Symulacje Komputerowe

Raport: **1**

Temat sprawozdania **Generowanie zmiennych losowych**
Nazwisko i Imię prowadzącego kurs **dr Michał Balcerek**

Wykonawca:	
Imię i Nazwisko, nr indeksu	Adrianna Ziobroniewicz, 262227 Mateusz Stasiak
Wydział	Wydział matematyki, W13
Termin zajęć:	Wtorek, 11^{15}
Numer grupy ćwiczeniowej	T00-70
Data oddanie sprawozdania:	4 stycznia 2023
Ocena końcowa	

Adnotacje dotyczące wymaganych poprawek oraz daty otrzymania poprawionego sprawozdania

1. Metoda odwrotnej dystrybuanty

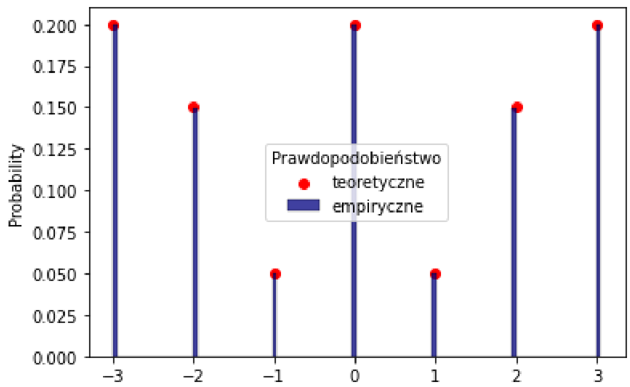
1.1. Rozkład dyskretny

Algorytm:
1. Generujemy $U \sim U(0,1)$.
2. Wyznaczamy $j \in N$, takie że

$$\sum_{i=1}^{j-1} p_i < U \leq \sum_{i=1}^{j} p_i \iff F_X\left(x_{j-1}\right) < U \leq F_X\left(x_j\right).$$

3. Zwróć $X = x_j$.

Niech X będzie dyskretną zmienną losową o rozkładzie $P\left(X = -3\right) = P\left(X = 3\right) = 0.2$, $P\left(X = -2\right) = P\left(X = 2\right) = 0.15$, $P\left(X = -1\right) = P\left(X = 1\right) = 0.05$, $P\left(X = 0\right) = 0.2$.



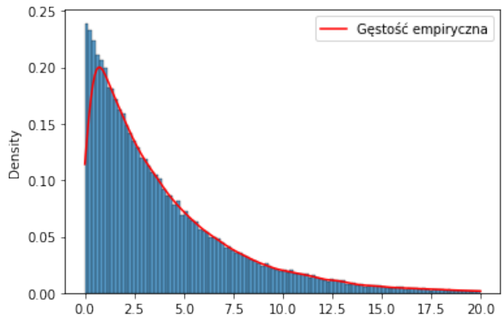
Rysunek 1. Rozkład prawdopodobieństwa

Rozkład prawdopodobieństwa dla wygenerowanego wektora zmiennych losowych jest bardzo zbliżony do rozkładu teoretycznego.

1.2. Rozkład ciągły

Algorytm:
1. Generujemy $U \sim U(0,1)$.
2. Wstawiamy $X = F_X^{-1}(U)$.

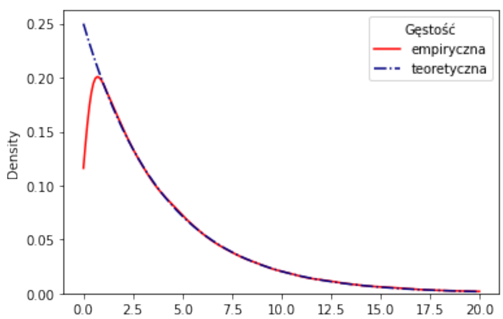
Chcemy wygenerować zmienne z rozkładu $\mathrm{Exp}\left(\frac{1}{4}\right)$. Stwórzmy wektor długości 10^5 zawierający zmienne losowe z rozkładu jednostajnego na przedziale $(0,1)$. Następnie odwracamy dystrybuantę F_X i wstawiamy każdy element powyższego wektora jako argument otrzymanej funkcji.



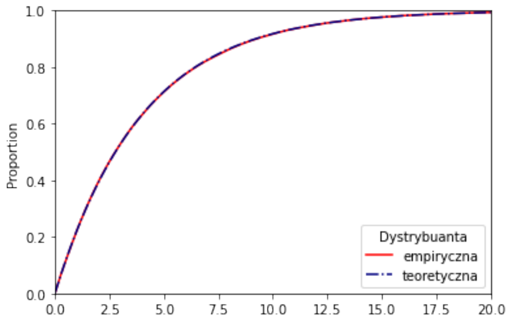
(a) Histogram wektora zmiennych

	Wartości empiryczne	Wartości teoretyczne
count	100000.000000	100000.000000
mean	4.004558	4.000000
std	4.015639	4.000000
min	0.000034	NaN
25%	1.144984	1.150728
50%	2.764760	2.772589
75%	5.541320	5.545177
max	44.237265	NaN

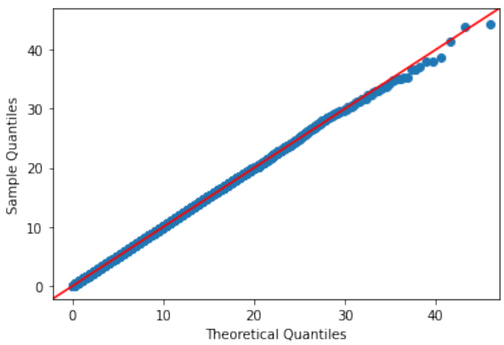
(b) Najważniejsze statystyki opisowe



(c) Porównanie funkcji gęstości



(d) Porównanie dystrybuant



(e) Wykres kwantylowy

Kształt histogramu odpowiada funkcji gęstości empirycznej. Wartości najważniejszych statystyk opisowych próby są zbliżone do wartości teoretycznych. Wykresy gęstości i dystrybuant się pokrywają, a wykres kwantylowy z argumentem $\text{Exp}\left(\frac{1}{4}\right)$ stanowi linię prostą nachyloną do osi OX pod kątem 45 stopni.

2. Metoda Boxa-Müllera

Służy do generowania zmiennych z rozkładu normalnego.

$$X \sim N(\mu,\sigma) \qquad\qquad f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

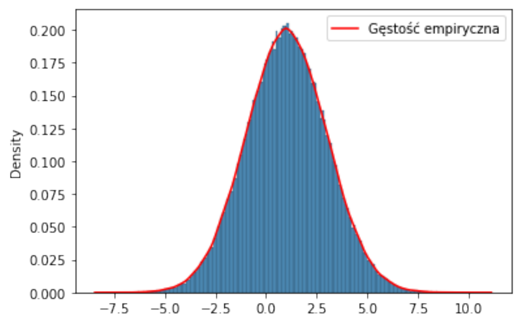
Weźmy $X_0 \sim N(0,1)$, wtedy $X \overset{\mathrm{d}}{=} \sigma X_0 + \mu$.

Algorytm:
1. Generujemy U_1, U_2 iid, $U_i \sim U(0,1)$.
2. Wstawiamy

$$X \overset{\text{def}}{=} \sqrt{-2\ln U_1} \cdot \cos(2\pi U_2)$$
$$Y \overset{\text{def}}{=} \sqrt{-2\ln U_1} \cdot \sin(2\pi U_2).$$

Wtedy $X \perp\!\!\!\perp Y$ oraz $X,Y \sim N(0,1)$.

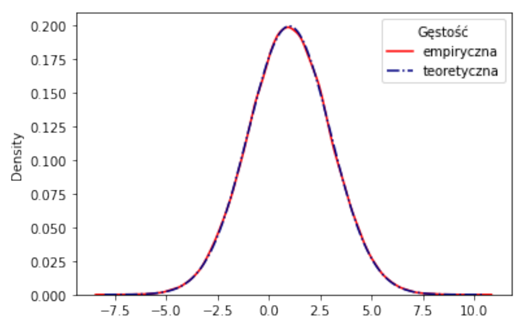
Chcemy wygenerować zmienne z rozkładu $N(1,2)$.



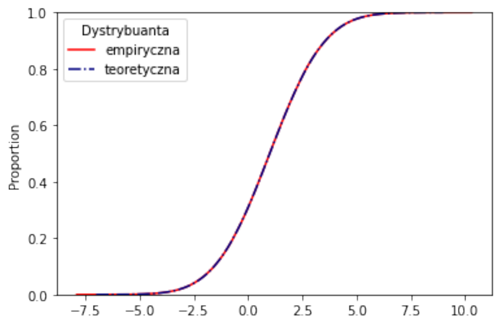
(f) Histogram wektora zmiennych

	Wartości empiryczne	Wartości teoretyczne
count	200000.000000	200000.00
mean	0.999033	1.00
std	2.002314	2.00
min	-7.903699	NaN
25%	-0.350898	-0.35
50%	0.994261	1.00
75%	2.350894	2.35
max	10.318967	NaN

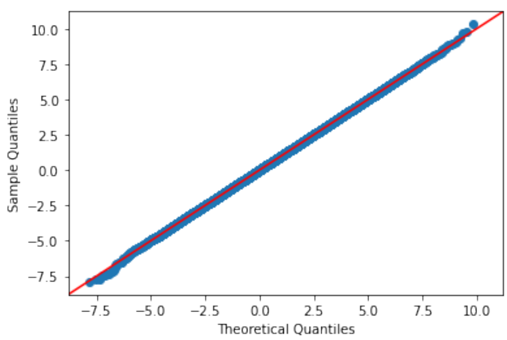
(g) Najważniejsze statystyki opisowe



(h) Porównanie funkcji gęstości



(i) Porównanie dystrybuant



Kształt histogramu odpowiada funkcji gęstości empirycznej. Wartości najważniejszych statystyk opisowych próby są zbliżone do wartości teoretycznych. Wykresy gęstości i dystrybuant się pokrywają, a wykres kwantylowy z argumentem $N(1{,}2)$ stanowi linię prostą nachyloną do osi OX pod kątem 45 stopni.

3. Metoda biegunowa Boxa-Müllera

Metoda Boxa-Müllera jest mało wydajna. Obliczanie powyższych funkcji trygonometrycznych oparte jest na aproksymacji wielomianami. Możemy przekształcić tę metodę za pomocą współrzędnych biegunowych.

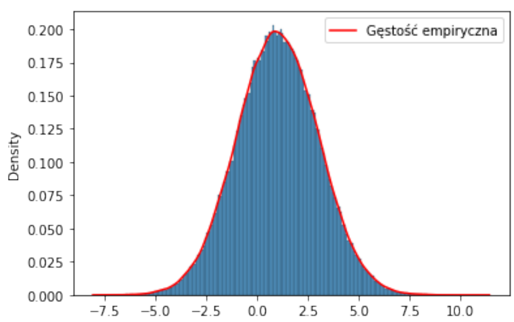
Załóżmy, że wektor losowy (V_1,V_2) ma rozkład jednostajny w kole jednostkowym. $R^2 \to U_1$

$$X \overset{\text{def}}{=} \sqrt{-2\ln U_1} \cdot \cos(2\pi U_2) \overset{\text{d}}{=} \sqrt{-2\ln R^2} \cdot \cos(\alpha) \overset{\text{d}}{=} \sqrt{-2\ln R^2} \cdot \frac{V_1}{R}$$

$$Y \overset{\text{def}}{=} \sqrt{-2\ln U_1} \cdot \sin(2\pi U_2) \overset{\text{d}}{=} \sqrt{-2\ln R^2} \cdot \sin(\alpha) \overset{\text{d}}{=} \sqrt{-2\ln R^2} \cdot \frac{V_2}{R}.$$

Algorytm:
1. Generujemy V_1, V_2 iid, $V_i \sim U(-1{,}1)$.
2. Wyznaczamy $R^2 = V_1^2 + V_2^2$.
3. Jeśli $R^2 > 1$ to wracamy do 1.
4. Wstawiamy
$$X \overset{\text{def}}{=} \sqrt{-2\ln R^2} \cdot \frac{V_1}{R}$$
$$Y \overset{\text{def}}{=} \sqrt{-2\ln R^2} \cdot \frac{V_2}{R}.$$
   Wtedy $X \perp\!\!\!\perp Y$ oraz $X,Y \sim N(0{,}1)$.

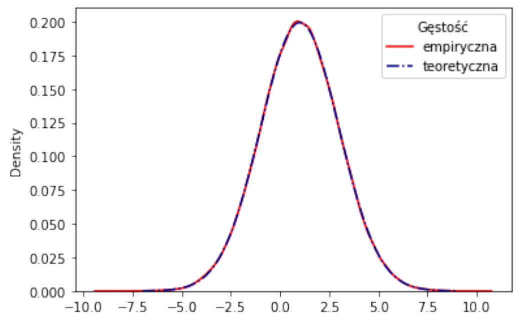
Chcemy wygenerować zmienne z rozkładu $N(1{,}2)$.



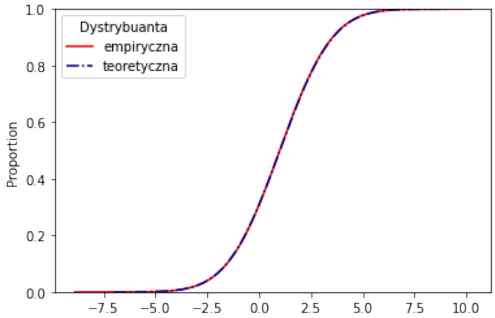
(j) Histogram wektora zmiennych

	Wartości empiryczne	Wartości teoretyczne
count	200000.000000	200000.00
mean	0.998047	1.00
std	1.994536	2.00
min	-8.898640	NaN
25%	-0.346410	-0.35
50%	0.996731	1.00
75%	2.339869	2.35
max	10.222987	NaN

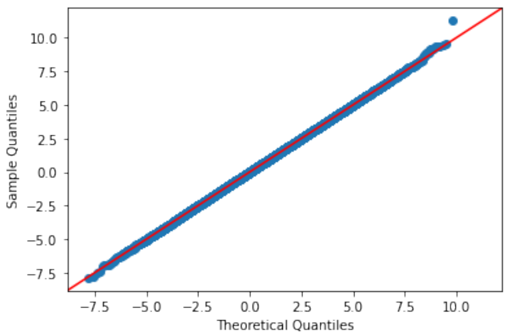
(k) Najważniejsze statystyki opisowe



(l) Porównanie funkcji gęstości



(m) Porównanie dystrybuant



(n) Wykres kwantylowy

Kształt histogramu odpowiada funkcji gęstości empirycznej. Wartości najważniejszych statystyk opisowych próby są zbliżone do wartości teoretycznych. Wykresy gęstości i dystrybuant się pokrywają, a wykres kwantylowy z argumentem $N(1,2)$ stanowi linię prostą nachyloną do osi OX pod kątem 45 stopni.

4. Metoda akceptacji-odrzucenia

Metoda akceptacji-odrzucenia polega na generowania zmiennej losowej dyskretnej X o rozkładzie

$$p_i = P(X = i), i = 1,2,\ldots$$

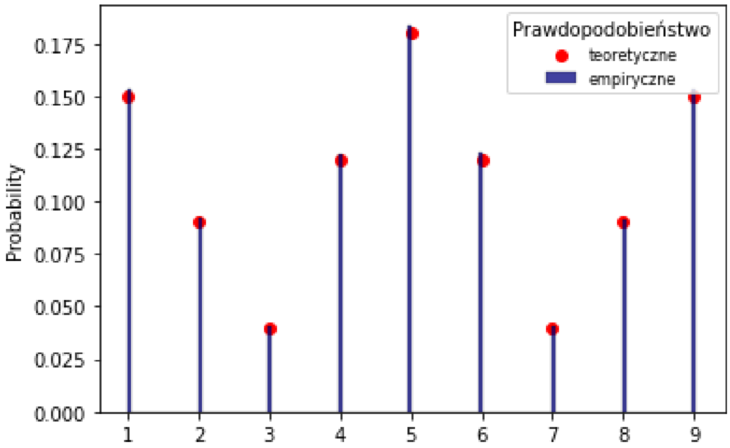
$$\sum_i p_i = 1, p_1 \geqslant 0$$

.

4.1. Algorytm - rozkład dyskretny

1. Generujemy jedną realizację Y.
2. Generujemy $U \sim U(0,1)$ U niezależne od Y.
3. Jeśli $U \leqslant \frac{P_x}{cq_x}$ zwróć $X = Y$ w przeciwnym razie wróć do 1.

Wykonaliśmy próbę dla $P(X = 1) = 0.15, P(X = 2) = 0.09, P(X = 3) = 0.04, P(X = 4) = 0.12, P(X = 5) = 0.18, P(X = 6) = 0.12, P(X = 7) = 0.04, P(X = 8) = 0.09, P(X = 9) = 0{,}15$.



Rysunek 2. Rozkład prawdopodobieństwa

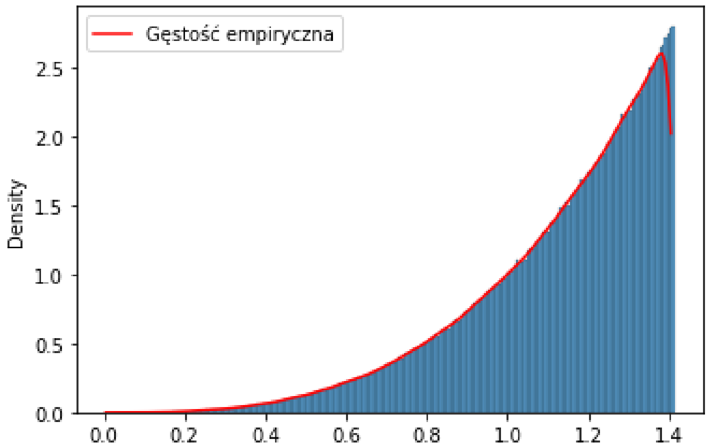
Rozkład prawdopodobieństwa dla wygenerowanego wektora zmiennych losowych jest bardzo zbliżony do rozkładu teoretycznego.

4.2. Algorytm - rozkłady ciągłe

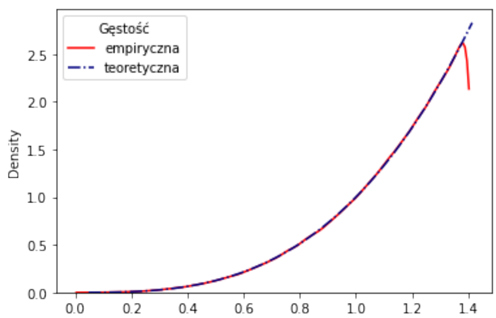
Prawdopodobieństwo sukcesu pojedynczej iteracji algorytmu wynosi $\frac{1}{c}$. Liczba powtórzeń algorytmu ma rozkład Geom$(\frac{1}{c})$. Średnia liczba powtórzeń algorytmu to c.

Algorytm:
1. Generujemy $U_1 \sim U(a,b)$ i $U_2 \sim U(O,m)$, U_1 niezależne od U_2.
2. Jeśli $f(U_1) \geqslant U_2$ zwracamy $X = U_1$, w przeciwnym razie wróć do podpunktu 1.

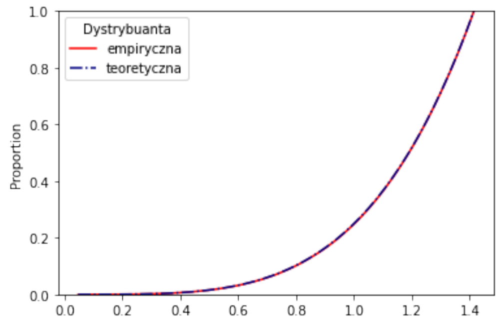
Wygenerowaliśmy próbę dla $f(x) = x^3$ oraz $g(x) = \sqrt[2]{\frac{1}{2}}$.



Rysunek 3. Histogram wektora zmiennych



(a) Porównanie funkcji gęstości



(b) Porównanie dystrybuant

Kształt histogramu odpowiada funkcji gęstości empirycznej, a wykresy gęstości i dystrybuant się pokrywają.

5. Metoda splotowa

Metoda ta pozwala wygenerować pewną zmienną losową X, przy pomocy innych zmiennych, które potrafimy efektywnie generować, i zsumowaniu ich. Załóżmy, że

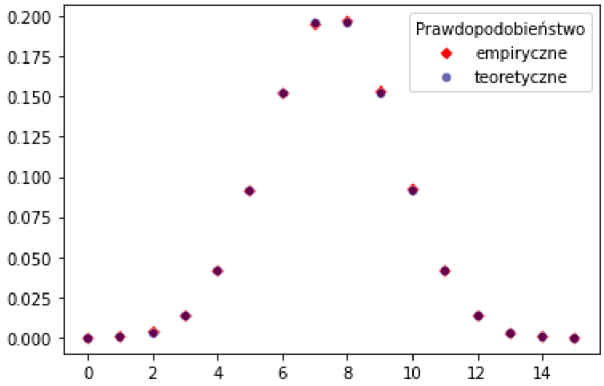
$$X \overset{d}{=} Y_1 + Y_2 + \cdots + Y_n,$$

gdzie Y_i to zmienne losowe niezależne.

5.1. Algorytm - rozkład dyskretny

1. Generujemy $U_1, \ldots, U_n$, iid. $U_1 \sim (0,1)$
2. Wstaw $X = (U_1 \leqslant p) + \ldots + (U_n \leqslant p)$

Aby zobaczyć wyniki działań skorzystaliśmy z rozkładu Bernoulliego.

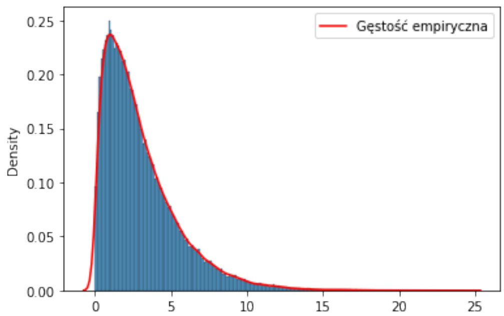


Rysunek 4. Prawdopodobieństwo zmiennych losowych

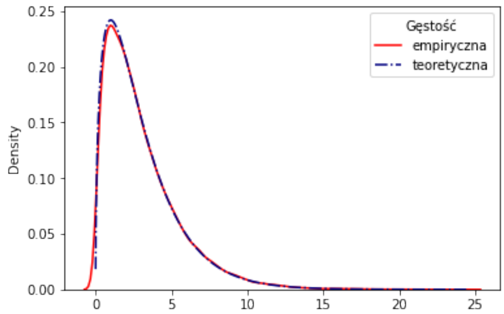
5.2. Algorytm - rozkład ciągły

1. Generujemy $Y_1, Y_2, \ldots, Y_n$.
2. Zwracamy $X = Y_1 + Y_2 + \cdots + Y_n$.

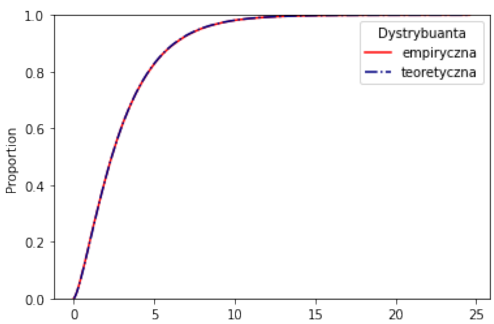
Nasze wyniki zostały uzyskane za pomocą rozkładu chi-kwadrat, $X_i \sim N(0{,}1)$, o 3 stopniach swobody.



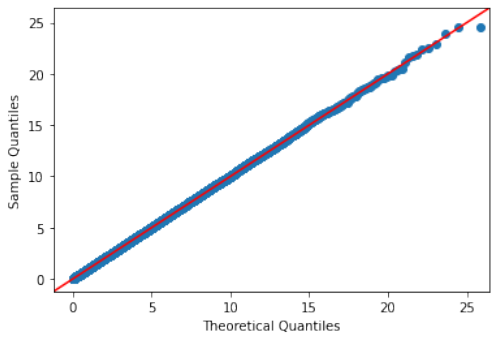
(a) Histogram wektora zmiennych



(b) Porównanie funkcji gęstości



(c) Porównanie dystrybuant



(d) Wykres kwantylowy

Kształt histogramu odpowiada funkcji gęstości empirycznej. Wykresy gęstości i dystrybuant się pokrywają, a wykres kwantylowy z argumentem chi^2(3) stanowi linię prostą nachyloną do osi OX pod kątem 45 stopni.

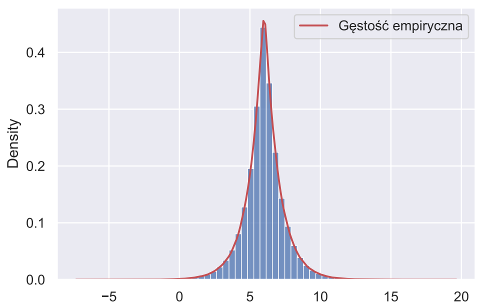
6. Metoda kompozycji

Załóżmy, że X ma dystrybuantę postaci $F_x(x) = \sum_{i=1}^{n} p_i F_i(x)$, gdzie $p_i > 0$, $\sum_{i=1}^{n} p_i = 1$ oraz F_i to dystrybunaty pewnych zmiennych losowych Y_i.

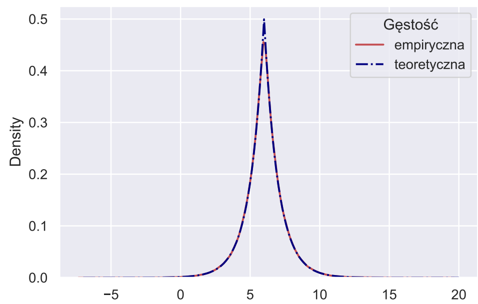
6.1. Algorytm - metoda kompozycji

1. Generujemy zmienną losową I o rozkładzie $P(I = i) = p_i, \quad i = 1,2,...,n$ $I \perp\!\!\!\perp Y_i$.
2. Generujemy zmienną losową Y_I.
3. Wstaw $X = Y_I$.

Wykonaliśmy metodę dla rozkładu ciągłego dla rozkładu ciągłego Laplace'a z parametrem $\lambda = \frac{1}{b}$, gdzie $b = 1$.



(e) Histogram wektora zmiennych



(f) Porównanie funkcji gęstości

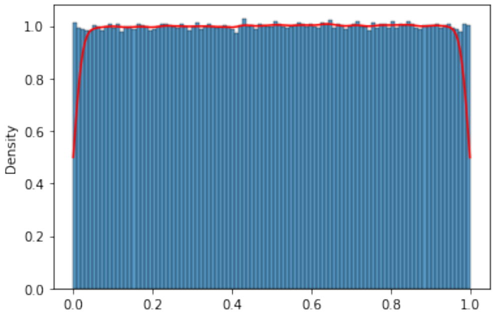
Kształt histogramu odpowiada funkcji gęstości empirycznej, a wykresy gęstości się pokrywają.

7. Liniowy generator kongruentny

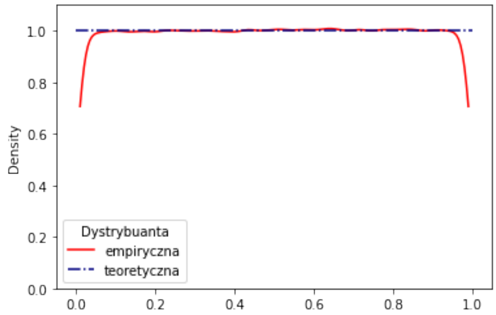
Liczba losowa jest, to liczba r należącą do pewnego zbioru wartości $r_1, ..., r_n$ wybieranych z pewnym prawdopodobieństwem, z kolei liczby pseudolosowe wyglądają jak losowe, lecz tworzy się je algorytmicznie. Oznacza to, iż znając wzór generacyjny oraz kilka kolejnych liczb pseudolosowych możemy bez problemu wygenerować wszystkie dalsze - tej cechy nie posiadają liczby losowe, w przeciwnym razie totolotek straciłby sens.
Liniowy generator kongruentny zaproponowany przez D. H. Lehmera w 1951 roku. Większość dzisiejszych procedur do generowania liczb pseudolosowych jest oparta na tym generatorze.

$$x_{n+1} = (ax_n + b)(\text{mod} \quad m)$$

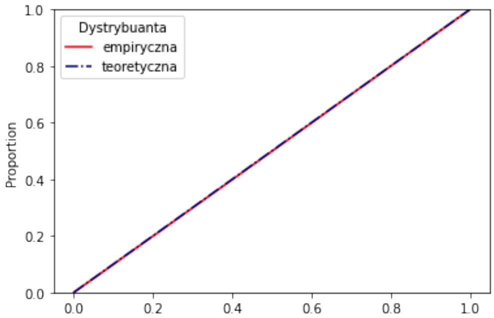
Zwykle wybieramy m najbliższe największej liczbie całkowitej dostępnej dla komputera.



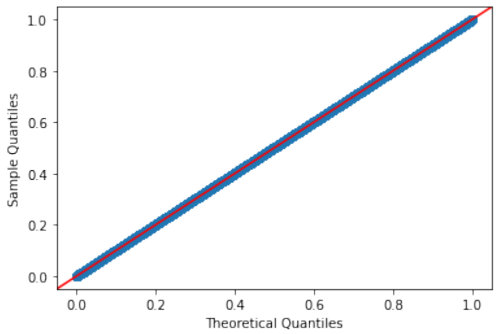
(g) Histogram wektora zmiennych



(h) Porównanie funkcji gęstości



(i) Porównanie dystrybuant



(j) Wykres kwantylowy

Kształt histogramu odpowiada funkcji gęstości empirycznej. Wykresy gęstości i dystrybuant się pokrywają, a wykres kwantylowy z argumentem $U(0{,}1)$ stanowi linię prostą nachyloną do osi OX pod kątem 45 stopni.

8. Wnioski

Za pomocą wyżej wymienionych metod jesteśmy w stanie efektywnie generować realizacje zmiennych losowych o różnych rozkładach. Pozwalają nam one w pewien sposób "obejść" praktyczną niemożność komputera do wykonywania operacji prawdziwie losowych. Wykresy dystrybuant, gęstości oraz wartości statystyk opisowych dla danych rozkładów możemy wysymulować na tyle dokładnie, ażeby móc wykorzystywać potencjał komputera do obliczeń, w których korzystamy z teorii rachunku prawdopodobieństwa.