AGH UNIVERSITY OF SCIENCE AND TECHNOLOGY

WYDZIAŁ FIZYKI I INFORMATYKI STOSOWANEJ
KIERUNEK STUDIÓW: FIZYKA TECHNICZNA

METODY MONTE CARLO

Laboratorium 2

Generatory liczb pseudolosowych o zadanym rozkładzie w jednym wymiarze

---

zrealizował
Przemysław Ryś

Kraków, 4 Marzec 2024

1 Opis zagadnienia

Na zajęciach skonstruowaliśmy generator jednowymiarowy o funkcji gęstości prawdopodobieństwa $f(x) = \frac{1+x-x^3}{4}$ dla $x \in [0,1]$ oraz dystrybuancie $F(x) = \frac{x^2}{4} + \frac{x^4}{5}$ dla $x \in [0,1]$.

Do generowania liczb pseudolosowych użyjemy schematów dla:

a) Rozkładu złożonego,

b) Łańcucha Markowa,

c) Metody eliminacji.

1.1 Rozkład złożony

Dystrybuanta rozkładu jest wielomianem, więc możemy spróbować zapisać ją w postaci rozkładu złożonego:

$$F(x) = \sum_{i=1}^{n} g_i H_i(x),$$

gdzie g_i to dystrybuanta rozkładu dyskretnego, a $H_i(x)$ to dystrybuanty rozkładów podlegających superpozycji. Musimy przekształcić $F(x)$ do postaci akceptowalnej:

$$F(x) = \frac{1}{4}x + \frac{2x^2 - x^4}{5},$$

skąd odczytujemy:

$$g_1 = \frac{4}{5}, \quad H_1 = x,$$

$$g_2 = \frac{1}{5}, \quad H_2 = 2x^2 - x^4.$$

Teraz należy znaleźć funkcje odwrotne H_1 i H_2 aby otrzymać liczbę losową $X = x$. H_1 jest liniowa, więc od razu dostajemy:

$$H_1(x) = x = U \sim U(0,1).$$

Dla H_2:

$$H_2(x) = 2x^2 - x^4 = U \sim U(0,1).$$

Mamy równanie 4 stopnia:

$$x^4 - 2x^2 + U = 0,$$

które rozwiązujemy podstawiając $y = x^2$ i rozwiązując równanie kwadratowe:

$$y^2 - 2y + U = 0, \quad y_1, y_2 = 1 \pm \sqrt{1-U}.$$

Ponieważ $y_1 > 1$ (pamiętamy, że generujemy $x \in [0,1]$), wybieramy y_2 i liczymy x:

$$x = \sqrt{1 - \sqrt{1-U}}.$$

1.2 Łańcuch Markowa

W tej metodzie generujemy ciąg $\{X_0, X_1, X_2, \ldots\}$, gdzie związek pomiędzy ostatnim elementem X_i a kolejnym X_{i+1} określamy na podstawie prawdopodobieństwa przejścia, które spełnia warunek detailed balance:

$$\frac{1}{\Delta} T(X_{i+1}|X_i) = \frac{1}{2\Delta},$$

i prawdopodobeińistwa akceptacji nowego stanu (liczby):

$$p_{\text{acc}} = \min\left\{1, \frac{T(X_i|X_{i+1}) f(X_{i+1})}{T(X_{i+1}|X_i) f(X_i)}\right\}.$$

Algorytm Metropolisa generowania nowego elementu w łańcuchu:

$$\begin{cases} x_{\text{new}} = X_i + (2U_1 - 1)\Delta, & \text{gdy } x_{\text{new}} \in [0,1] \text{ i } U_2 p_{\text{acc}}, \\ X_{i+1} = X_i, & \text{w przeciwnym przypadku.} \end{cases}$$

1.3 Metoda eliminacji

W metodzie tej wykorzystujemy funkcję gęstości prawdopodobieństwa $f(x)$, którą ograniczamy od góry inną funkcją o rozkładzie $g(x)$, dla której dysponujemy generatorem G. Algorytm generowania ciągu liczb metodą eliminacji:

$$\mathrm{U}_1 \sim U(0,1),$$
$$\mathrm{G}_2 \sim G, \quad \text{np. } G = 1.15 \cdot U(0,1),$$
$$\begin{cases} G_2 f(U_1) \Rightarrow X = U_1, \\ G_2 > f(U_1) \Rightarrow \text{losujemy nową parę } U_1, G_2. \end{cases}$$

1.4 Test χ^2

Test χ^2 jest jednym z najczęściej stosowanych testów statystycznych służących do oceny zgodności pomiędzy obserwowanymi danymi a oczekiwanym rozkładem. Jest to test nielosowy, który pozwala zbadać, czy rozkład empiryczny danych różni się istotnie od założonego teoretycznego rozkładu.

Wartość statystyki testowej dla testu χ^2 wyraża się wzorem:

$$\chi^2_{k-1} = \sum_{i=1}^{k} \frac{(n_i - p_i N)^2}{p_i N} \tag{1}$$

gdzie:

- k - liczba przedziałów (binów) w analizowanym zbiorze danych,
- p_i - prawdopodobieństwo, że zmienna losowa znajdzie się w przedziale i-tym (do wyznaczenia używa się dystrybuanty),
- n_i - ilość obserwacji (liczb pseudolosowych), które znalazły się w i-tym przedziale,
- N - całkowita ilość obserwacji.

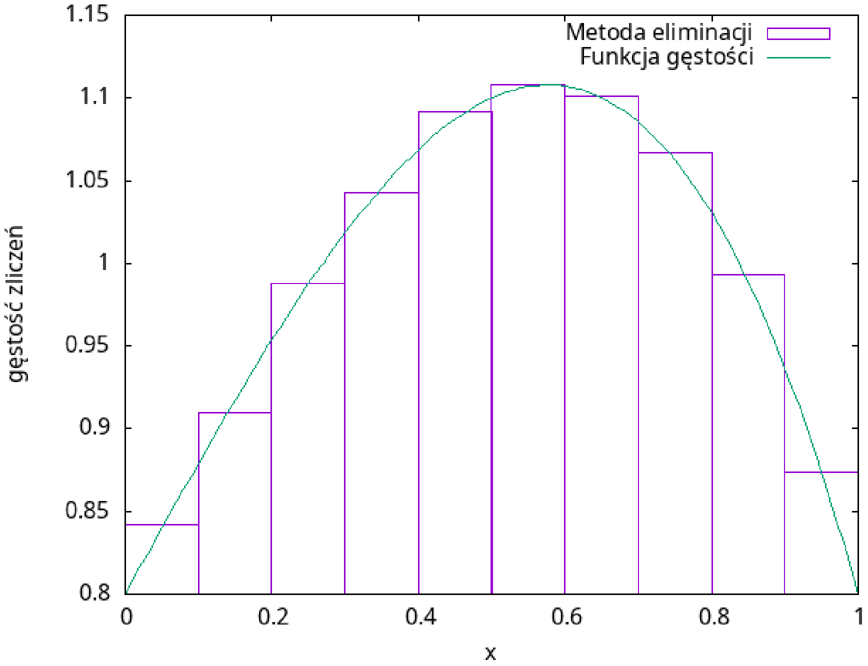
Statystyka χ^2 porównywana jest z wartością graniczną z rozkładu χ^2 z $k-1$ stopniami swobody. Na podstawie porównania uzyskanej wartości χ^2 z wartością graniczną można stwierdzić, czy hipoteza zerowa (H_0), zakładająca zgodność danych z teoretycznym rozkładem $F(x)$, jest odrzucana czy nie.

Przyjmując poziom istotności równy $\alpha = 0.05$, odrzucamy hipotezę H_0, jeśli wartość χ^2 przekracza odpowiednią wartość graniczną dla $k-1$ stopni swobody przy danym poziomie istotności.

2 Wyniki

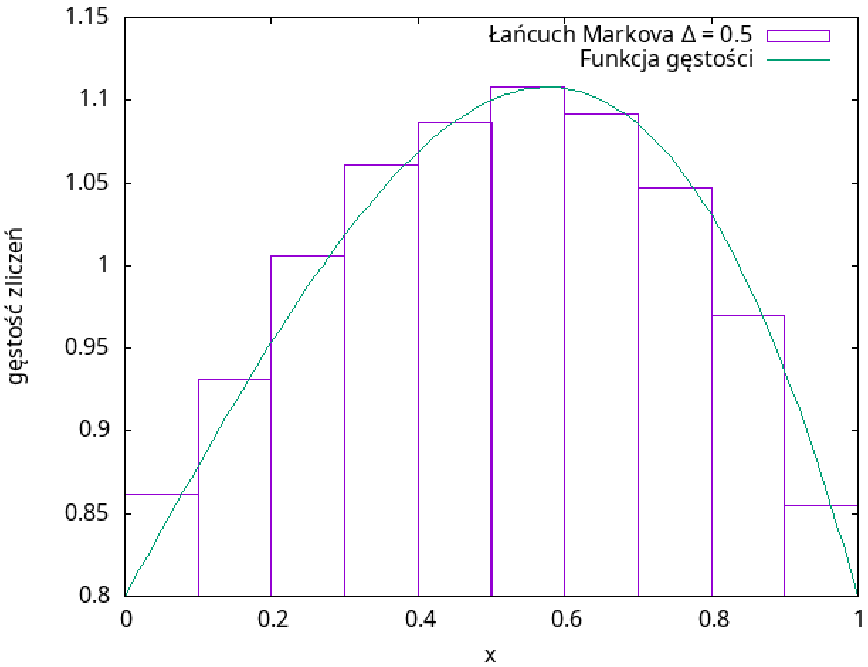
Histogramy w porównaniu z krzywą teoretyczną

Dla każdej z metod generacji liczb pseudolosowych wygenerowano histogramy danych w porównaniu z odpowiadającą im krzywą teoretyczną w programie gnuplot.



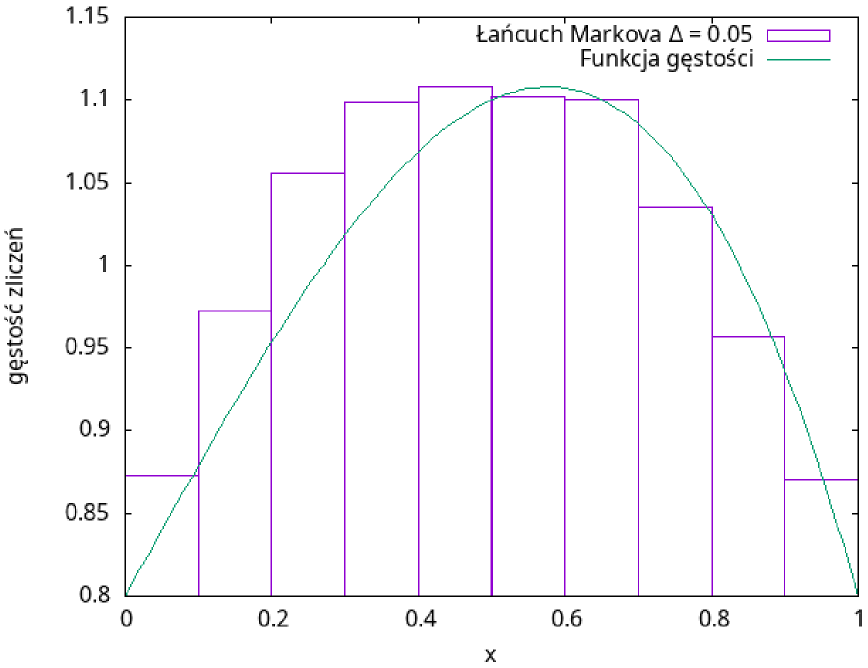
Rys. 1: Histogram liczb pseudolosowych generowanych metodą eliminacji wraz z naniesioną krzywą teoretyczną gęstości prawdopodobieństwa

Dla metody złożonej zauważamy, że histogram danych pokrywa się w przybliżeniu z krzywą teoretyczną, co sugeruje, że dane generowane tą metodą są zgodne z oczekiwanym rozkładem.



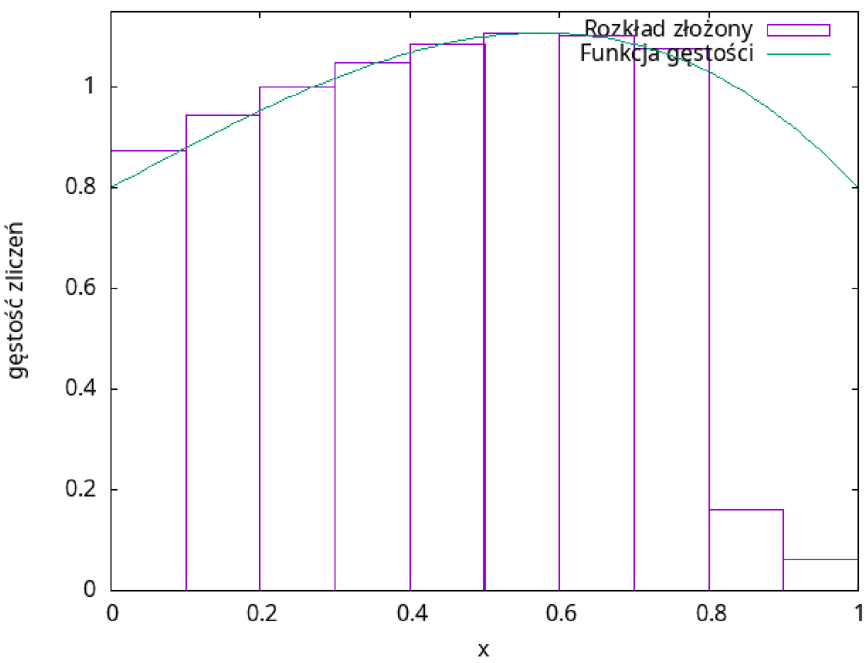
Rys. 2: Histogram liczb pseudolosowych generowanych metodą łańcucha Markova z parametrem $\Delta = 0.5$ wraz z naniesioną krzywą teoretyczną gęstości prawdopodobieństwa

Dla metody Łańcucha Markowa z $\Delta = 0.5$, histogram danych również dobrze odwzorowuje krzywą teoretyczną, co sugeruje, że generowane dane są zgodne z oczekiwanym rozkładem. W porównaniu z metodą rozkładu złożonego, wyższe stają się słupki dla mniejszych wartości x.



Rys. 3: Histogram liczb pseudolosowych generowanych metodą łańcucha Markova z parametrem $\Delta = 0.05$ wraz z naniesioną krzywą teoretyczną gęstości prawdopodobieństwa

Dla metody Łańcucha Markowa z $\Delta = 0.05$, obserwujemy podobne zachowanie histogramu danych w stosunku do krzywej teoretycznej. W porównaniu z tą samą metodą dla parametru $\Delta = 0.0$, słupki z lewej strony wychodzą ponad wartości krzywej teoretycznaj, co sprawia, że wylosowany ciąg staje się mniej efektywny względem oczekiwań.



Rys. 4: Histogram liczb pseudolosowych generowanych metodą rozkładu złożonego wraz z naniesioną krzywą teoretyczną gęstości prawdopodobieństwa

Dla metody Eliminacji, histogram danych również dobrze odzwierciedla oczekiwaną krzywą teoretyczną, ale do pewnego momentu. Prawdopodobnie związane jest to z popełnuionym w programie błędem, niemniej autor nie był w stanie go zlokalizować.

Wyniki testu χ^2

Dla każdej metody wykonano test χ^2 i porównano uzyskane wyniki z wartością graniczną rozkładu, przyjmując poziom istotności równy $\alpha = 0.05$. Wartość statystyki testowej dla $k-1$ stopni swobody obliczono ze wzoru 1.

Poniżej przedstawiono otrzymane wyniki testu χ^2 dla poszczególnych metod:

χ^2 dla metody złożonej:	168969
χ^2 dla metody Łańcucha Markova Delta = 0.5:	221.339
χ^2 dla metody Łańcucha Markova Delta = 0.05:	1087.99
χ^2 dla metody Eliminacji:	14.1666
Wartość graniczna χ^2 :	16.919

Na podstawie porównania uzyskanych wartości χ^2 z wartością graniczną można stwierdzić:

- Hipoteza H_0 dla metody złożonej zostaje odrzucona.
- Hipoteza H_0 dla metody Łańcucha Markova Delta = 0.5 zostaje odrzucona.
- Hipoteza H_0 dla metody Łańcucha Markova Delta = 0.05 zostaje odrzucona.
- Hipoteza H_0 dla metody Eliminacji nie zostaje odrzucona.

3 Wnioski

Metoda złożona oraz metoda Łańcucha Markowa dla $\Delta = 0.5$ generują próbki, których rozkłady istotnie różnią się od teoretycznego rozkładu $F(x)$. Metoda Łańcucha Markowa dla $\Delta = 0.05$ daje jeszcze bardziej odległe od teoretycznego rozkładu próbki niż poprzednie metody, co może wskazywać na większą niestabilność tej metody dla mniejszych wartości Δ. Metoda Eliminacji wydaje się być najbardziej zbliżona do teoretycznego rozkładu $F(x)$, ponieważ generowane próbki mają znacznie mniejszą wartość statystyki testowej χ^2 i nie ma podstaw do odrzucenia hipotezy zerowej H_0. Wybór odpowiedniej metody generacji próbek powinien być uzależniony od konkretnego zastosowania i wymagań dotyczących jakości generowanych próbek.