Politechnika Wrocławska

Wykonawcy: Julia Grzegorzewska 262314, Wiktoria Fimińska 262283
Prowadzący: dr inż. Michał Balcerek
Termin zajęć: Wtorek 9:15

Symulacje komputerowe laboratorium

Metody generowania zmiennych losowych

27 kwietnia 2022

Spis treści

1. Abstrakt

Raport przedstawia opisy metod generowania zmiennych losowych poprzez pokazanie ogólnych algorytmów ich tworzenia wraz z przykładami dla rozkładu dyskretnego i ciągłego. Oprócz tego zawiera opis liniowego generatora kongruentnego oraz wyjaśnienie, czym są liczby "pseudolosowe". Przykłady generowane są w języku programistycznym *Python*, a otrzymane rozkłady porównywane są z rozkładami teoretycznymi dla różnych długości prób.

Materiały i informacje bazowane były na wykładach prof. dr. hab. inż. Marcina Magdziarza, Politechnika Wrocławska.

2. Metody generowania zmiennych losowych

2.1. Metoda odwracania dystrybuanty

Metoda, jak sama nazwa wskazuje, polega na odwróceniu dystrybuanty danego rozkładu. To znaczy, że biorąc $X = F^{-1}(U)$, gdzie $U \sim \mathcal{U}(0,1)$, otrzymamy zmienną losową X o dystrybuancie F. W sytuacji, gdy F nie jest ściśle odwrotna, to definiujemy tzw. uogólnioną dystrybuantę odwrotną, która wyraża się następującym wzorem $\widetilde{F}_x^{-1}(y) = inf\,\{x \in \mathbb{R} : F_x(x) \geqslant y\}$.

2.1.1. Rozkład dyskretny

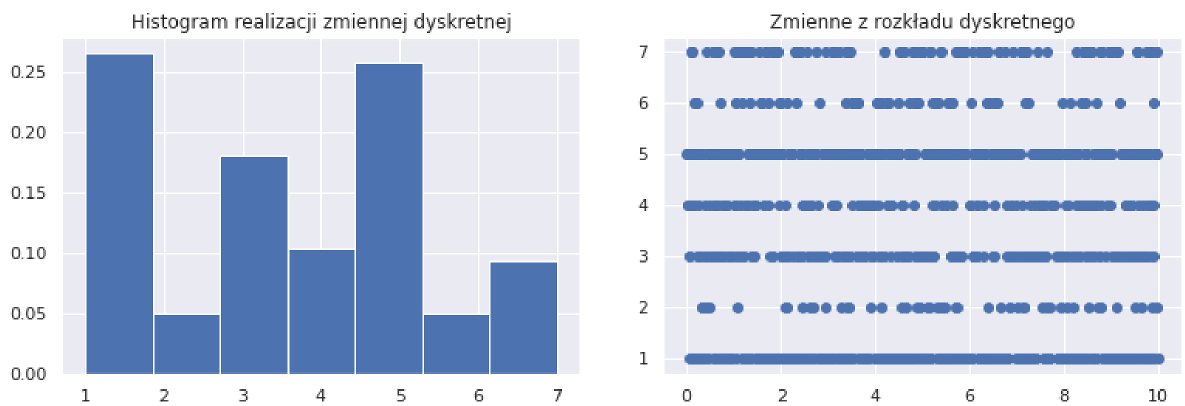
Algorytm:

1.) generuj $U \sim \mathcal{U}(0,1)$
2.) wyznacz $i \in \mathbb{N}$ takie, że:

$$\sum_{k=1}^{i-1} p_k < U \leqslant \sum_{k=1}^{i} p_k$$

3.) zwróć $X = X_i$

Przykład:
Korzystając z powyższej metody generujemy 10^3 realizacji dyskretnej zmiennej losowej X o rozkładzie $P(X=1) = 0.25, P(X=2) = 0.05, P(X=3) = 0.2, P(X=4) = P(X = 7) = 0.1, P(X = 5) = 0.26, P(X=6) = 0.04$.



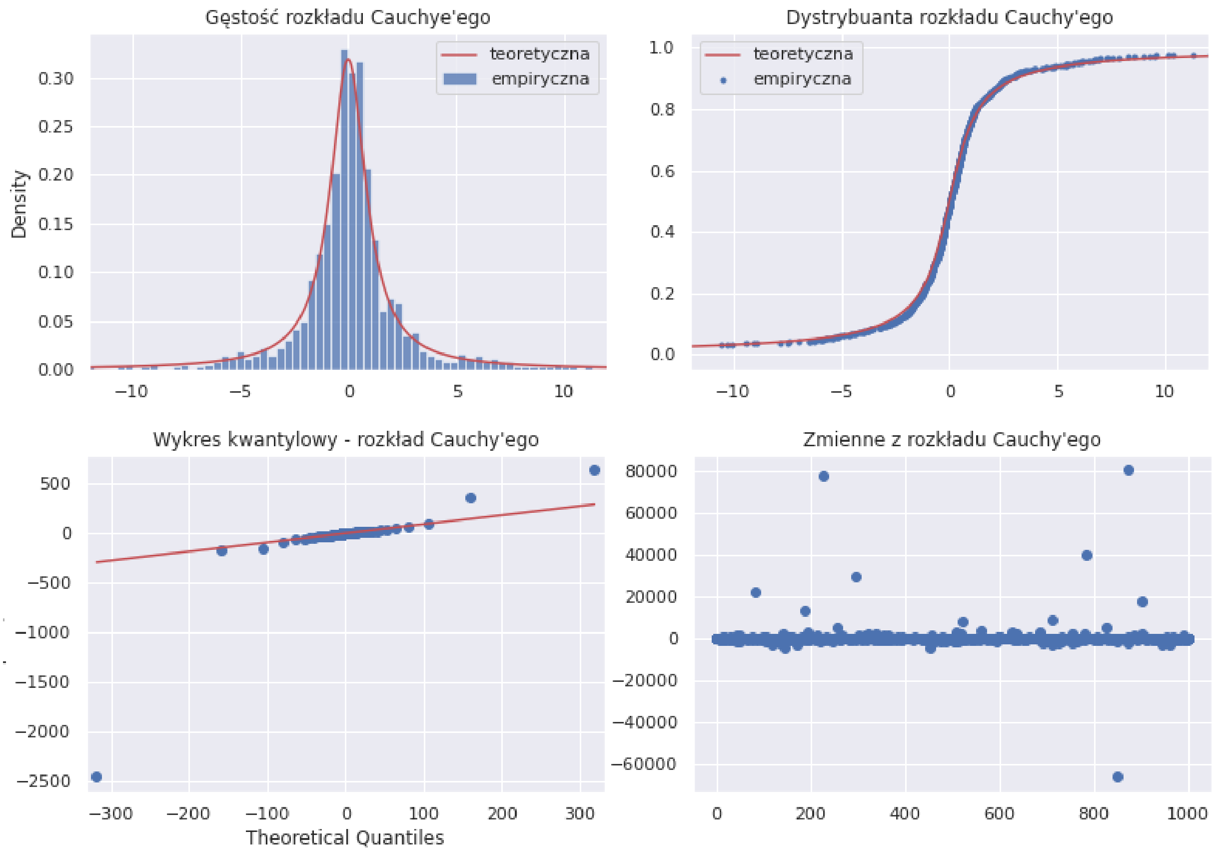
Rysunek 1

2.1.2. Rozkład ciągły

Algorytm:

1.) wyznacz $F_x^{-1}(y)$
2.) generuj $U \sim \mathcal{U}(0,1)$
3.) wstaw $X = F_x^{-1}(U)$

Przykład:
Generujemy 10^3 realizacji zmiennej losowej X z rozkładu Cauchy'ego, $X \sim \mathcal{C}(0,1)$.



Rysunek 2

2.2. Metoda akceptacji-odrzucenia

Metoda ta wykorzystywana jest głównie do generowania zmiennych z rozkładu, w którym dystrybuanta odwrotna jest ciężka lub niemożliwa do wyznaczenia. Wykorzystuje się w niej prawdopodobieństwo geometryczne oraz metodę Monte Carlo.

2.2.1. Rozkład dyskretny

Założenia:

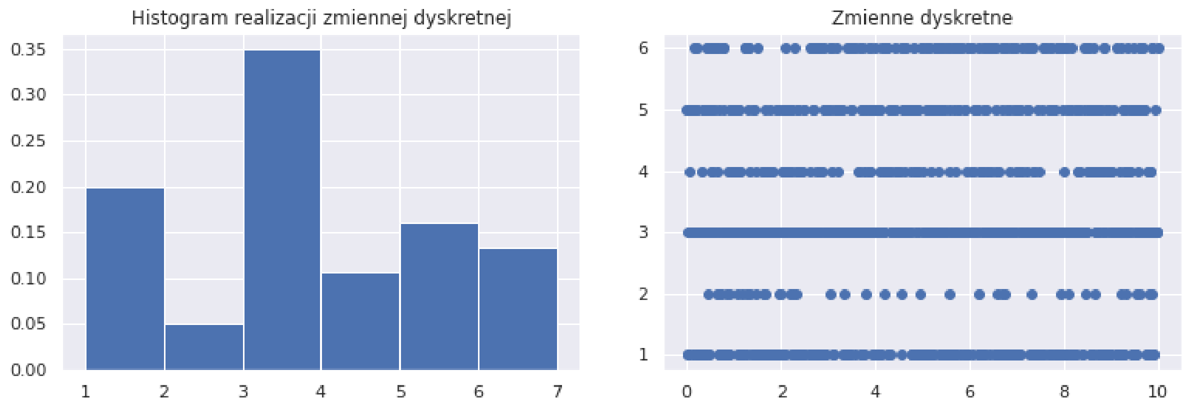
— potrafimy efektywnie generować zmienną losową Y o rozkładzie $q_i = P(Y = i)$;
— X i Y przyjmują wartości z tego samego zbioru;
— potrafimy wyznaczyć stałą $0 < c < \infty$, t.że $\max_i \frac{p_i}{q_u} \leqslant c$.

Należy wyznaczać możliwie najmniejsze c, ponieważ $\frac{1}{c}$ jest prawdopodobieństwem sukcesu w pojedyńczej iteracji, a c średnią liczbą powtórzeń algorytmu.

Algorytm:

1.) generuj Y
2.) generuj $U \sim \mathcal{U}(0,1)$, $U \perp Y$
3.) jeśli $U \leqslant \frac{p_Y}{c \cdot q_Y}$ zwróć $X = Y$, w przeciwnym razie wróć do 1.)

Przykład:
Korzystając z powyższej metody generujemy 10^3 realizacji dyskretnej zmiennej losowej X o rozkładzie $P(X=1) = 0.2, P(X=2) = 0.05, P(X=3) = 0.35, P(X=4) = 0.1, P(X=5) = 0.18, P(X=6) = 0.12$. Jako rozkład pomocniczy bierzemy rozkład jednostajny oraz wyznaczamy, że $c = \frac{0.35}{\frac{1}{6}} = 2.1$.



Rysunek 3

2.2.2. Rozkład ciągły

Założenia:

— potrafimy efektywnie generować zmienną losową Y o gęstości $g(x)$;
— X i Y mają te same zbiory wartości
— potrafimy wyznaczyć stałą $0 < c < \infty$, t.że $\sup_{x \in \mathbb{R}} \frac{f(x)}{g(x)} \leqslant c$.

Algorytm:

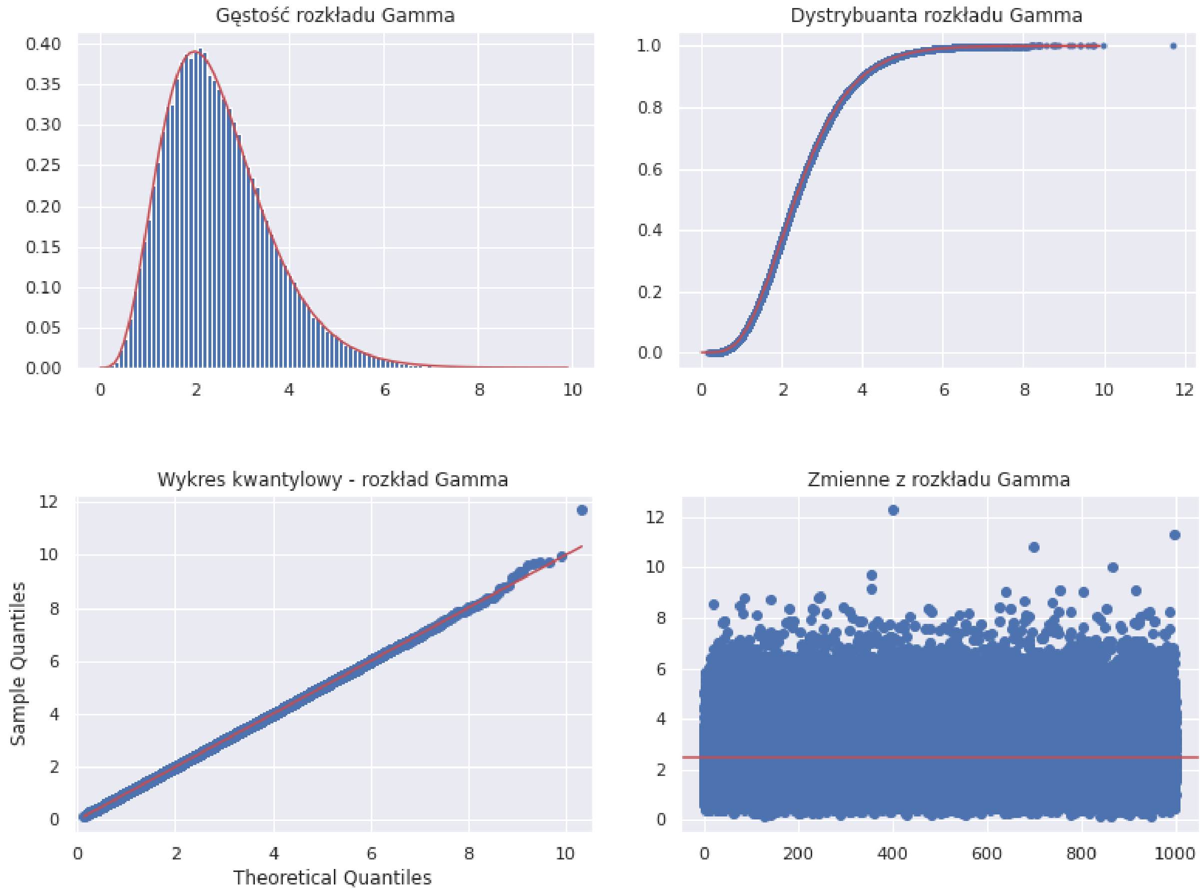
1.) generuj Y
2.) generuj $U \sim \mathcal{U}(0,1), U \perp Y$
3.) jeśli $U \leqslant \frac{f(Y)}{c \cdot g(Y)}$ zwróć $X = Y$, w przeciwnym razie wróć do 1.)

Szczególny przypadek: Gdy X ma ograniczoną gęstość o wartościach dodatnich tylko na odcinku $[a,b]$, to powyższy algorytm może wyglądać w następujący sposób:

1.) generuj $U_1 \sim \mathcal{U}(a,b)$ i $U_2 \sim \mathcal{U}(0,m)$, $U_1 \perp U_2$, gdzie $m = \max_{x \in [a,b]} f(x)$
2.) jeśli $f(U_1) \geqslant U_2$ zwróć $X = U_1$, w przeciwnym razie wróć do 1.)

Powyższe algorytmy generują zmienną losową X o gęstości $f(x)$

Przykład:
Generujemy realizacje zmiennej losowej X z rozkładu Gamma, $X \sim \Gamma(5, 2)$ o gęstości $f(x) = 2e^{-2x} \cdot \frac{(2x)^4}{\Gamma(5)} = \frac{4}{3}e^{-2x}x^4$ dla $x > 0$. Jako rozkład pomocniczy bierzemy rozkład wykładniczy z parametrem $\lambda = 4$. Wartość c została wyliczona w programie i wynosi około 2.67.



Rysunek 4

Na ostatnim wykresie zaznaczona została empiryczna wartość średnia, która wynosi około 2.5 i zgadza się z wartością teoretyczną równą $\frac{\alpha}{\lambda} = \frac{5}{2} = 2.5$.

2.3. Metoda splotowa

Metoda ta służy, do generowania zmiennych losowych, które powstają poprzez sumowanie innych zmiennych (gęstość sumy niezależnych zmiennych losowych nazywamy splotem funkcji).

Założenia:

— zmienna losowa X może byc zapisana w postaci $X = Y_1 + \ldots + Y_n$;
— Y_i są niezależnymi zmiennymi losowymi;
— potrafimy generować Y_i.

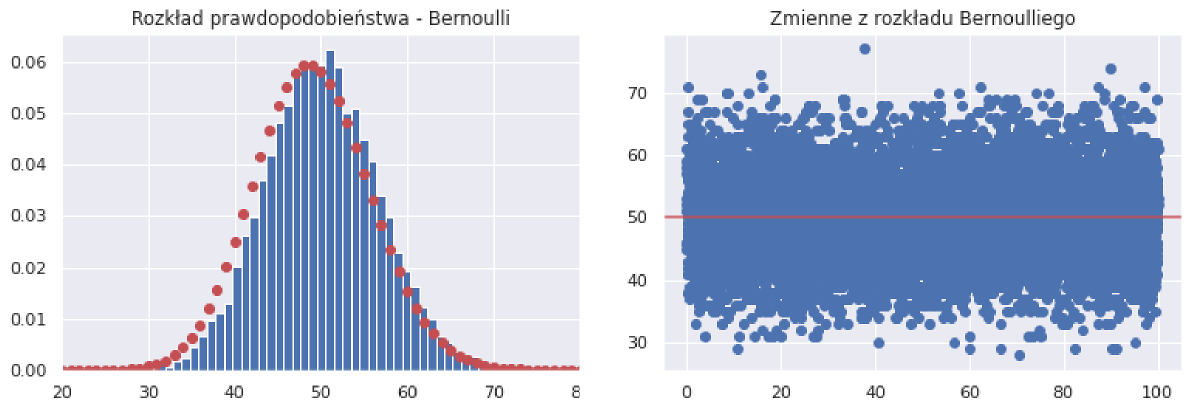
Algorytm:

1.) generuj niezależne $Y_1, \ldots, Y_n$
2.) wstaw $X = Y_1 + \ldots + Y_n$

2.3.1. Rozkład dyskretny

Przykład:
Generujemy realizacje zmiennej X z rozkładu dwumianowego $X \sim \mathcal{B}(n,p)$, gdzie $n = 500$, $p = 0.1$ oraz $P(X = i) = \binom{n}{i} p^i (1-p)^{n-i}, i = 0, 1, \ldots, n$. Y_i to wynik i-tej próby Bernoulliego. Algorytm w tym przypadku wygląda następująco:

1.) generuj niezależne $U_i, U_i \sim \mathcal{U}(0,1)$
2.) wstaw $X = (U_1 \leqslant p) + \ldots + (U_n \leqslant p)$

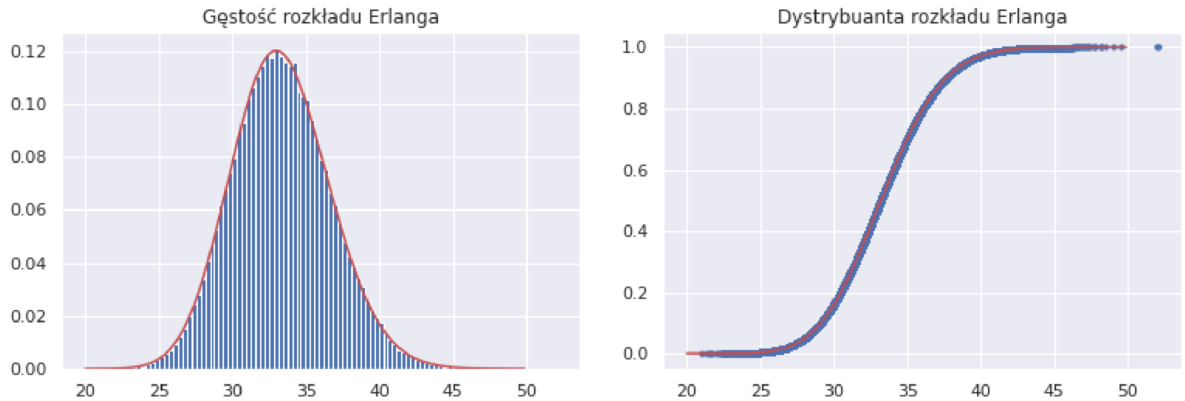


Rysunek 5

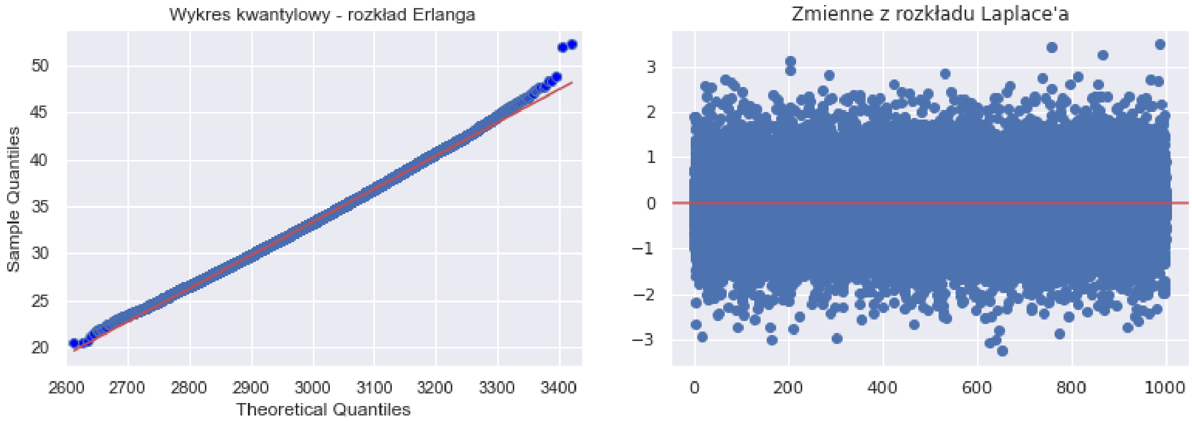
Na ostatnim wykresie zaznaczona została empiryczna wartość średnia, która wynosi 500 i zgadza się z wartością teoretyczną równą $np = 500 \cdot 0.1 = 50$.

2.3.2. Rozkład ciągły

Przykład:
Generujemy realizacje zmiennej losowej X z rozkładu Erlanga, $X \sim \Gamma(100, 3)$ o gęstości $f(x) = \frac{(\lambda x)^{n-1}}{(n-1)!} \lambda e^{-\lambda x}$ dla $x > 0$. Y_i mają rozkład wykładniczy z parametrem λ.



Rysunek 6



Rysunek 7

Na ostatnim wykresie zaznaczona została empiryczna wartość średnia, która wynosi około 33.3, co zgadza się z wartością teoretyczną równą $\frac{100}{3} \approx 33.3$.

2.4. Metoda mieszaniny rozkładów (kompozycji)

2.4.1. Rozkład ciągły

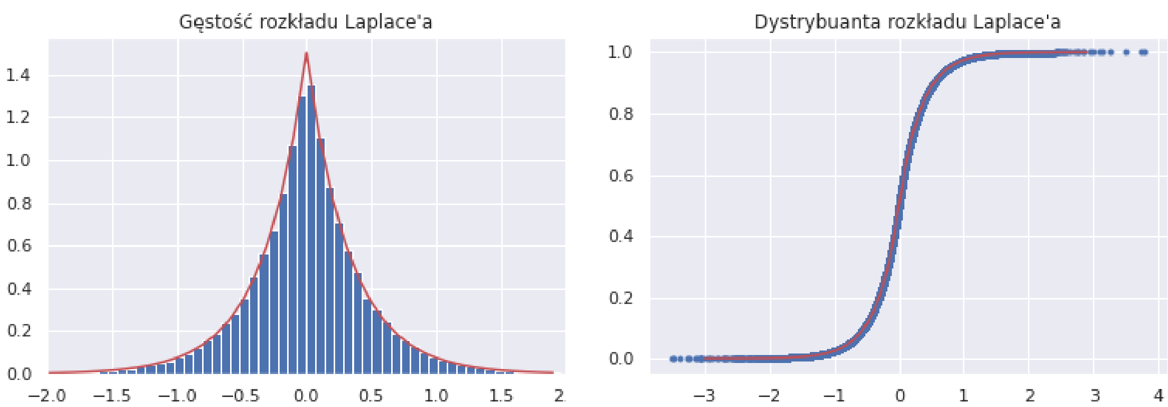
Założenia:

- zmienna losowa X ma dystrybuantę $F_X(x) = \sum_{i=1}^{n} p_i F_i(x)$, gdzie $p_i > 0, \sum p_i = 1$ oraz F_i to dystrybuanty pewnych zmiennych losowych Y_i
- potrafimy generować Y_i

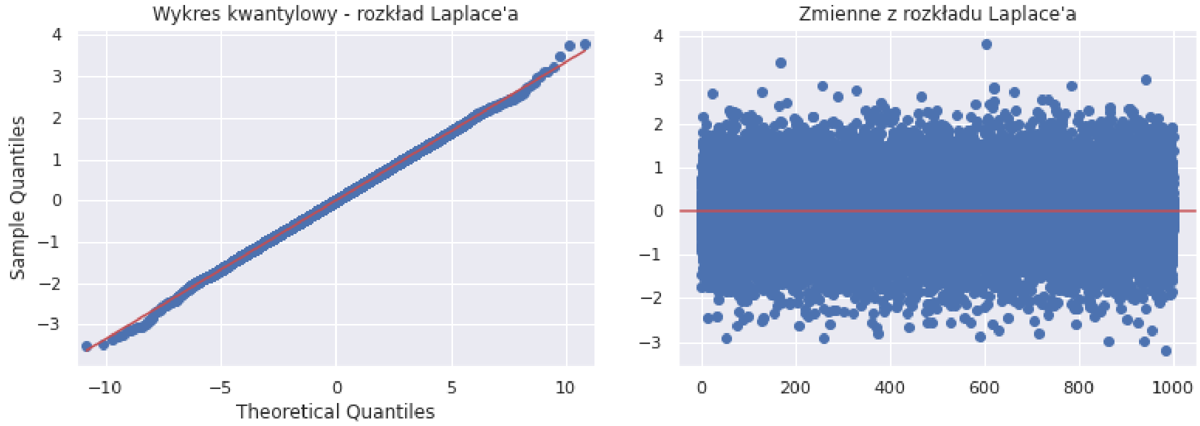
Algorytm [1]:

1.) generuj zmienną losową I o rozkładzie $P(I = i) = p_i, i \in \mathbb{N}, I \perp Y_i$
2.) generuj zmienną losową Y_I
3.) wstaw $X = Y_I$

Przykład:
Generujemy realizacje zmiennej losowej z rozkładu Laplace'a [2], $X \sim \mathcal{L}(0, 3)$ o gęstości $f(x) = \frac{3}{2} e^{-3|x|}$.



Rysunek 8



Rysunek 9

Na ostatnim wykresie zaznaczona została empiryczna wartość średnia, która wynosi około 0, co zgadza się z wartością teoretyczną równą parametrowi m, czyli 0.

3. Metody generowania rozkładu normalnego

Będziemy generować zmienne z rozkładu normalnego standardowego $X_0 \sim \mathcal{N}(0,1)$. Wtedy, aby wygenerować rozkład o średniej μ i wariancji σ stosujemy wzór $X = \sigma X_0 + \mu$

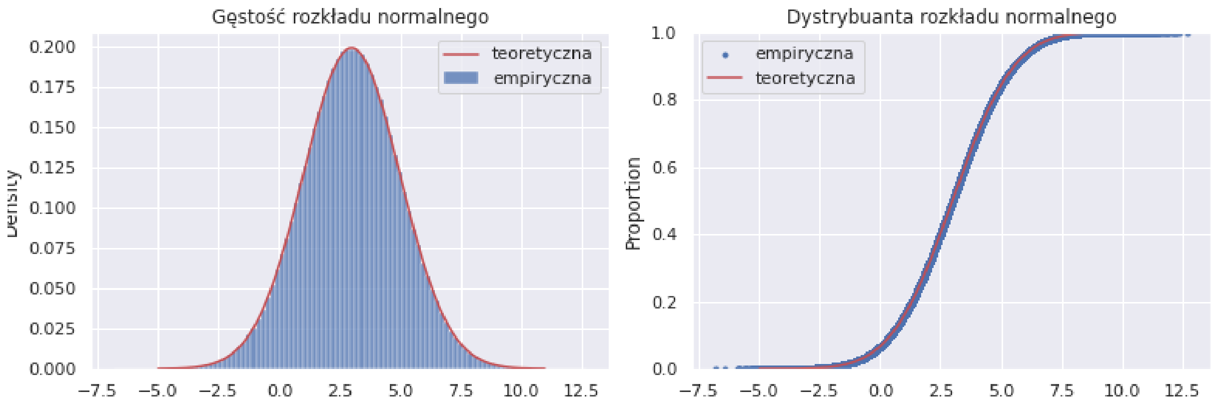
3.1. Metoda Boxa–Mullera

Algorytm:

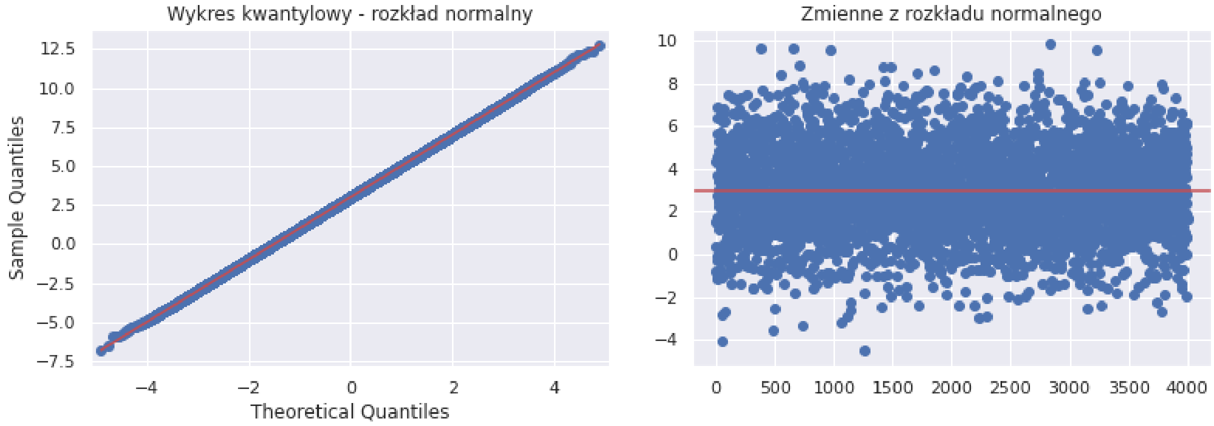
1.) generuj niezależne U_1, U_2, gdzie $U_i \sim \mathcal{U}(0,1)$
2.) wstaw $X = \sqrt{-2\ln U_1}\cos(2\pi U_2)$ oraz $Y = \sqrt{-2\ln U_1}\sin(2\pi U_2)$

Wtedy X i Y są niezależne oraz $X, Y \sim \mathcal{N}(0,1)$.

Przykład:
Generujemy zmienną z rozkładu normalnego o średniej $\mu = 3$ i wariancji $\sigma = 2$, $X \sim \mathcal{N}(3,2)$



Rysunek 10



Rysunek 11

Na ostatnim wykresie zaznaczona została empiryczna wartość średnia, która wynosi około 3, co zgadza się z wartością teoretyczną równą parametrowi μ.

3.2. Metoda biegunowa

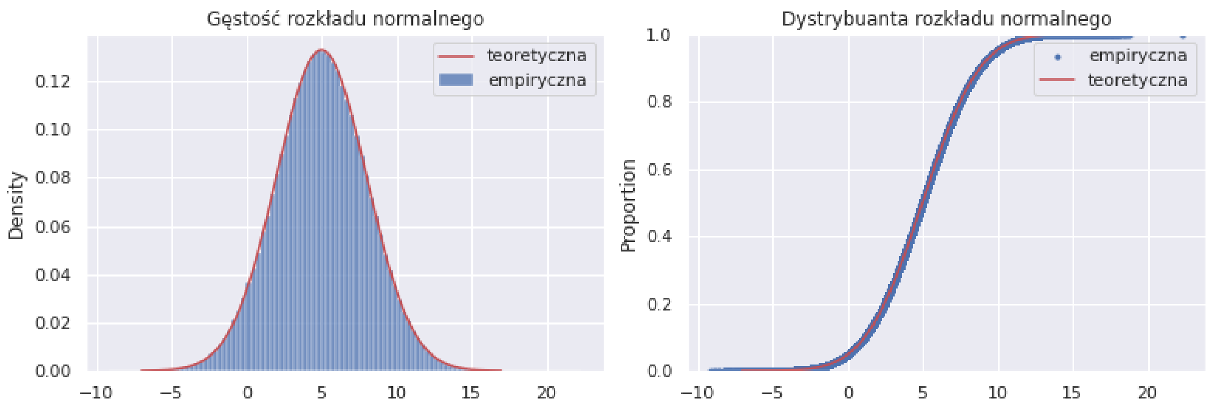
Chcemy pozbyć się sin i cos z metody Boxa-Mullera. W tym celu generujemy wektor (V_1, V_2) o rozkładzie jednostajnym w kole jednostkowym, gdzie $V_1 = R\cos\alpha, V_2 = R\sin\alpha$.

Algorytm:

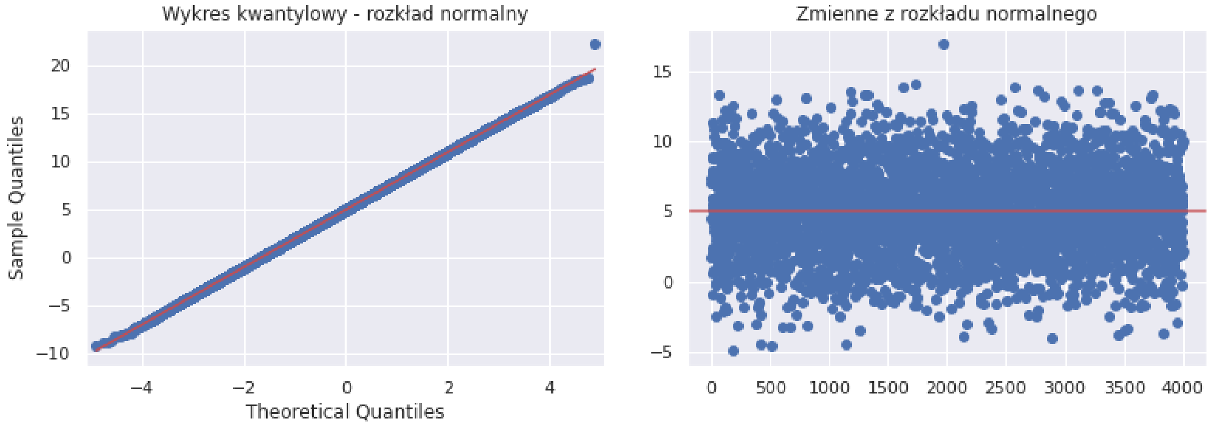
1.) generuj niezależne $V_1, V_2 \sim U(-1, 1)$
2.) wyznacz $R^2 = V_1^2 + V_2^2$
3.) jeśli $R^2 > 1$ wróć do 1.), w przeciwnym wypadku wstaw
$X = \sqrt{\frac{-2\ln R^2}{R^2}} V_1$ oraz $Y = \sqrt{\frac{-2\ln R^2}{R^2}} V_2$

Wtedy X i Y są niezależne oraz $X, Y \sim \mathcal{N}(0, 1)$.

Przykład:
Ponownie generujemy rozkład normalny, jednak tym razem o średniej $\mu = 5$ i wariancji $\sigma = 3$, $X \sim \mathcal{N}(5, 3)$.



Rysunek 12



Rysunek 13

Na ostatnim wykresie zaznaczona została empiryczna wartość średnia, która wynosi około 5, co zgadza się z wartością teoretyczną równą parametrowi μ.

4. Liniowy generator kongruentny

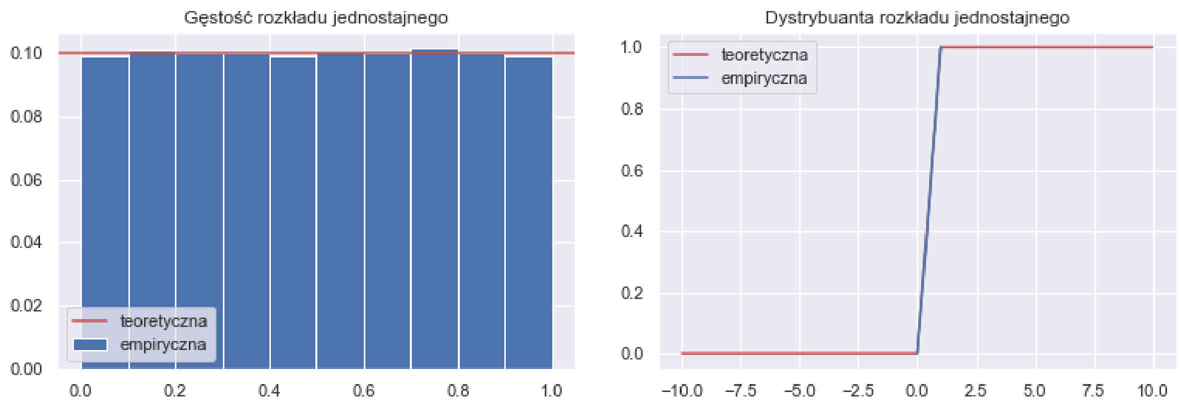
4.1. Opis

W większości symulacji komputerowych ciężko jest generować całkowicie losowe liczby. Zamiast tego korzystamy z liczb *pseudolosowych*, które tworzymy używając określonych algorytmów. Na pierwszy rzut oka wyglądają na przypadkowe, jednak są owocem pewngo schematu.

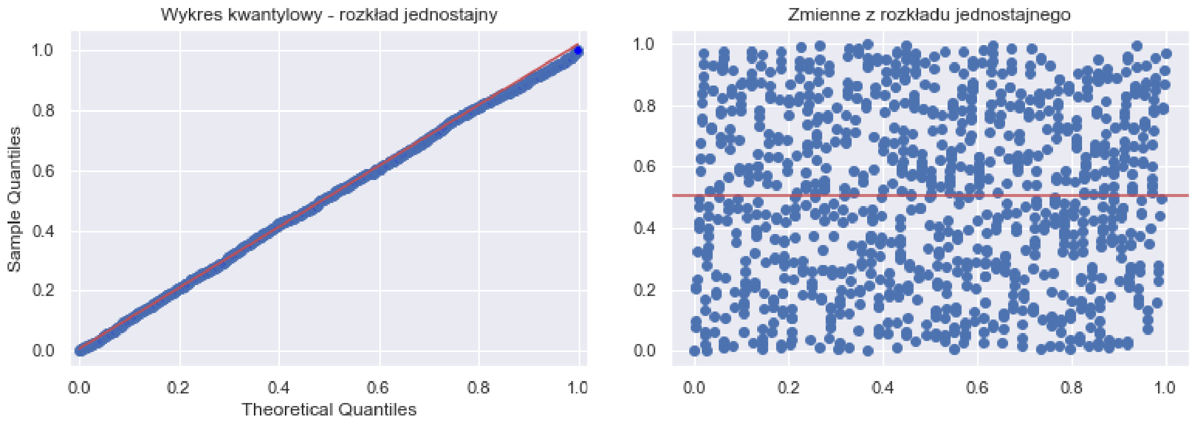
Liniowy generator kongruentny jest jednym z najbardziej powszechnych i najstarszym algorytmem generującym liczby *pseudolosowe*. Może on stworzyć wiele liczb w krótkim czasie, więc jest korzystnym wyborem w symulacjach, w których potrzebujemy wiele wartości. [3]

4.2. Przykład

Możemy zdefiniować następujący generator liczb pseudolosowych: $Z_{i+1} = (a \cdot Z_i) \mod m$, gdzie Z_0 to ziarno, $a = 7^5$, $m = 2^{31} - 1$ $U_i := \frac{Z_i}{m}$ Wtedy nasza wygenerowana zmienna będzie miała rozkład jednostajny $U_i(0, 1)$. [3]



Rysunek 14



Rysunek 15

Literatura

[1] https://mst.mimuw.edu.pl/lecture.php?lecture=sst&part=Ch3#S4

[2] http://prac.im.pwr.wroc.pl/~agniesz/rachunek_prawd_MAT1332/files/RPr_MAT1332_wyklad_6a_rozklady_probabilistyczne_prezentacja.pdf

[3] https://www.math.arizona.edu/~tgk/mc/book_chap3.pdf