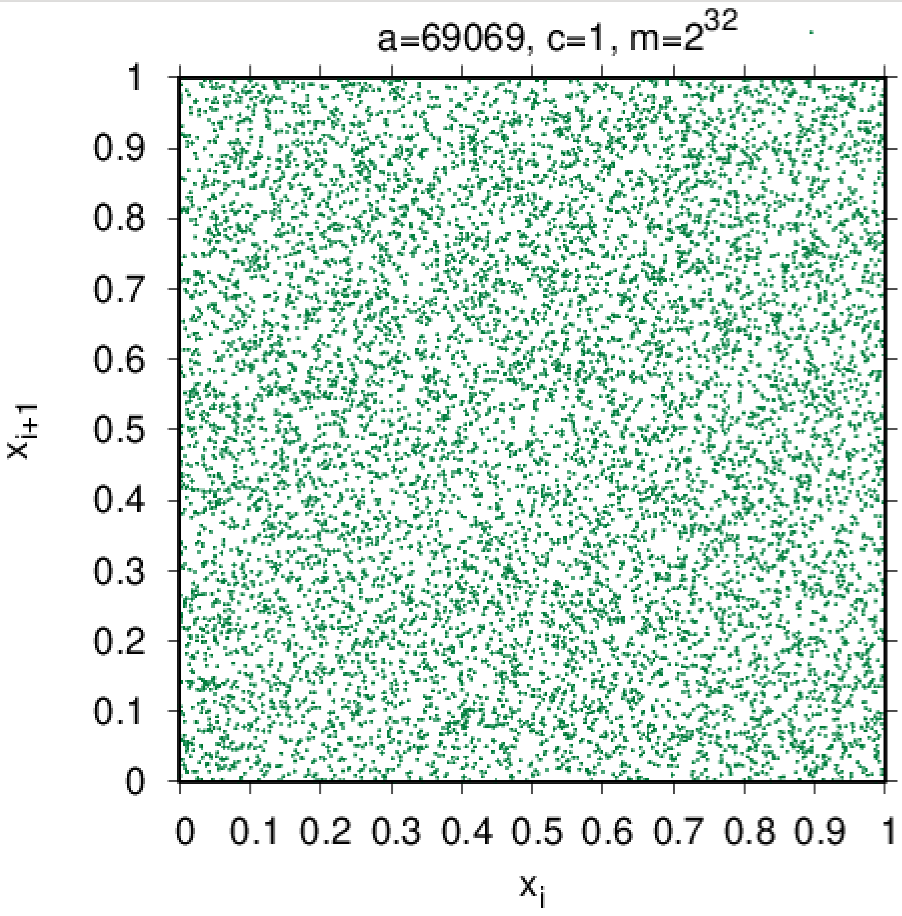
Dodatkowo wypisano wartości średnie, odchylenia standardowe i zbadano hipotezę H_0.

Parametr	Wartość średnia	Odchylenie standardowe
Pierwszy generator	0.498266	0.28712
Drugi generator	0.503806	0.28807

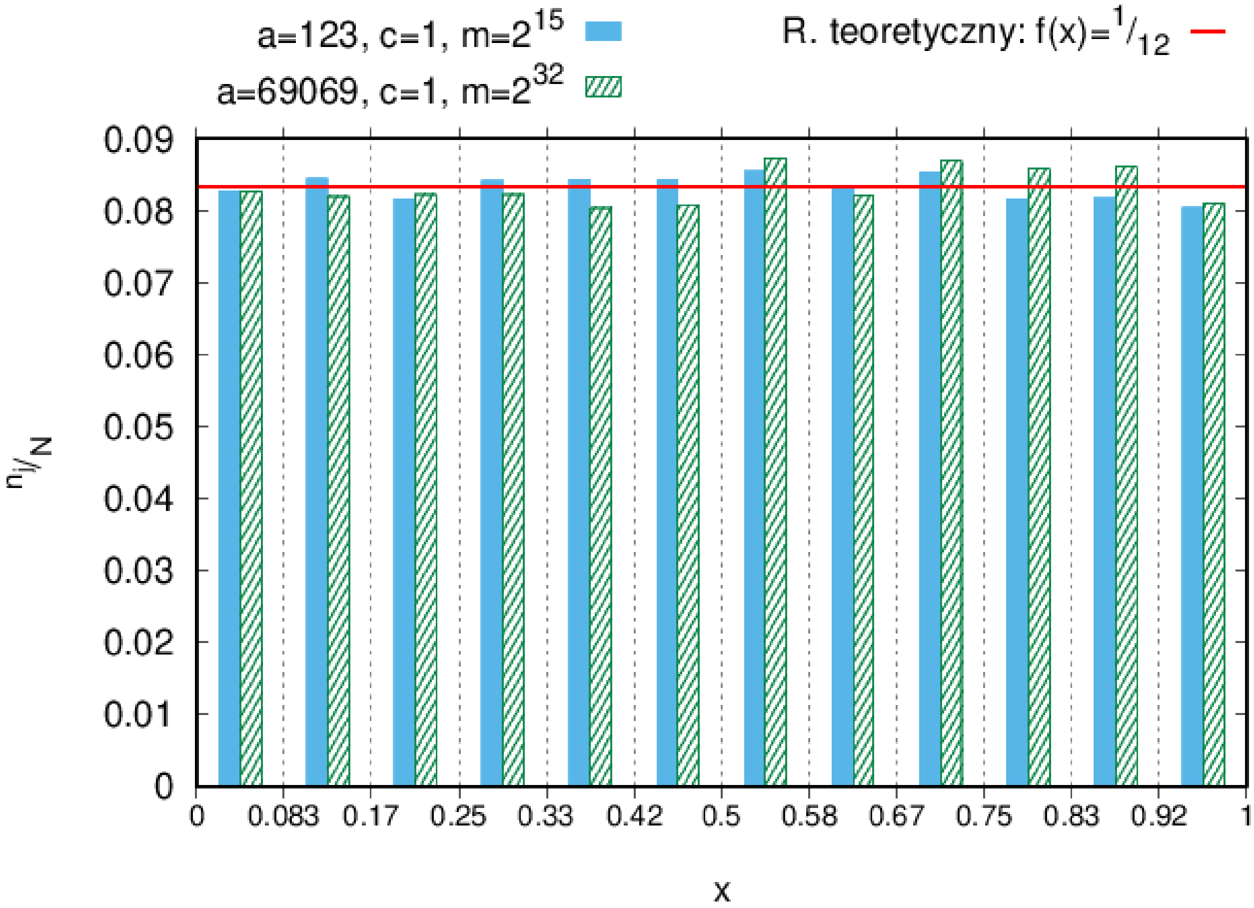
Wartości teoretyczne: $\mu = 0.5,\ \sigma = \sqrt{\frac{1}{12}} \approx 0.288675.$



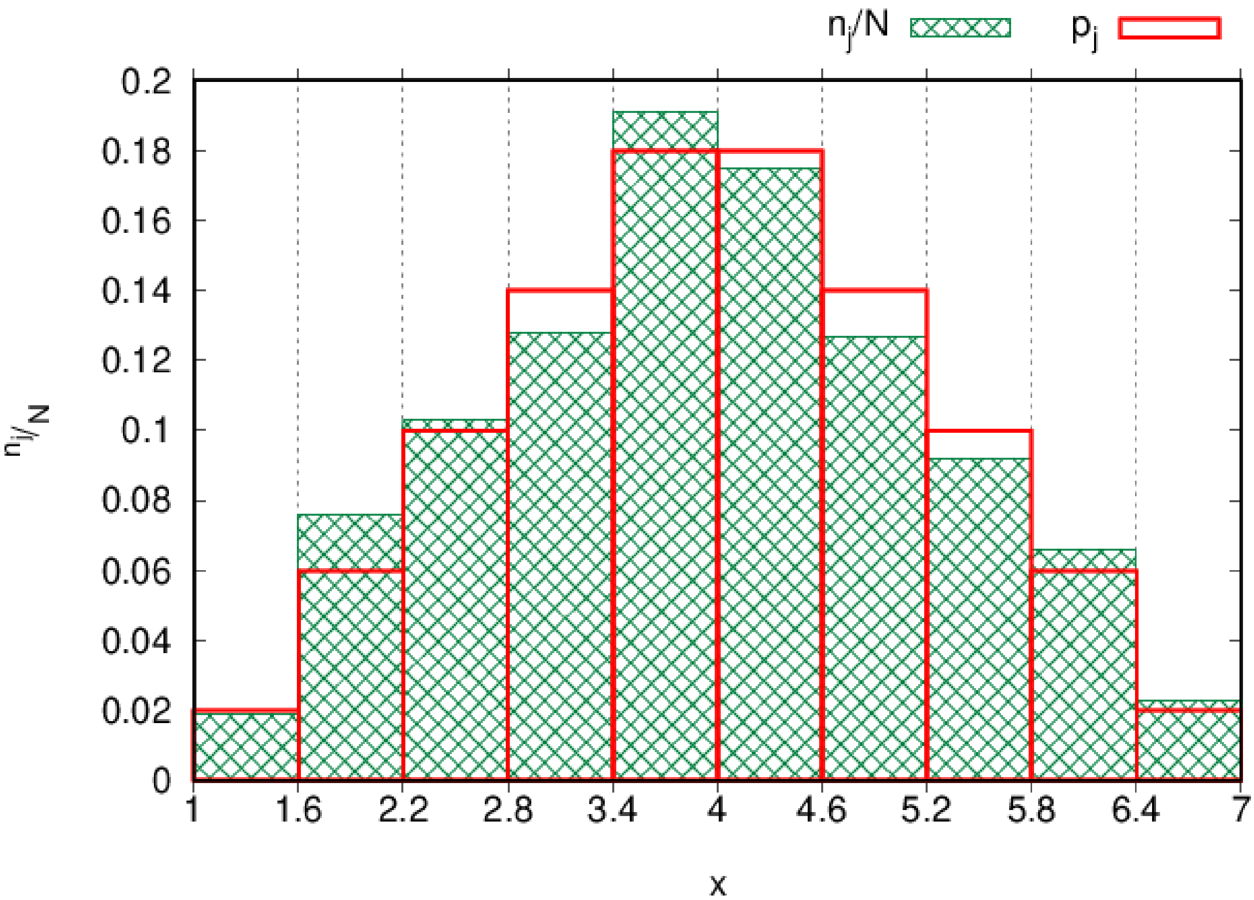
Wykres 1: Zależność $x_{i+1}(x_i)$ dla wyników uzyskanych za pomocą pierwszego generatora mieszanego.



Wykres 2: Zależność $x_{i+1}(x_i)$ dla wyników uzyskanych za pomocą drugiego generatora mieszanego.



Wykres 3: Historgram gęstości prawdopodobieństwa dla dwóch generatorów mieszanych - $k = 12$ podprzedziałów.



Wykres 4: Historgram gęstości prawdopodobieństwa dla rozkładu trójkątnego - $k = 10$ podprzedziałów.

Obliczony stopień swobody: $\nu = 7$.
Dla $\nu = 7, \alpha = 0.05$ odczytana wartość: $\varepsilon = 14.06$.
Obliczona wartość statystyki testowej: $\chi^2 = 9.14349$.

Zachodzi nierówność: $\chi^2 < \varepsilon \Longrightarrow$ hipoteza H_0 nie została odrzucona na poziomie istotności $\alpha = 0.05$.

3. Wnioski

Na podstawie wartości uzyskanych parametrów μ, σ - bardzo zbliżonych do wartości teoretycznych dla rozkładu jednorodnego, możemy wnioskować o poprawnym działaniu obu generatorów mieszanych. Potwierdza to też histogram, którego wartości w każdym podprzedziale są zbliżone do teoretycznej wartości gęstości rozkładu równej $1/12$.

Patrząc na zależności $x_{i+1}(x_i)$ możemy jednak zaobserwować, że lepsze wyniki daje drugi generator. Liczby rozsiane są bardziej chaotycznie i nie da się łatwo dostrzec między nimi pewnych zależności, w przeciwieństwie do pierwszego generatora gdzie układają się czasami w łatwo widoczne linie. Prawdopodobnie jest to spowodowane znacząco większą wartością m w drugim zestawie parametrów.

Obserwując histogram dla rozkładu trojkątnego widzimy, że wyniki w pewnym przybliżeniu pokrywają się z oczekiwaniem teoretycznym. Nadal dobrze widoczny jest spodziewany kształt trójkąta. Również statystyka testowa potwierdza hipotezę o poprawności wygenerowanego rozkładu.