Identyfikacja i modelowanie statystyczne
Generator liczb pseudolosowych

Termin zajęć:
08.03.2022

Autorzy:
Rafał Rzewucki 248926

Prowadzący: dr heb. inż. Paweł Wachel

9 marca 2022

1 Generator liczb pseudolosowych - przebieg piłokształtny

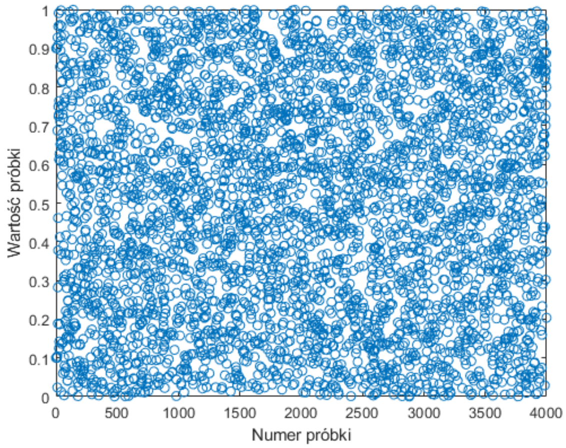
1.1 Zadanie

Zadaniem było przygotowanie własnej implementacji generatora liczb pseudolosowych z rozkładu jednostajnego w zakresie $[0, 1]$. Generator miałbyć oparty na przekształceniu piłokształtnym, które opisane jest równaniem

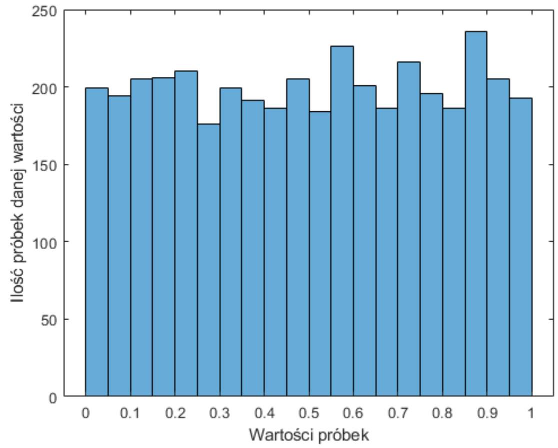
$$X_{n+1} = X_n \cdot z - \lfloor X_n \cdot z \rfloor$$

1.2 Wyniki działania generatora

Generator w założeniu ma zwracać liczby z zakresu $[0, 1]$, co oznacza, że jego wartość początkowa musi również być z tego zakresu.

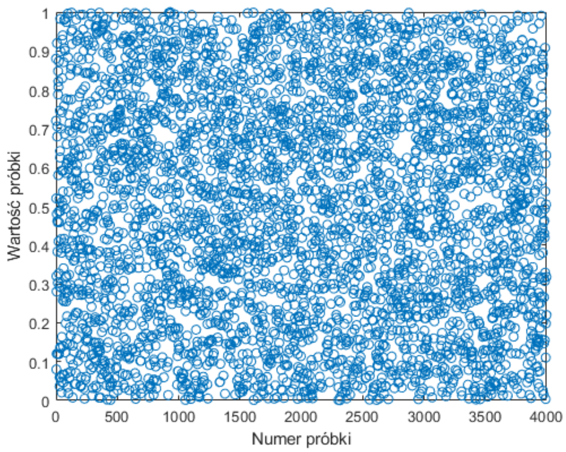


(a) Rozkład liczb losowych dla wartości początkowej x=0.1

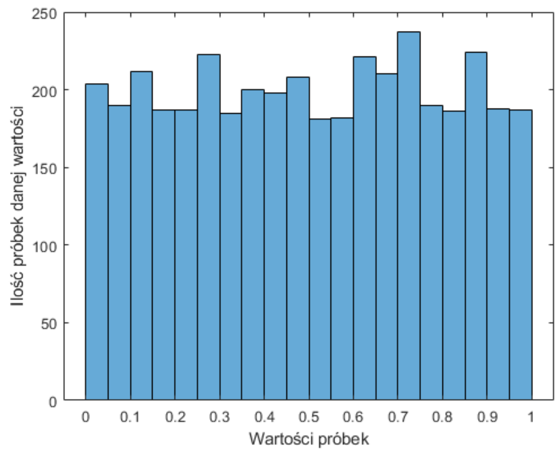


(b) Histogram dla wartości początkowej x=0.1

Modyfikując wartość początkową jesteśmy w stanie zmienić otrzymywany rozkład generowanych liczb.

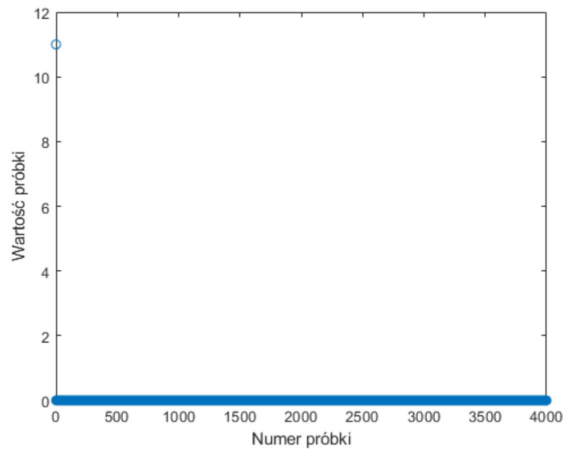


(c) Rozkład liczb losowych dla wartości początkowej x=0.72

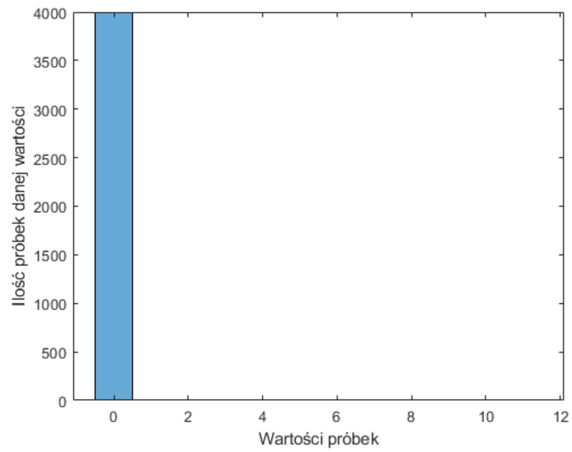


(d) Histogram dla wartości początkowej x=0.72

Gdy wartość począkowa wykroczy poza zakres $[0, 1]$ wyjście genratora będzie zerowe, oprócz wartości początkowej zadanej na jego wejście.

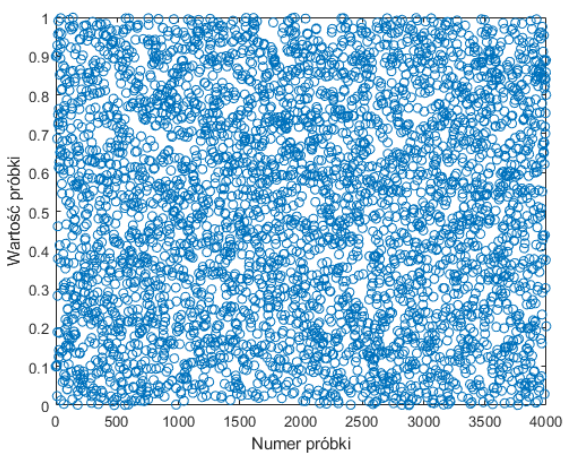


(e) Rozkład liczb losowych dla wartości początkowej x=11

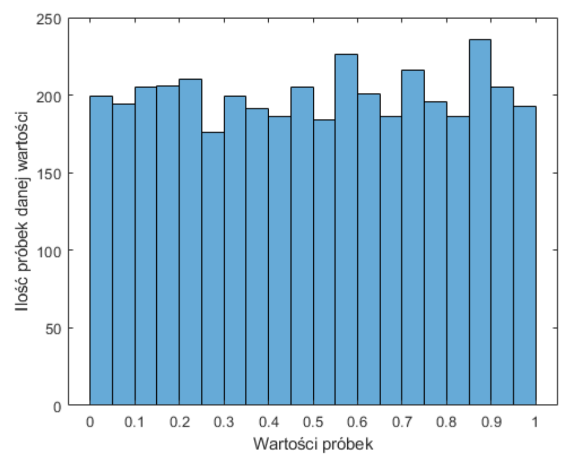


(f) Histogram dla wartości początkowej x=11

Zwiększając liczbę otrzymywanych wyników nie wpływamy bezpośrednio na ostateczne wartości jakie otrzymujemy z generatora



(g) Rozkład liczb losowych dla wartości początkowej x=0.1 - 40 000 wyników



(h) Histogram dla wartości początkowej x=0.1 - 40 000 wyników

Zauważyć można, że lepsze wyniki z generatorów opartych na przebiegu piłokształtnym uzyskiwane są, gdy liczba zębów jest wartością nieparzystą, a najlepsze, gdy liczba zębów jest liczbą pierwszą. Zwiększa to znacząco okres generatora, co sprawia, że liczby rzadziej się powtarzają.

Okres generatora został wyznaczony na podstawie analizy tablicy otrzyanych wyników. Przyjęto wartość początkową równą 0.2 oraz 3 zęby piły w genratorze. Na podstawie takich danych wygenerowano 100 próbek, których wartości przedstawiono na zrzucie z programu MATLAB poniżej:

```
result =

  Columns 1 through 14

    0.3300    0.9900    0.9700    0.9100    0.7300    0.1900    0.5700    0.7100    0.1300    0.3900    0.1700    0.5100    0.5300    0.5900

  Columns 15 through 28

    0.7700    0.3100    0.9300    0.7900    0.3700    0.1100    0.3300    0.9900    0.9700    0.9100    0.7300    0.1900    0.5699    0.7098

  Columns 29 through 42

    0.1294    0.3881    0.1643    0.4929    0.4788    0.4364    0.3092    0.9275    0.7826    0.3477    0.0430    0.1289    0.3867    0.1602
```

Rysunek 1: Zrzut wyników generowania z programu MATLAB w celu zbadania okresowości

Zauważyć można, że początkowa sekwencja liczb $0.33, 0.99, 0.91$ zaczyna powtarzać się już w 21 wyniku generatora.

2 Generator liczb pseudolosowych oparty na specyficznym równaniu

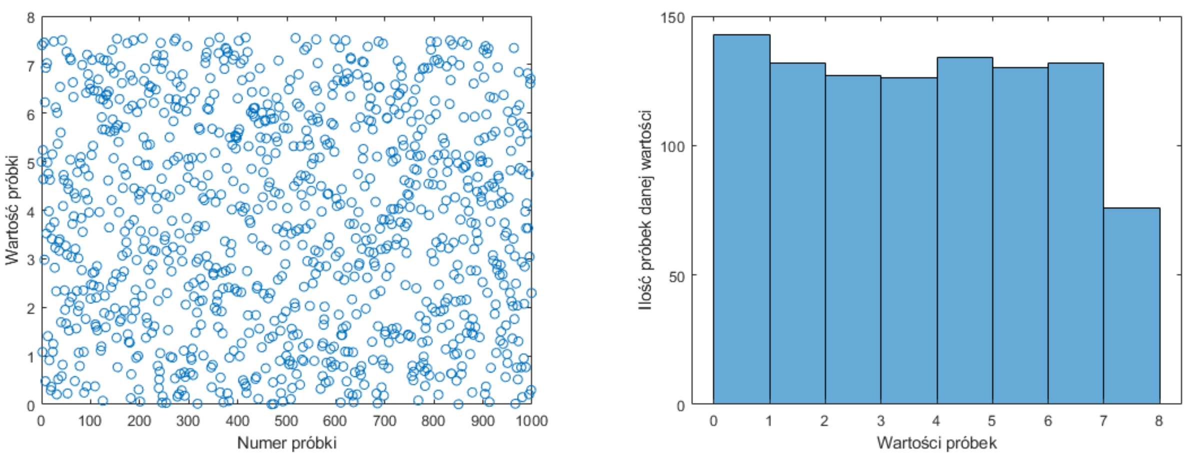
2.1 Zadanie

Zadaniem było przygotowanie własnej implementacji generatora liczb pseudolosowych opartego na równaniu

$$X_{n+1} = (a_0 X_n + a_1 X_{n-1} + \cdots + a_k X_{n-k} + c) mod\, m$$

2.2 Wyniki działania generatora

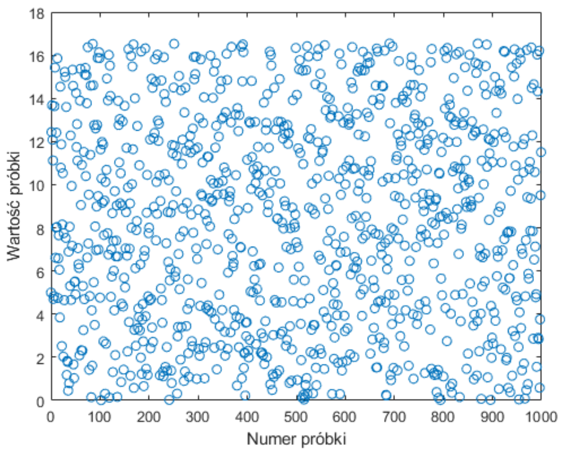
W badaniu własności generatora przyjęto generowanie 1000 próbek przy każdym podejściu. Głównymi parametrami dla tego typu generatora jest wartość początkowa, określająca próbkę od której rozpoczyna się sekwencja liczb oraz liczba oznaczona we wzorze jako m, jest to wartość, która uczestniczy w operacji modulo i określa nam zakres wartości otrzymywanych wyników, przy prawidłowym ustawieniu wartości początkowej będzie to zakres $[0, m]$.



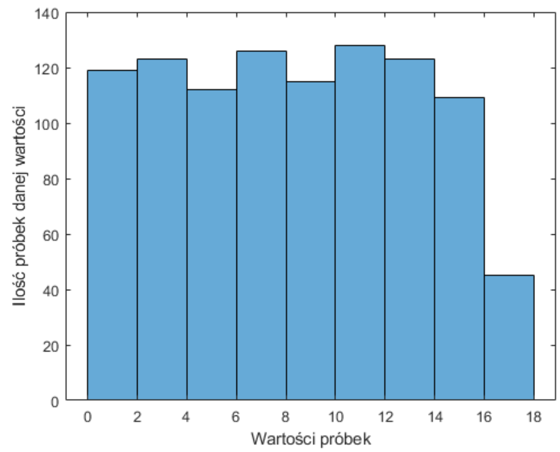
(a) Rozkład liczb losowych dla wartości x=5 i m=7.56

(b) histogram dla wartości x=5 i m=7.56

Gdy zwiększymy wartość współczynnika m, to zgodnie z założeniami zwiększy się również zakres generowanych liczb

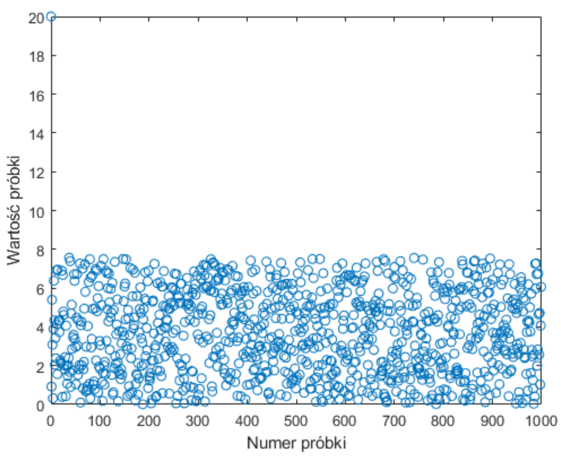


(c) Rozkład liczb losowych dla wartości x=5 i m=16.56

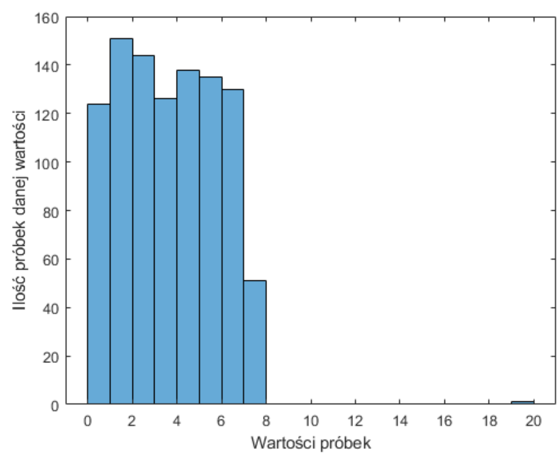


(d) histogram dla wartości x=5 i m=16.56

Prawidłowe ustawienie watości początkowej jest tutaj kluczowe dla uzyskania poprawności wyników, gdyż jeśli ustawimy wartość początkową większą od wartości współczynnika m, otrzymamy niespójne wyniki, w których wartość początkowa będzie wykraczać poza zakładany zakres.



(e) Rozkład liczb losowych dla wartości x=20 i m=7.56



(f) histogram dla wartości x=20 i m=7.56

3 Generator liczb pseudolosowych - metoda odwracania dystrybuanty

Zadaniem było przygotowanie implementacji generatora liczb pseudolosowych opartego na gęstości prawdopodobieństwa

3.1 Gęstość numer 1

3.1.1 Obliczenia

$$f(x) = \begin{cases} 2x \quad dla \quad x \in [0,1] \\ 0 \quad dla \quad x \in (-\infty, 0) \cup (1, \infty) \end{cases} \tag{1}$$

W kolejnym kroku została wyznaczona całka w zakresie $x \in [0,1]$

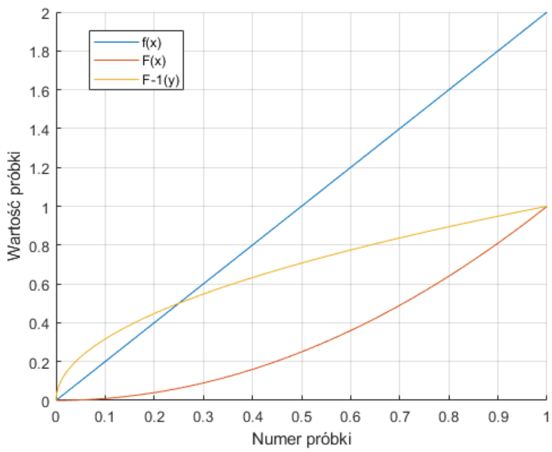
$$F(x) = \int_{-\infty}^{x} f(t)dt = 0 + \int_{0}^{x} 2tdt = 2 \cdot \left[\frac{t_2}{2}\right]_0^x = x^2 \tag{2}$$

$$F(x) = \begin{cases} 0 \quad dla \quad x \in (-\infty, 0) \\ x^2 \quad dla \quad x \in [0,1] \\ 1 \quad dla \quad x \in (1, \infty) \end{cases} \tag{3}$$

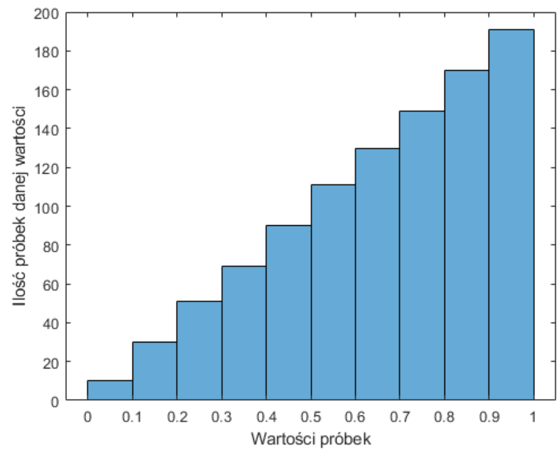
i na tej podstawie wyznaczono odwrotną dystrybuantę dla przedziału $x \in [0, 1]$

$$F^{-1}(y) = \sqrt{y}. \tag{4}$$

3.1.2 Wyniki działania generatora



(g) Wykresy poszczególnych funkcji



(h) histogram

3.2 Gęstość numer 2

3.2.1 Obliczenia

Z gęstości prawdopodobieństwa

$$f(x) = \begin{cases} x+1 \quad dla \quad x \in (-1, 0) \\ -x+1 \quad dla \quad x \in [0, 1) \\ 0 \quad dla \quad x \notin (-1, 1) \end{cases} \tag{5}$$

wyznaczono dystybuante

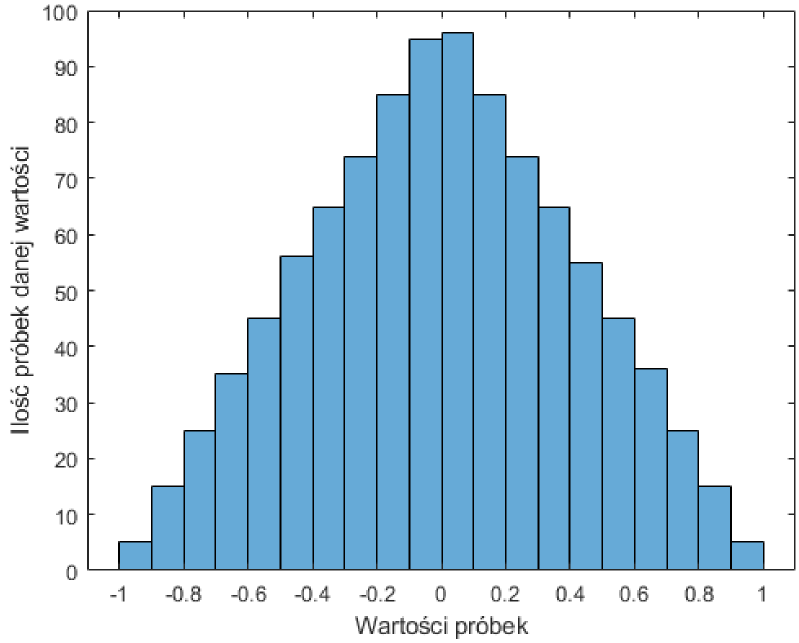
$$F(x) = \begin{cases} \frac{1}{2} - \frac{x^2}{2} + x \quad dla \quad x \in [0, 1) \\ \frac{1}{2} + \frac{x^2}{2} + x \quad dla \quad x \in (-1, 0) \\ 0 \quad dla \quad x \in (-\infty, -1] \\ 1 \quad dla \quad x \in [1, \infty) \end{cases} \tag{6}$$

, którą następnie poddano operacji odwracania funkcji, co dało nam

$$F^{-1}(y) = \begin{cases} \sqrt{2y} - 1 \quad dla \quad x \in [0, \frac{1}{2}] \\ 1 - \sqrt{2 - 2y} \quad dla \quad x \in (\frac{1}{2}, 1] \end{cases} \tag{7}$$

3.2.2 Wynik działania generatora

Korzystając ze wzoru uzyskanego powyżej, potraktowano go jako genrator. Histogram otrzymanych wyników został zamieszczony poniżej.



(i) Histogram

3.3 Gęstość numer 3

3.3.1 Obliczenia

W tym przykładzie została wyznaczona odwrotna dystrybuanta generująca liczby losowe z rozkładu wykładniczego opisanego wzorem

$$f(x) = e^{-x} \tag{8}$$

, w którym $x \geqslant 0$. Dystrybuanta tej gęstości wynosi

$$F(x) = 1 - e^{-x} \tag{9}$$

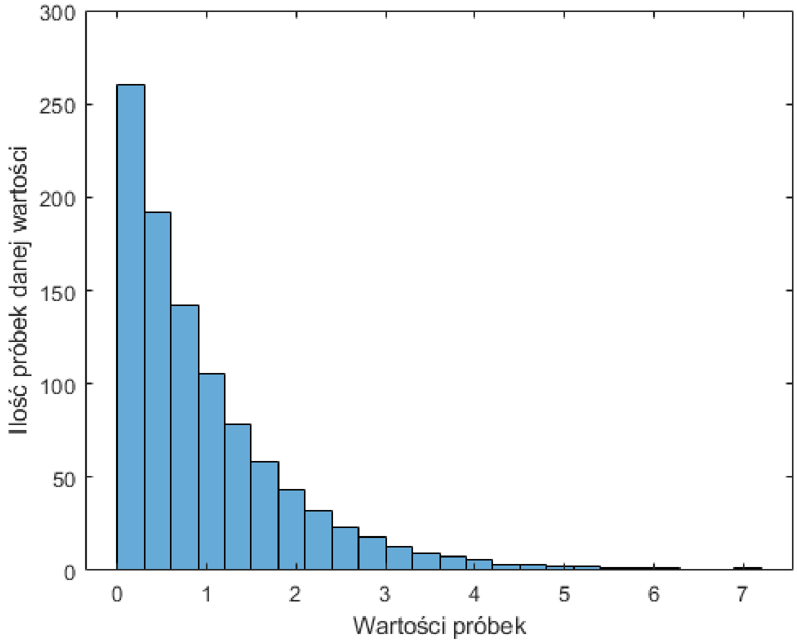
co daje odwrotną dystrybuantę

$$F^{-1}(y) = -ln(1 - y) \tag{10}$$

dla $y \in [0, 1]$

3.3.2 Wynik działania generatora

Na podstawie powyższego wzoru wygenerowano wyniki, których histogram został zamieszczony poniżej.



(j) Histogram

3.4 Gęstość numer 4

3.4.1 Obliczenia

W ostatnim przykładzie został zaimplementowany generator bazujący na rozkładzie Laplacea, którego dystrybuanta wynosi

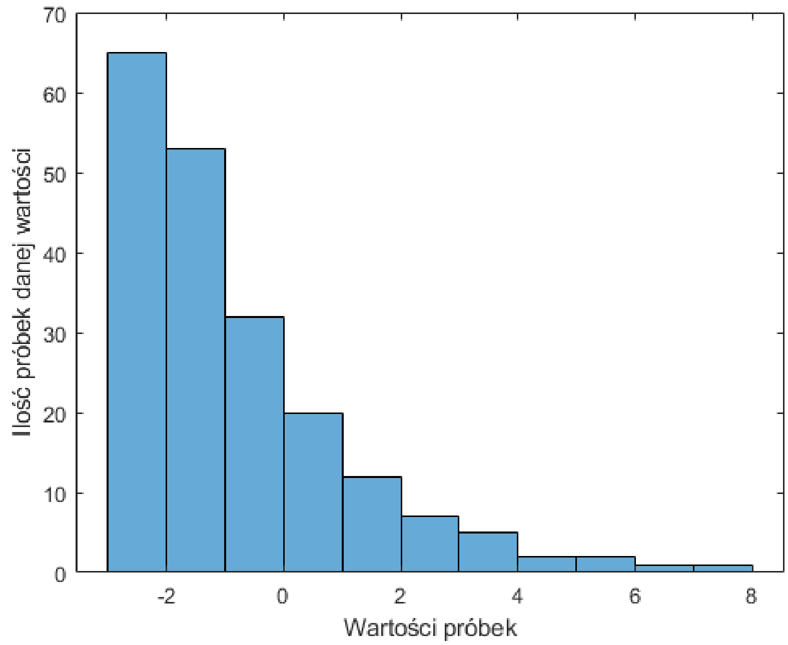
$$F(x) = \begin{cases} \frac{1}{2}e^{(\frac{x-u}{b})} & dla \quad x \leqslant u \\ 1 - \frac{1}{2}e^{(-\frac{x-u}{b})} & dla \quad x \geqslant u \end{cases} \tag{11}$$

, z której wyznaczono odwrotną dystrybuantę

$$F^{-1}(y) = ln(1-y) \tag{12}$$

3.4.2 Wynik działania generatora

Korzystając ze wzoru na odwróconą dystrybuantę wyznaczono histogram generatora bazującego na powyższym rozkładzie.



(k) Histogram

4 Wnioski

Generatory o rozkładzie równomiernym są bardzo ważnym elementem wielu generatorów liczb pseudolosowych ponieważ stanowią podstawę do budwy bardziej złożonych generatorów. Prostota ich działania powoduje, że są one stosunkowo proste w użyciu i nie zużywają wielu zasobów obliczeniowych.
Analizując generatory o rozkładzie jednostajnym można zauważyć, że dla większej liczby próbek otrzymujemy równomierny ilościowo rozkład, co zmniejsza ryzyko, że uzyskamy dwa wyniki z podobnego zakresu liczb.
Okres generatora piłokształtnego został wyznaczony w sposób przybliżony. Dla wielu tysięcy próbek bardzo trudno jest znaleźć powtarzające się wartości, dlatego rozpatrywano tylko 100 pierwszych próbek, dla bardzo małej liczby zębów. Zwiększając liczbę zębów generatora, można w prosty sposób modyfikować jego własności.
Metoda odwrotnej dystrybuanty opiera się na zupełnie innym podejściu, niż metoda generowania wyników na podstawie przebiegu piłokształtnego, dlatego można dopasować własności konkretnych generatorów do swoich potrzeb. Zaletą metody odwrotnej dystrybuanty jest fakt, że pozwala ona zmiennej losowej w rozkładzie znajdującym się w przedziale $[0, 1]$ wygenerować rozkład z całej dostępnej dziedziny. Ze względu na własności samej gęstości prawdopodobieństwa, rozkład dla wartości o małej wariacji będzi zbliżony do generatora jednostkowego. Metoda ta nie jest jednak uniwersalna i nie można jej zastosować dla dystrybuant, dla których nie możliwe jest wykonanie operacji odwrócenia. W lepszych przypadkach odwrócenie jest bardzo skomplikowane, więc można poszukać prostrzej metody, która da nam zbliżone rezultaty.