Wydział: Fizyki i Informatyki Stosowanej
Kierunek: Informatyka Stosowana
Rok: 2020/21
Semsetr: letni
Typ: stacjonarne
Nr albumu: 401984
Data: 07.06.2021

Sprawozdanie - Laboratorium nr 14
Generowanie ciągu liczb pseudolosowych o rozkładzie jednorodnym i trójkątnym.

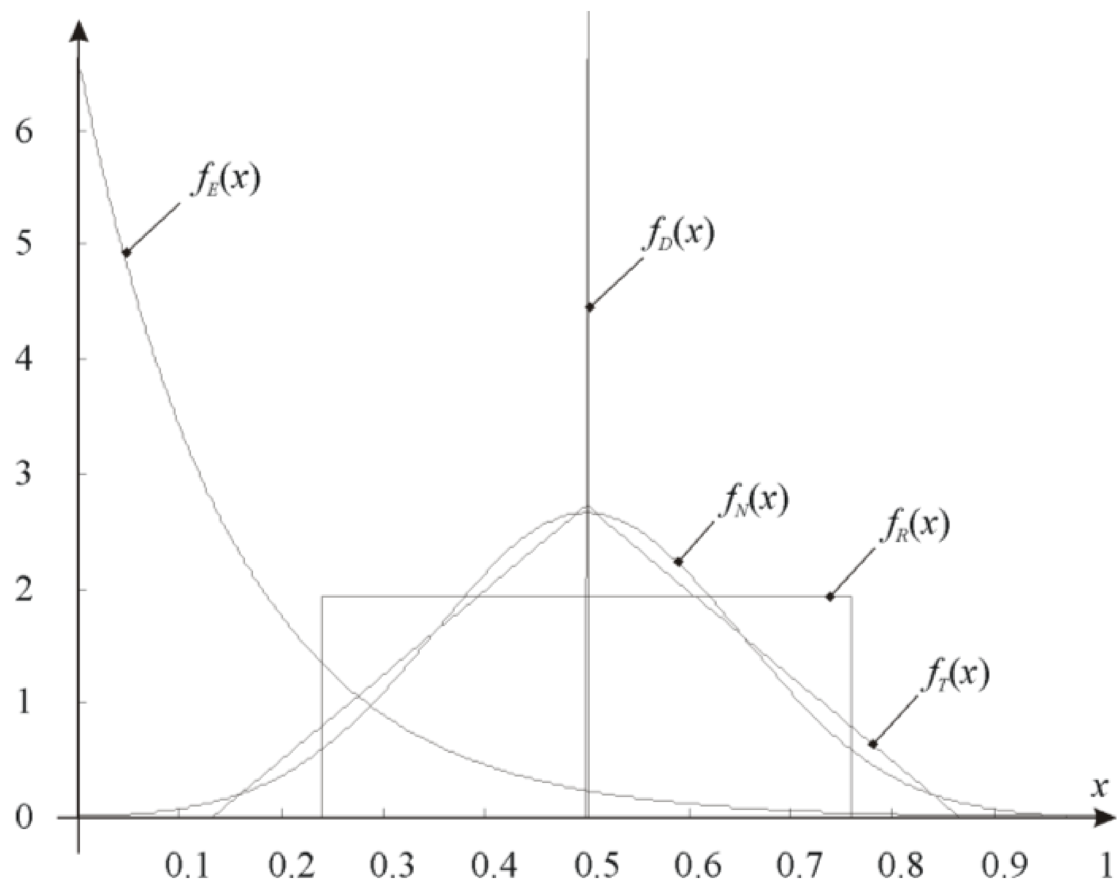
Spis treści

opracował:
Tomasz Szkaradek

1 Wstęp teoretyczny

1.1 Rozkład prawdopodobieństwa

Rozkład prawdopodobieństwa – miara probabilistyczna określona na zbiorze wartości pewnej zmiennej losowej (wektora losowego), przypisująca prawdopodobieństwa wartościom tej zmiennej. Formalnie rozkład prawdopodobieństwa można rozpatrywać bez odwołania się do zmiennych losowych.



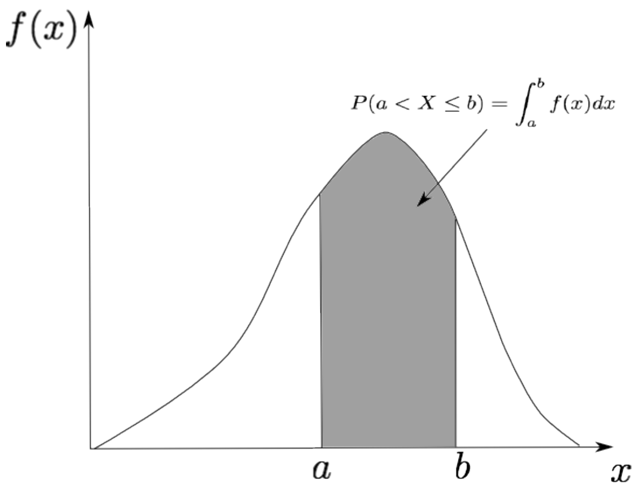
Rysunek 1: Wybrane rozkłady gęstości prawdopodobieństwa:
$f_N(x)$ – rozkład normalny,
$f_E(x)$ – rozkład wykładniczy,
$f_R(x)$ – rozkład jednostajny,
$f_T(x)$ – rozkład trójkątny,
$f_D(x)$ – rozkład delty Diraca dla zmiennej pewne

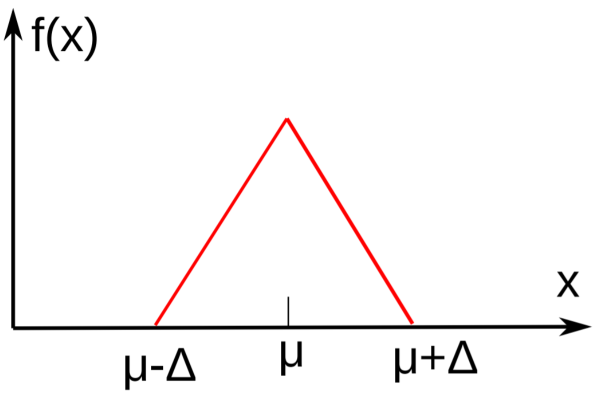
1.2 Rozkład trójkątny

Rozkład trójkątny - ten typ rozkładu jest użyteczny do opisu zmiennych losowych o stałej gęstości prawdopodobieństwa w obrębie określonego przedziału Funkcję gęstości prawdopodobieństwa dla rozkładu trójkątnego $T(\mu, \Delta)$ definiujemy następująco:

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

gdzie:
μ to środek rozkładu,
Δ to jego szerokość.





Jeśli $\xi_1 \in U(0,1)$ i $\xi_2 \in U(0,1)$ to zmienną o rozkładzie trójkątnym oraz parametrach μ i Δ generujemy stosując formułę

$$x = \mu + (\xi_1 + \xi_2 - 1) \cdot \Delta \tag{2}$$

Dystrybuanta tego rozkładu jest następująca

$$F(x) = \begin{cases} -\frac{1}{\Delta^2}(\frac{-x^2}{2} + \mu x) + \frac{x}{\Delta} - (-\frac{1}{\Delta^2} \cdot (\frac{-(\mu-\Delta)^2}{2} + \mu(\mu-\Delta)) + \frac{(\mu-\delta)}{\delta}) & x < \mu \\ -\frac{1}{\delta^2} \cdot (\frac{x^2}{2} - \mu x) + \frac{x}{\Delta} - (-\frac{1}{\Delta^2}(\frac{\mu^2}{2} - \mu^2) + \frac{\mu}{\Delta}) + \frac{1}{2} & x \geqslant \mu \end{cases} \tag{3}$$

1.3 Generator mieszany

Najbardziej znanym sposobem generowania liczb pseudolosowych jest metoda opracowana przez Lehmer'a w 1951 zwana liniowym generatorem kongruentnym. Polega ona na obliczaniu kolejnych liczb pseudolosowych: $x_1, x_2, \cdots x_n$ o zakresie wartości $0, \cdots, m-1$.Z generatora opracowanego przez Lehmer'a możemy otrzymać generator mieszany który wyraża się wzorem:

$$x_{i+1} = (a \cdot x_i + c) mod m \tag{4}$$

gdzie:
x_i kolejne pseudolosowe wartości
x_1 jest początkową wartością, którą inicjuje się generator - tzw. ziarnem
a, c, m parametry generatora

1.4 Test χ^2

Test zgodności χ^2 jest to najczęściej stosowany test nieparametryczny. Służy on do weryfikowania hipotezy, że obserwowana cecha X w zbiorowości generalnej ma określony typ rozkładu, np. dwumianowy, Poissona, normalny itd. Postać statystyki sprawdzającej hipotezy H_0

$$\chi^2 = \sum_{n=1}^{K} \frac{(n_i - n * p_i)^2}{n * p_i} \tag{5}$$

gdzie: n_i to ilość liczb znajdujących się w podprzedziale o indeksie i,
p_i to prawdopodobieństwo teoretyczne że zmienna losowa X znajdzie się w i-tym przedziale

$$p_i = F(x_i, max) - F(x_i, min) \tag{6}$$

F(X) jest wartością dystrybuanty(dla np. rozkładu trójkątnego obliczamy ją zgodnie ze wzorem [3])
Liczbę stopni swobody określamy jako $v = K - r - 1$
gdzie:
K jest liczbą podprzedziałów,
r = 2 jest liczbą parametrów testowanego rozkładu (σ i Δ)

2 Zadanie do wykonania

2.1 Rozkład jednorodny

Naszym pierwszym zadaniem w trakcie laboratoriów było napisać generator mieszany dla rozkładu jednorodnego korzystając z wzoru podanego na wstępie teoretycznym musimy wylosować $n = 10^4$ liczby pseudolosowych za pomocą naszego programu, startując od liczby $x_0 = 10$

$$x_{i+1} = (a \cdot x_i + c) mod m \tag{7}$$

o parametrach:

1. $a = 123$
   $c = 1$
   $m = 2^{15}$

2. $a = 69069$
   $c = 1$
   $m = 2^{32}$

Następnie dla obu tych przypadków rysujemy:
Wykres zależności $X_{i+1} = f(X_i)$
gdzie $X_i = x_1/(m+1)$ z warunku normalizacji do rozkładu $U(0,1)$
Histogram (dla K = 12 podprzedziałów) rozkładu gęstości prawdopodobieństwa dla $n = 10^4$ liczb pseudolosowych o rozkładzie równomiernym (oba przypadki).
Na sam koniec zadania 1 obliczamy wartości μ, σ i porównujemy z teoretycznymi.

2.2 Rozkład trójkątny

W następnym zadaniu będziemy się zajmować rozkładem trójkątnym: Na samym początku generujemy $n = 10^3$ liczby pseudolosowych według wzoru podanego we wstępie teoretycznym tj,

$$x = \mu + (\xi_1 + \xi_2 - 1) \cdot \Delta \tag{8}$$

dla parametrów $\Delta = 3$ $\sigma = 4$
Następnie dzielimy nasz przedział $[\mu - \Delta, \mu + \Delta]$ w którym znajdują się nasze liczby na K=10 równych podprzedziałów a następnie zliczamy ile liczb wpadło do każdego z tych podprzedziałów.

Dla rozkładu trójkątnego przeprowadzamy test χ^2 czyli określamy wartość statystyki testowej

$$\chi^2 = \sum_{n=1}^{K} \frac{(n_i - n * p_i)^2}{n * p_i} \tag{9}$$

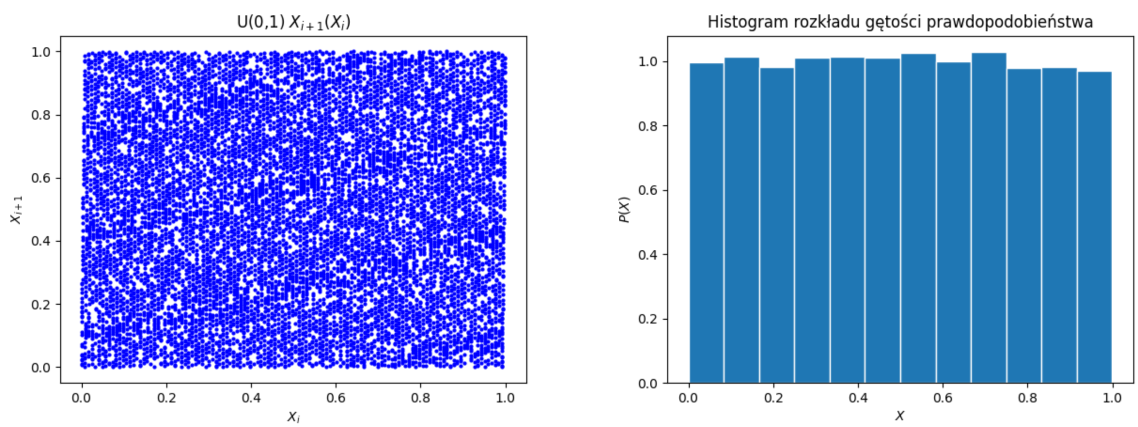
W wykonujemy wykres gęstości naszego prawdopodobieństwa oraz histogram pokazujący wartości $\frac{n_i}{n}$ oraz p_i dla każdego z podprzedziałów.
Następnie testujemy hipotezę H_0 czy wygenerowany rozkład jest rozkładem $T(\mu, \Delta)$ wobec H_1 że nie jest to prawdą.
Korzystając z odpowiednich tabel statystycznych proszę sprawdzić czy nasza hipoteza jest prawdziwa na poziomie istotności $\alpha = 0.05$ (α jest prawdopodobieństwem pierwszego rodzaju czyli prawdopodobieństwem odrzucenia hipotezy H_0 gdy ta jest prawdziwa). W tym celu definiujemy obszar krytyczny testu:

$$\Phi = X : \chi^2(X) > \epsilon \tag{10}$$

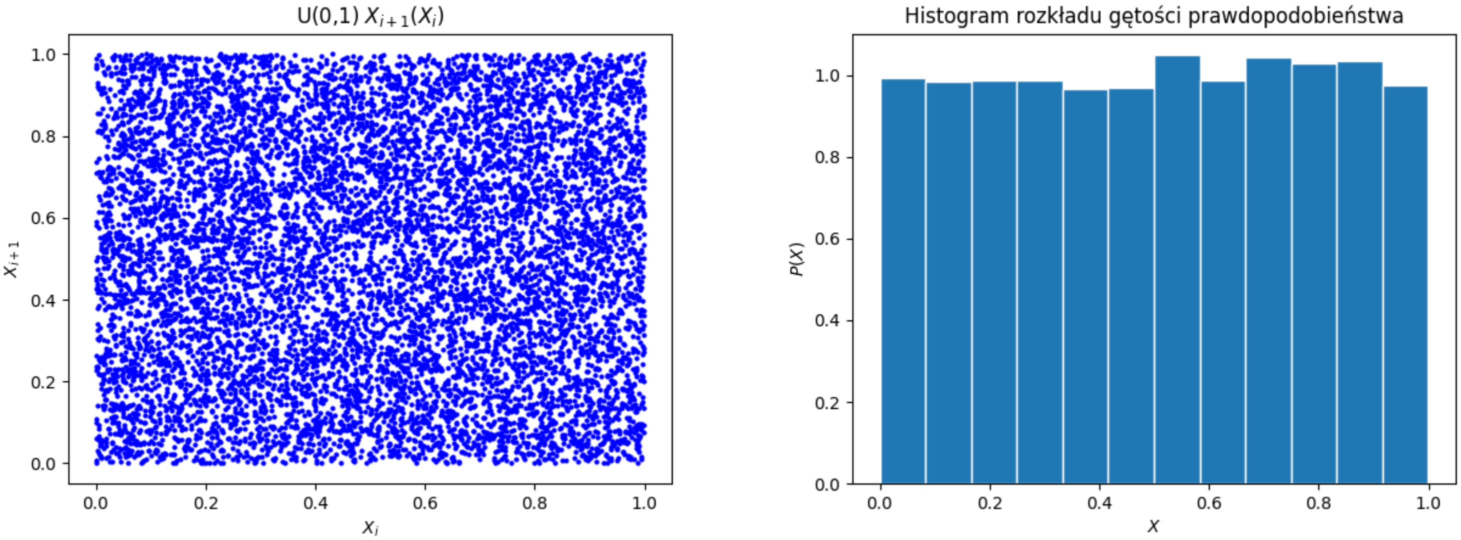
gdzie: $X = x_1, x_2, \cdots, x_n$ jest ciągiem liczb pseudolosowych, $\chi^2(X)$ wartością statystyki dla danego ciągu X, ϵ jest poziomem krytycznym danego rozkładu dla określonej liczby stopni swobody (określone w wstępie teoretycznym) i założonego poziomu istotności (należy odczytać z tabel statystycznych). Jeśli $\chi^2 < \epsilon$ to stwierdzamy że dla zadanego poziomu istotności nie ma podstaw do odrzucenia hipotezy H_0.

3 Wyniki

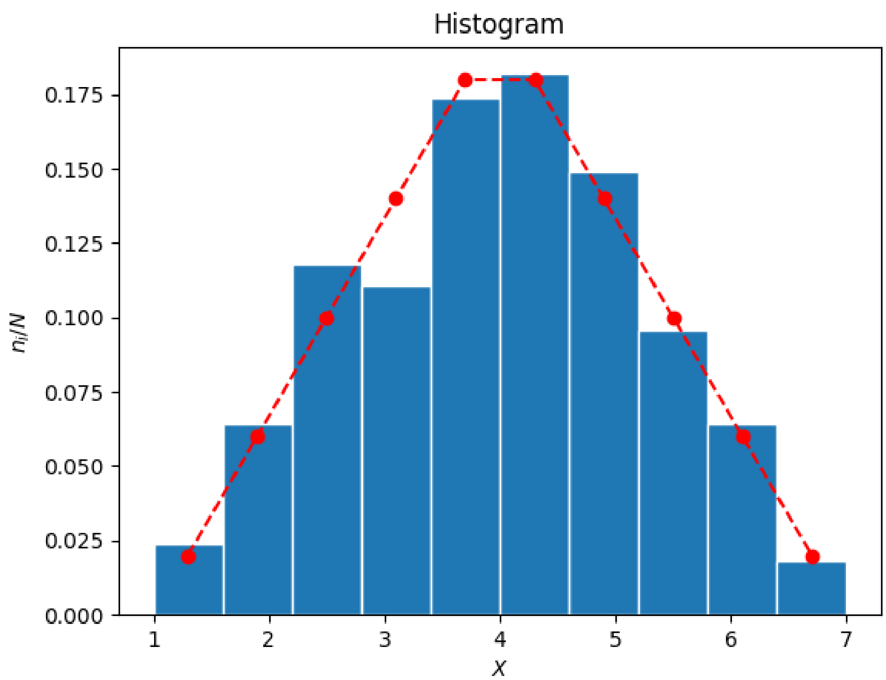
Cały program został napisany w języku Python.



Rysunek 3: Wartości $\Delta_1 = 0.498$, $\sigma_1 = 0.287$



Rysunek 4: Wartości $\Delta_2 = 0.503$, $\sigma_2 = 0.288$



Rysunek 5: Histogram dla rozkładu trójkątnego

4 Podsumowanie

Podsumowując można powiedzieć że nie jesteśmy w stanie wygenerować ciągu liczb prawdziwie losowych z wykorzystaniem programów komputerowych wygenerowane liczby będą pseudolosowe. Pomiędzy kolejno wygenerowanymi liczbami zawsze będzie istniała jakaś zależność, jednak dzięki skorzystaniu z odpowiednich parametrów i funkcji będzie ona bardzo trudna do ustalenia i w konsekwencji ciąg liczbowy będzie trudny do odróżnienia od prawdziwie losowych liczb.
Możemy to wywnioskować na podstawie wygenerowanych wykresów 3 oraz 4 na których widać iż liczby nie pokrywają w pełni określonego obszaru.
Generator mieszany teoretycznie generuje liczby o rozkładzie jednorodnym, jednak dla małej liczby wylosowanych punktów pojawiają się pewne odchylenia w histogramach jednak będą się one zmniejszać wraz z liczbą generowanych punktów. Wyniki obliczonych wartości Δ oraz σ w obu przypadkach nieznacznie odbiegały(dla ustalonej dokładności ε_1 i ε_2) od wartości teoretycznych tj.

$$\Delta = \frac{0+1}{2} = 0.5 \tag{11}$$

$$\sigma = \frac{(1-0)}{\sqrt{12}} = 0,289 \tag{12}$$

$$|\Delta - \Delta_1| = |0.5 - 0.498| < \varepsilon_1$$

$$|\sigma - \sigma_1| = |0,289 - 0.287| < \varepsilon_2$$

$$|\Delta - \Delta_2| = |0.5 - 0.503| < \varepsilon_1$$

$$|\sigma - \sigma_2| = |0,289 - 0.288| < \varepsilon_2$$

co świadczy o poprawności zaimplementowanych generatorów które dają poprawne rezultaty i jak najbardziej spełniają swoje zadanie.
Podsumowując, zaletą generatorów liniowych jest prosta implementacja i szybkość działania. Przy odpowiednim doborze parametrów oraz zmiennym ziarnie można uzyskać liczby, które wyglądają na losowe i na pierwszy rzut oka nie mają widocznych zależności między kolejnymi wartościami. Dalej jednak układają się na hiperpłaszczyznach i aby tego uniknąć, należałoby zastosować inne generatory. Z otrzymanych przez nas wyników z 2 zadania to jest rozkład trójkątny możemy stwierdzić iz wylosowane przez nas wartości wmiare dokładnie oddają rozkład trójkątny co tylko potwierdza wykres 5 w którym kolejne wartości n_i/N układają się w trójkąt i nie odbiegają stanowczo od teoretycznych (tj.czerwona linia). Natomiast wykonany przez nas test χ^2 zdaje się to tylko potwierdzać prawdziwość naszych wyników

$$\chi^2 = \sum_{n=1}^{K} \frac{(n_i - n * p_i)^2}{n * p_i} = 5.63 \tag{13}$$

Odczytując z tablic wartość $\epsilon = 20,28$ poziomu krytycznego naszego rozkładu
dla $K = 10$, $r = 2 \Longrightarrow \nu = K - r - 1$ na poziomie istotności $\alpha = 0.05$ Możemy stwierdzić że dla zadanego poziomu istotności nie ma podstaw do odrzucenia hipotezy H_0 (wygenerowany rozkład jest rozkładem $T(\mu, \Delta)$) ponieważ:

$$\chi^2 < \epsilon \quad \Rightarrow \quad 5.63 < 20,28 \tag{14}$$

5 Literatura

[1] Tomasz Chwiej, Generatory liczb pseudolosowych
http://home.agh.edu.pl/ chwiej/mn/generatory_1819.pdf

[2] Wikipedia, Rozkład prawdopodobieństwa
https://www.wikiwand.com/pl/Rozk%C5%82ad_prawdopodobie%C5%84stwa