

Raport nr 1 z przedmiotu:
"Symulacje komputerowe"
Temat: Generowanie zmiennych losowych różnymi metodami

Jakub Muzyka i Jacek Paździerkiewicz

April 28, 2022

Czym są liczby pseudolosowe?

Liczbami pseudolosowymi są liczby wyglądające jak losowe, które tworzy się algorytmicznie. Dzięki temu, znając tylko kilka kolejnych liczb pseudolosowych, możemy wygenerować wszystkie następne.

W naszym raporcie przedstawimy różne sposoby generowania zmiennych losowych:

1. Metoda odwrotnej dystrybucyjności dla rozkładów dyskretnych

Chcemy wygenerować wartość z rozkładu dyskretnego o prawdopodobieństwie

$$P(X = x_j) = p_j, \quad j = 0, 1, \dots, \sum_j p_j = 1$$

Żeby osiągnąć nasz cel, musimy generować zmienną $U \sim (0, 1)$ z rozkładu jednostajnego i odpowiednio zwracać wartość p_j , jeżeli U spełnia warunek nierówności:

$$\begin{cases} x_0 & \text{if } U < p_0 \\ x_1 & \text{if } p_0 \leq U < p_0 + p_1 \\ \vdots & \\ x_j & \text{if } \sum_{i=0}^{j-1} p_i \leq U < \sum_{i=0}^j p_i \\ \vdots & \end{cases}$$

Wiemy, że kiedy $0 < a < b < 1$ mamy $P(a \leq U < b) = b - a$, to:

$$P(X = x_j) = P\left(\sum_{i=0}^{j-1} p_i \leq U < \sum_{i=0}^j p_i\right) = p_j$$

Jest naszą szukaną dystrybucją.

Algorytm:

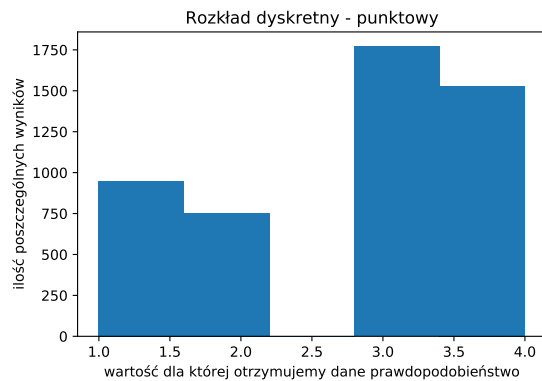
- (a) Generuj $U \sim (0, 1)$
- (b) Jeżeli $U < p_0$ zwróć $X = x_0$ i kończymy program
- (c) Jeżeli $U < p_0 + p_1$ zwróć $X = x_1$ i kończymy program
- (d) Jeżeli $U < p_0 + p_1 + p_2$ zwróć $X = x_2$ i kończymy program ... *etc.*

Przykładowa implementacja:

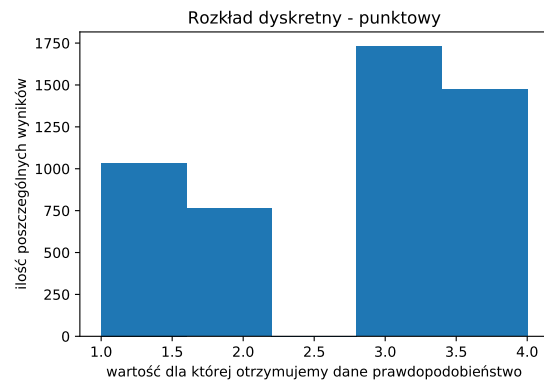
Weźmy rozkład dyskretny z konkretnymi prawdopodobieństwami:

$$P(X = 1) = 0.2, P(X = 2) = 0.15, P(X = 3) = 0.35, P(X = 4) = 0.3$$

Które zgodnie z definicją prawdopodobieństwa sumują się do 1.

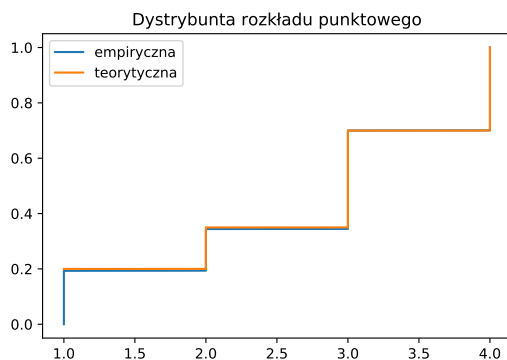


Wykres 1: dla 1000 losowań

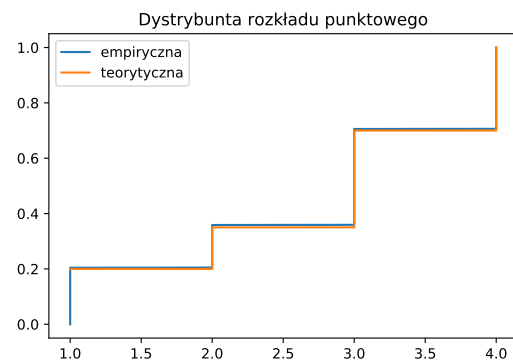


Wykres 2: dla 5000 losowań

Wykonujemy również test empiryczny dla dystrybucyj naszego rozkładu:



Wykres 3: dla 1000 losowań



Wykres 4: dla 5000 losowań

Widzimy, że zarówno dystrybucja empiryczna, jak i teoretyczna pokrywają się, a to oznacza, że nasza metoda odwrotnej dystrybucyj dla rozkładów dyskretnych jest poprawnie zaimplementowana.

2. Metoda odwrotnej dystrybucyj dla rozkładów ciągłych

Dążymy do tego, żeby wyprowadzić dystrybucję odwrotną ze wzoru $u = F(X)$:

$$X = F^{-1}(U); \quad U(0, 1)$$

Algorytm:

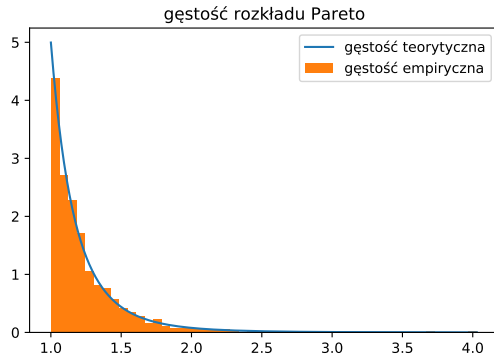
- (a) Generuj $U \sim (0, 1)$
- (b) Znajdź $x = F^{-1}(U)$ dla $u = F(x)$
- (c) Wstaw u do $x = F^{-1}(U)$ i zwróć wartość

Przykładowe implementacje: Użyjemy rozkładu Pareto, które gęstość i dystrybuanta wyglądają następująco:

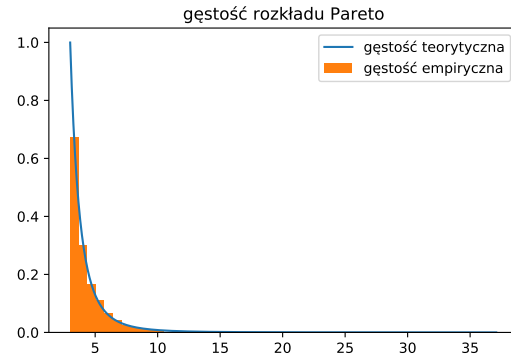
$$f(x) = \alpha \left(\frac{\lambda}{x}\right)^\alpha \frac{1}{x}$$

$$F(x) = 1 - \left(\frac{\lambda}{x}\right)^\alpha$$

Na początek porównujemy gęstość empiryczną i teoretyczną tego rozkładu:

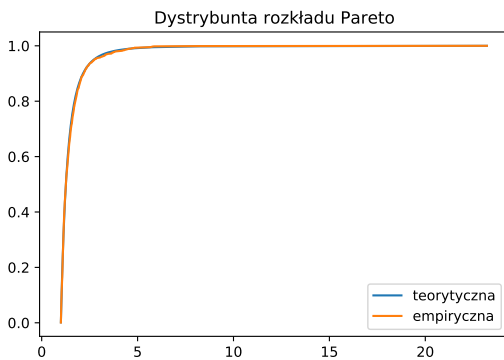


Wykres 5: dla $\alpha = 5, \lambda = 1$ i dla 1000 losowań

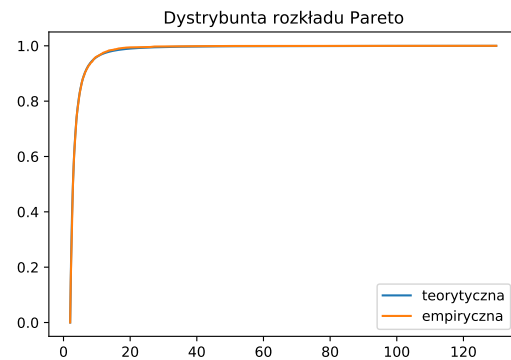


Wykres 6: dla $\alpha = 3, \lambda = 3$ i dla 3000 losowań

A następnie to samo dla dystrybuantry:



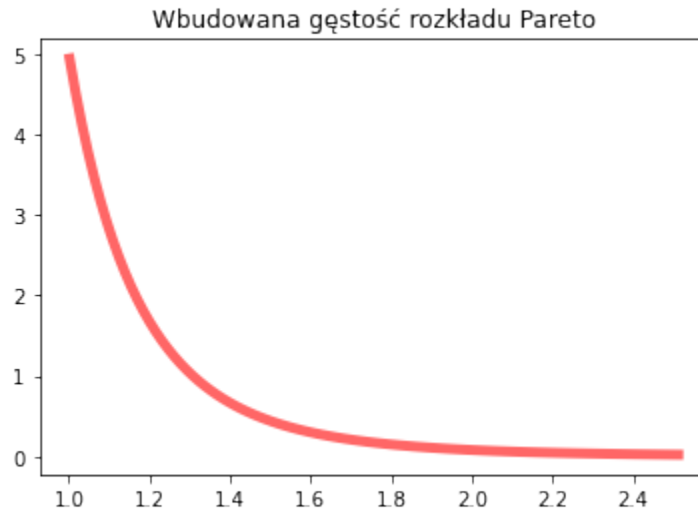
Wykres 7: dla $\alpha = 3, \lambda = 1$ i dla 1000 losowań



Wykres 8: dla $\alpha = 2, \lambda = 2$ i dla 3000 losowań

Dla sprawdzenia poprawności naszej metody implementujemy wbudowany rozkład Pareto:

Wnioski:



Możemy zauważyć, że zarówno gęstość teoretyczna i wbudowana rozkładu Pareto, przechodzą przez te same punkty. Dodatkowo dystrybucja empiryczna i teoretyczna, pokrywają się, co świadczy o poprawnej implementacji naszej metody.

3. Metoda kompozycji

Jeśli wiemy, że jesteśmy w stanie wyznaczyć dystrybuanty zmiennych losowych Y_i równych F_i , $i = 1, 2, 3, \dots$, to zakładamy, że zmienna X ma dystrybuantę postaci:

$$F_X(x) = \sum_{i=1}^n p_i F_i(x), \quad \text{gdzie } p_i > 0, \sum_{i=1}^n p_i = 1$$

oraz F_i , to dystrybuanty pewnych zmiennych losowych Y_i .

Jeśli X, Y_1, \dots, Y_n mają gęstości, to wyżej wymieniony wzór można równoważnie zapisać:

$$f_x(x) = \sum_{i=1}^n p_i f_i(x)$$

Algorytm dla rozkładu ciągłego:

- Generuj zmienną losową z rozkładu jednostajnego $U \sim (0, 1)$
- Przy pomocy metody odwrotnej dystrybuanty wyznacz wybrany rozkład jako funkcję z naszą losową wartością $U \sim (0, 1)$
- zmieniaj losowo znak funkcji
- zwróć wartość funkcji

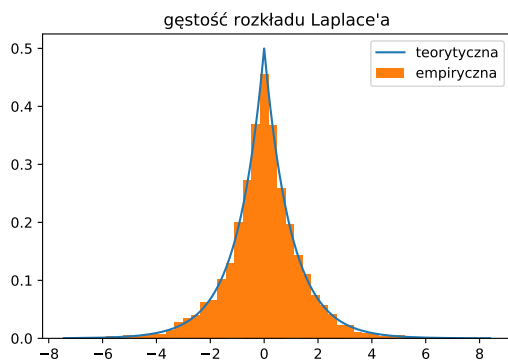
Przykładowe implementacje:

Użyjemy rozkładu Laplace'a, który jest różnicą dwóch rozkładów wykładniczych z gęstością i dystrybuantą daną wzorem:

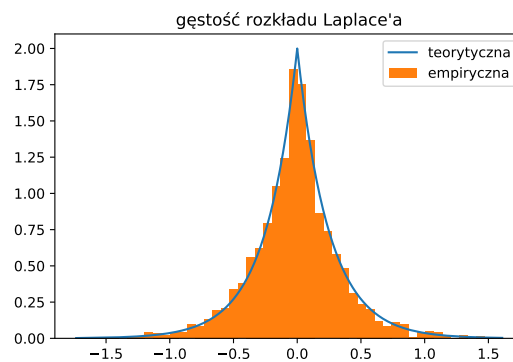
$$f(x) = \frac{\lambda}{2} e^{-\lambda|x|}$$

$$F(x) = \frac{1}{2} + \frac{1}{2} \operatorname{sgn}(x)(1 - e^{-\lambda|x|})$$

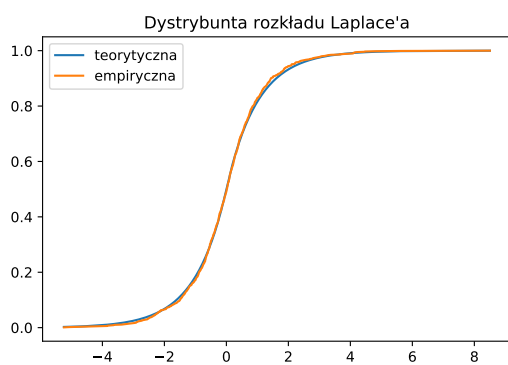
Komponujemy rozkład Laplace'a z dwóch rozkładów wykładniczych i porównujemy jego gęstość oraz dystrybuantę empiryczną i teoretyczną:



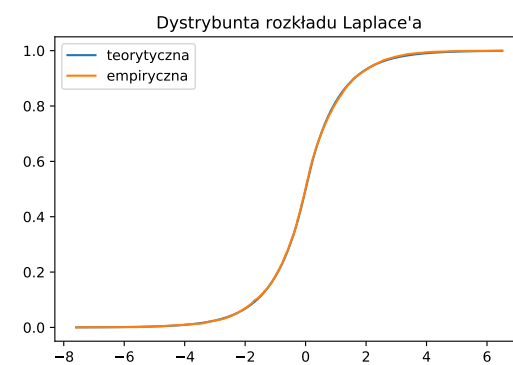
Wykres 9: dla $\lambda = 1$ i dla 5000 losowań



Wykres 10: dla $\lambda = 4$ i dla 3000 losowań

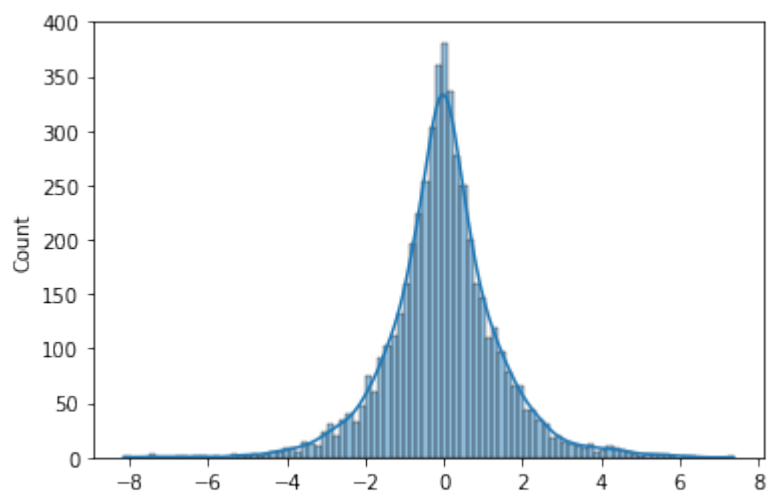


Wykres 11: dla 1000 losowań



Wykres 12: dla 5000 losowań

Dodatkowo, w ramach ciekawostki stworzyliśmy 'gładki' estymator gęstości teoretycznej:



Wnioski:

Tak jak w poprzednich przykładach, widzimy, że gęstości oraz dystrybuanty pokrywają się, dlatego metoda kompozycji dla rozkładu ciągłego — Laplace’a, jest poprawnie zaimplementowana.

4. Metoda Boxa - Müllera

Metoda ta opiera się na losowaniu wartości z rozkładu jednostajnego i podstawieniu do zdefiniowanych rozkładów X i Y :

$$X = \sqrt{-2 \log(U_1)} \cos(2\pi U_2)$$

$$Y = \sqrt{-2 \log(U_1)} \sin(2\pi U_2)$$

Wtedy otrzymujemy, że X i Y są niezależne i mają takie same rozkłady, oraz $X, Y \sim N(0, 1)$. Wyznaczyliśmy te zmienne ze współrzędnych biegunowych, niezależnych zmiennych X i Y :

$$R^2 = X^2 + Y^2$$

$$\tan \theta = \frac{Y}{X}$$

,gdzie $X = R \cos \theta, Y = R \sin \theta$, a gęstość łączna, to złożenie dwóch niezależnych rozkładów normalnych dla $X, Y \sim N(0, 1)$, czyli:

$$f(x, y) = \frac{1}{2\pi} e^{-\frac{(x^2+y^2)}{2}}$$

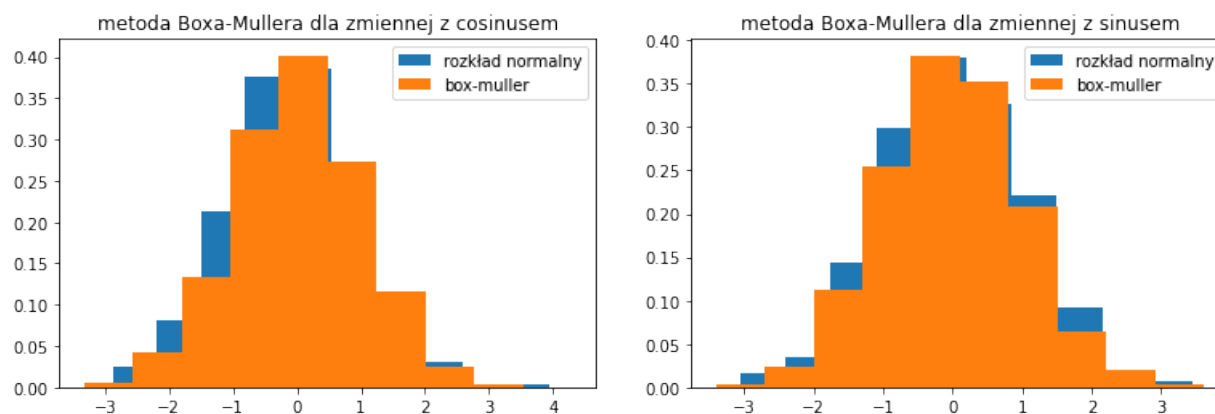
Algorytm:

(a) Generuj $U_1 \sim U(0, 1), U_2 \sim U(0, 1)$

(b) Wstaw $X = \sqrt{-2 \log(U_1)} \cos(2\pi U_2), Y = \sqrt{-2 \log(U_1)} \sin(2\pi U_2)$

Przykładowe implementacje:

Analizę naszych danych zaczęliśmy od porównania, wyprowadzonych już, dwóch gęstości, wyznaczonych metodą odwrotnej dystrybuanty:



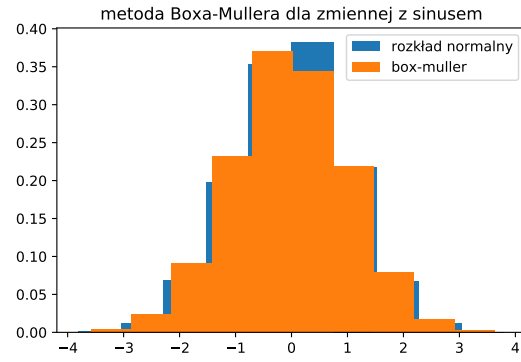
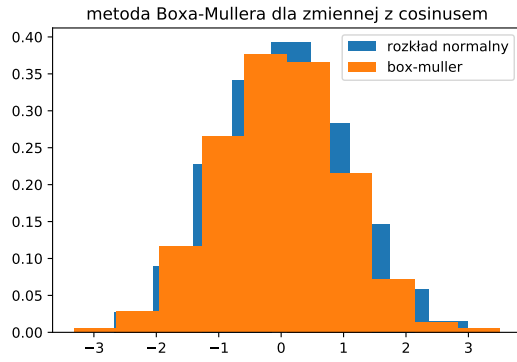
Wykres 13: dla 1000 losowań

Następnie porównujemy gęstości każdej zmiennej z teoretyczną gęstością rozkładu normalnego:

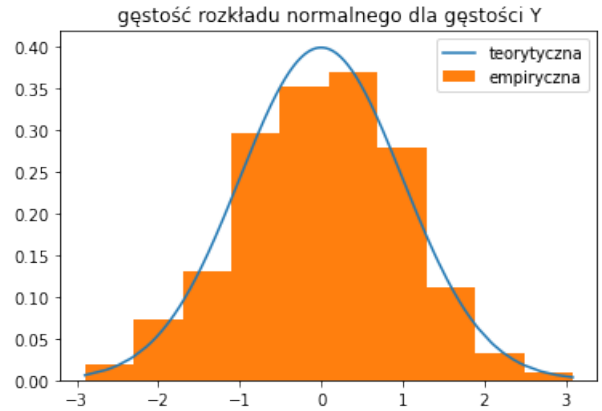
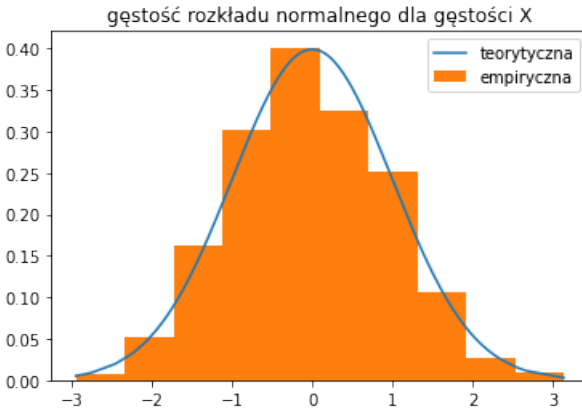
To samo dla dystrybuant:

Dodatkowo implementujemy wbudowany rozkład normalny oraz QQ-plot'y dla wyznaczonych przez nas gęstości i dystrybuanty, w celu potwierdzenia poprawności naszej metody:

Wnioski:



Wykres 14: dla 5000 losowań



Wykres 15: dla 1000 losowań

Metoda ta nie jest za bardzo optymalna, ponieważ musimy wygenerować wartości funkcji trygonometrycznej. Jednak jak możemy zauważyć po porównaniu - gęstości się pokrywają, a więc metoda jest prawidłowo zaimplementowana.

5. Metoda biegunowa

Metoda ta jest dość podobna do metody Boxa - Müllera. Rozpatrywanym obszarem jest koło jednostkowe, tylko tym razem weźmiemy rozkład wektora (V_1, V_2) , gdzie:

$$V_1 = R \cos \theta$$

$$V_2 = R \sin \theta$$

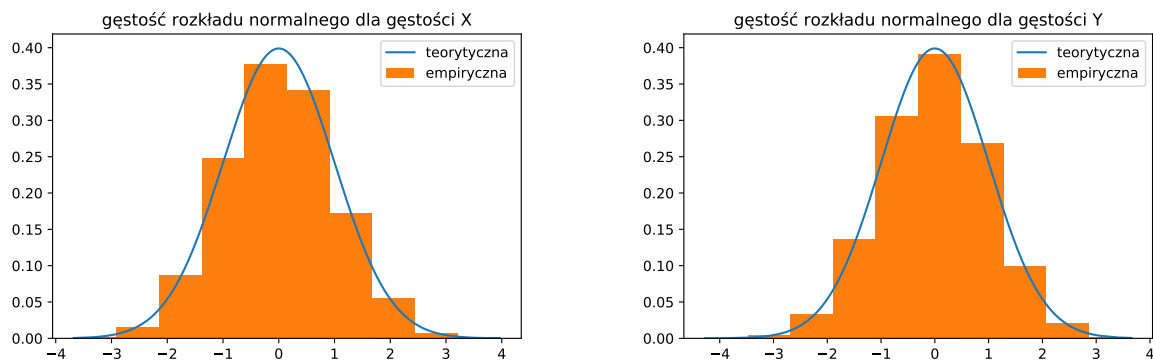
$$R = \sqrt{V_1^2 + V_2^2}$$

Po policzeniu dystrybuanty R i prawdopodobieństwa, że $P(R^2 \leq r, \alpha \leq a)$, gdzie $a = \text{const.}$, otrzymujemy, że α i R^2 mają rozkład:

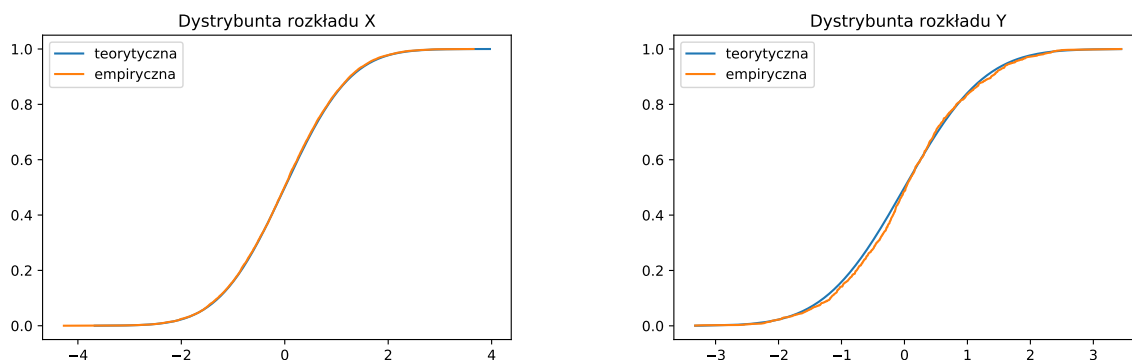
$$R^2 \sim U(0, 1)$$

$$\alpha \sim U(0, 2\pi)$$

Następnie, korzystając z wyliczonych wyżej zmiennych X i Y w metodzie Boxa - Müllera, możemy podstawić R^2 i α odpowiednio za argument logarytmu oraz funkcji trygonometrycznej. W wyniku



Wykres 16: dla 5000 losowań



Wykres 17: dla 5000 losowań

powyższego działania otrzymujemy następujące zmienne:

$$X = \sqrt{\frac{-2 \log R^2}{R^2}} V_1$$

$$Y = \sqrt{\frac{-2 \log R^2}{R^2}} V_2$$

Algorytm:

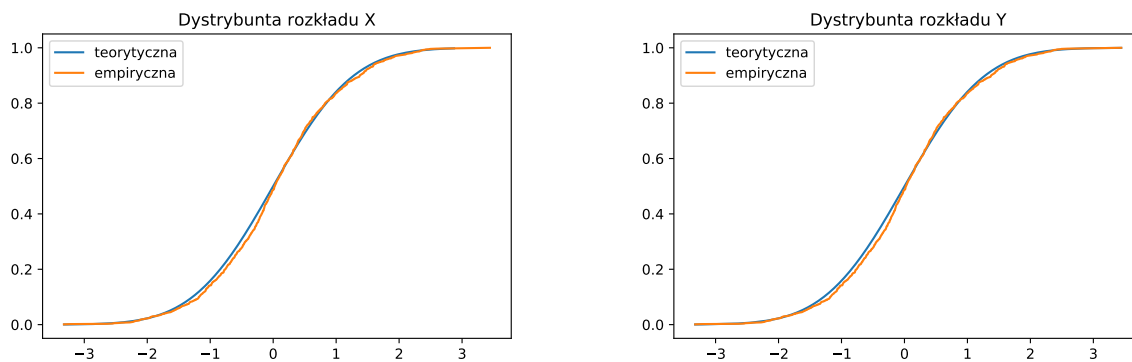
- Generuj $V_1 \sim U(-1, 1)$, $V_2 \sim U(-1, 1)$
- Wyznacz $R^2 = V_1^2 + V_2^2$
- Jeśli $R^2 > 1$ to wróć do a)
- wstaw wyliczoną wartość do wyżej wyznaczonej zmiennej X i Y

Przykładowe implementacje:

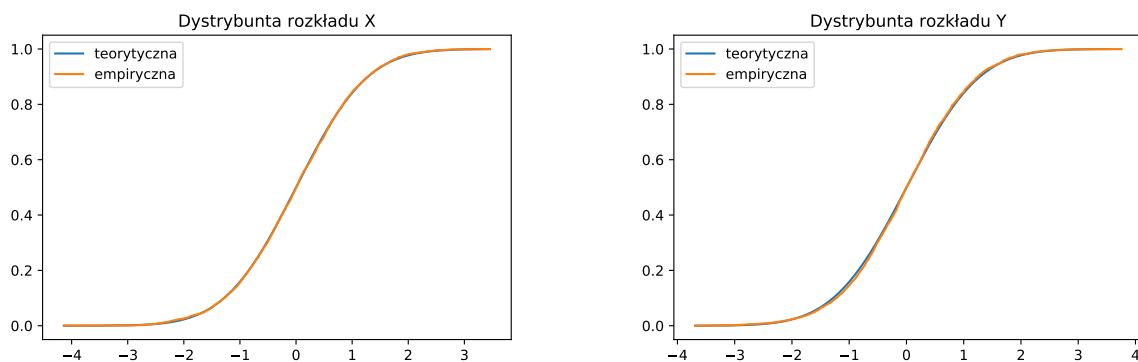
Najpierw pokażemy losowy rozkład powyższej metody dla próby różnej długości: Porównamy gęstość empiryczną i teoretyczną: I sprawdzenie poprawności naszej metody dla wykresu normalnego:

To samo dla dystrybuant:

Wnioski:

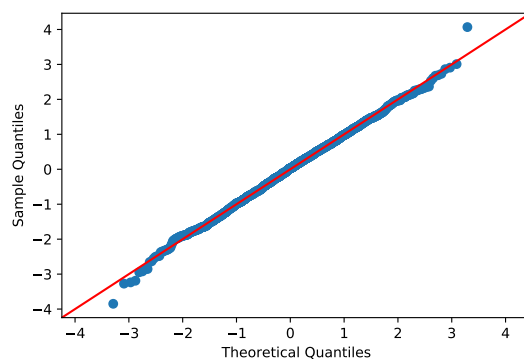
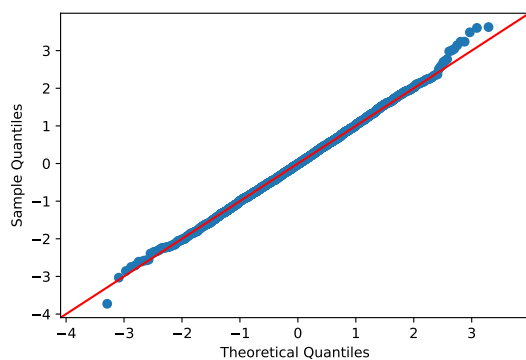


Wykres 18: dla 1000 losowań

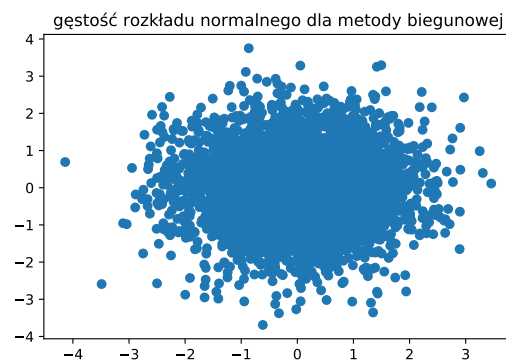
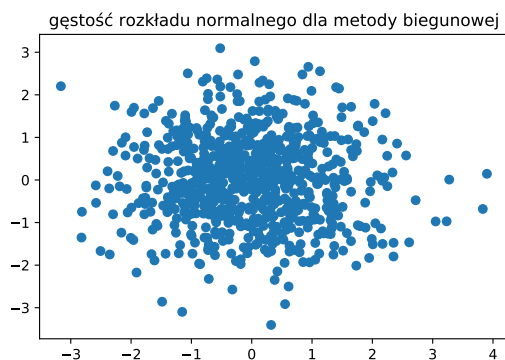


Wykres 19: Gęstości X i Y dla 5000 losowań

Metoda biegunowa jest szybsza i efektywniejsza od metody Boxa - Müllera. Widzimy również, że gęstości oraz dystrybuanty rozkładów X i Y, pokrywają się z rozkładem normalnym, co oznacza, że nasza implementacja jest poprawna. Im większa miara próby rozkładu, tym wykresy są bardziej zbliżone do wbudowanego rozkładu normalnego.

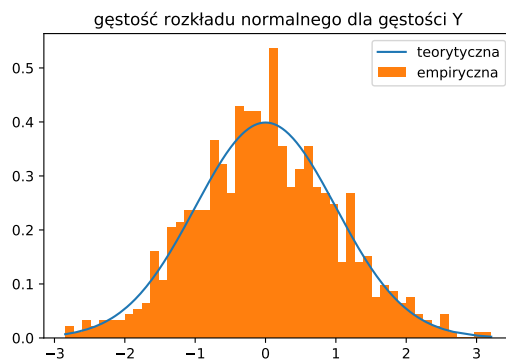
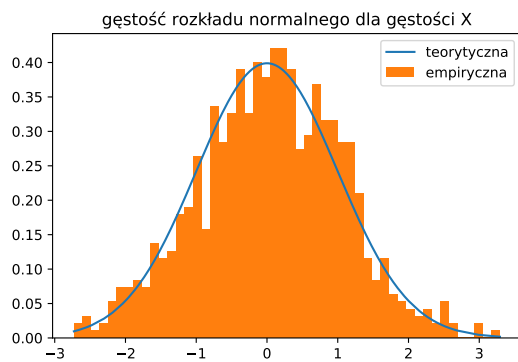


Wykres 20: Gęstości X i Y dla 1000 losowań

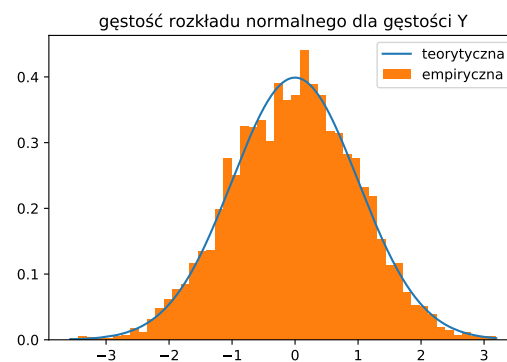
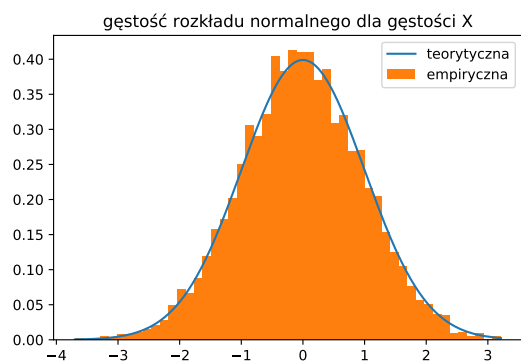


Wykres 21: Gęstości X i Y dla 1000 losowań

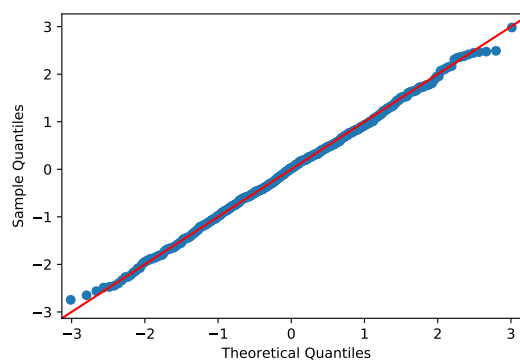
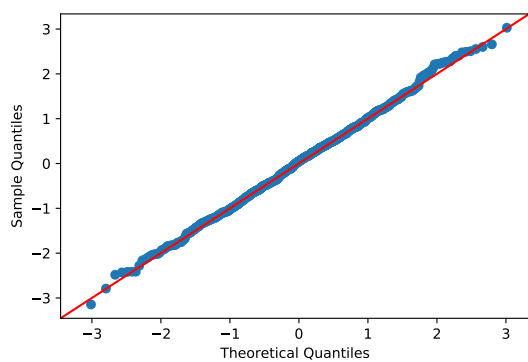
Wykres 22: Gęstości X i Y dla 5000 losowań



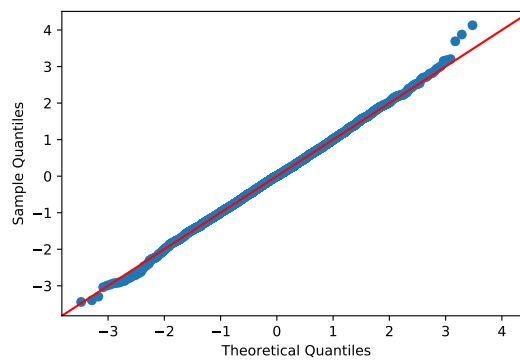
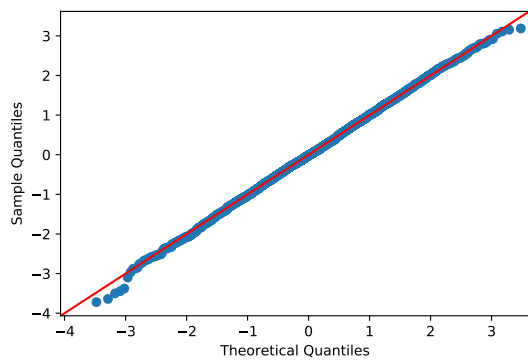
Wykres 23: Gęstości X i Y dla 1000 losowań



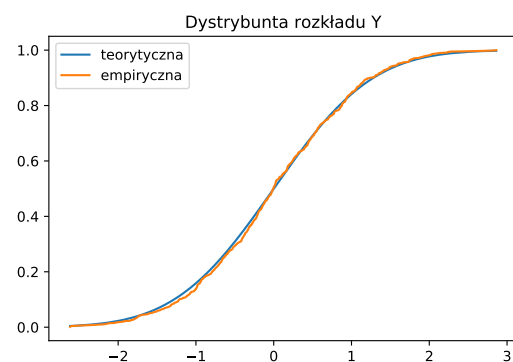
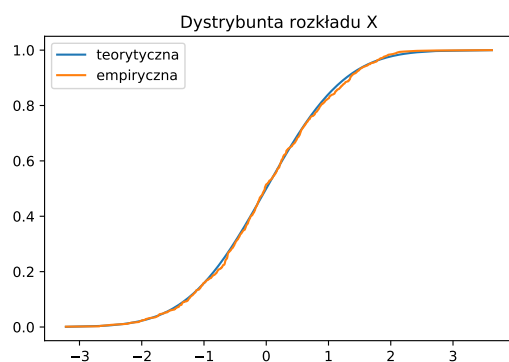
Wykres 24: Gęstości X i Y dla 5000 losowań



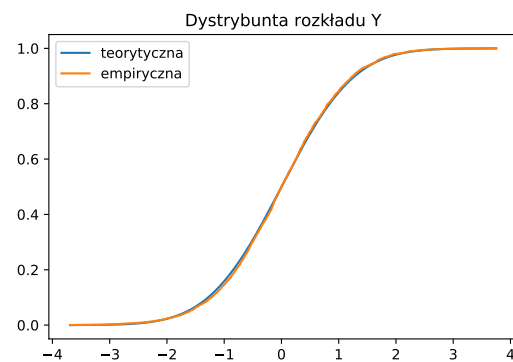
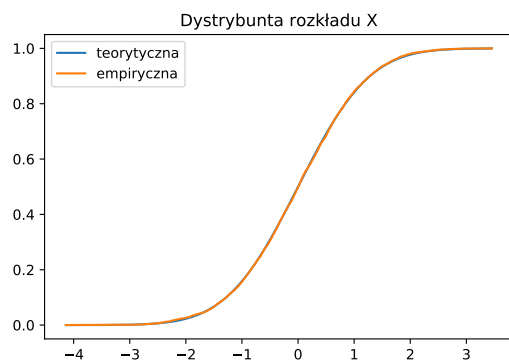
Wykres 25: Gęstości X i Y dla 1000 losowań



Wykres 26: Gęstości X i Y dla 5000 losowań



Wykres 27: dla 1000 losowań



Wykres 28: dla 5000 losowań