Projekt 2: Generatory liczb pseudolosowych o zadanym rozkładzie w jednym wymiarze

Kacper Połuszejko, 412183

Wstęp

Celem ćwiczenia było skonstruowanie jednowymiarowego generatora liczb pseudolosowych o funkcji gęstości prawdopodobieństwa:

$$f(x) = \frac{4}{5}(1 + x - x^3), \quad x \in [0, 1], \tag{1}$$

oraz dystrybuancie:

$$F(x) = \frac{4}{5}(x + \frac{x^2}{2} - \frac{x^4}{4}), \quad x \in [0, 1]. \tag{2}$$

Do generowania liczb pseudolosowych wykorzystano schematy dla: a) rozkładu złożonego, b) łańcucha Markowa, c) metody eliminacji.

1 Metodyka

1.1 Rozkład złożony

Dystrybuantę zapisujemy w postaci rozkładu złożonego o ogólnym wzorze:

$$F(x) = \sum_{i=1}^{n} g_i H_i(x), \quad g_i \in R, \quad H_i(x) : R \to R, \tag{3}$$

zatem w naszym przypadku:

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

skąd odczytujemy:

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

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

Funkcje odwrotne do $H1$ oraz $H2$:

$$x = U, \quad x = \sqrt{1 - \sqrt{1 - U}}, \quad U \sim U(0, 1). \tag{4}$$

W związku z tym, powyższy rozkład złożony wygenerowany został wg.następującego algorytmu:

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

$$X = \begin{cases} U_2, & \text{gdy } U_1 \leq g_1 \\ \frac{U_2}{\sqrt{1-\sqrt{1-U_2}}}, & \text{gdy } U_1 > g_1 \end{cases}$$

1.2 Łańcuch Markowa

W metodzie tej 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**

$$T(X_{i+1}|X_i) = T(X_i|X_{i+1}) = \frac{1}{2\Delta}, \quad \Delta \in [0,1]$$

i prawdopodobieństwa akceptacji nowego stanu (liczby)

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

gdzie $T(X_i|X_{i+1})/T(X_{i+1}|X_i)$ jest równe 1 dzięki spełnieniu warunku detailed balance.

Algorytm Metropolisa generowania nowego elementu w łańcuchu:

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

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

1.3 Metoda eliminacji

W tej metodzie wykorzystywana jest funkcja gęstości prawdopodobieństwa f(x), dodatkowo ograniczona od góry funkcją g(x), dla której dysponujemy generatorem **G**. Algorytm generowania ciągu liczb pseudolosowych przedstawia się w następujący sposób:

$$U_1 \sim U(0,1)$$

$$G_2 \sim \mathcal{G}, \quad \text{np. } \mathcal{G} = 1.15 \cdot U(0,1)$$

$$\begin{cases} G_2 \leq f(U_1) \implies X = U_1 \\ G_2 > f(U_1) \implies \text{losujemy nową parę } U_1, G_2 \end{cases}$$

1.4 Test 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:

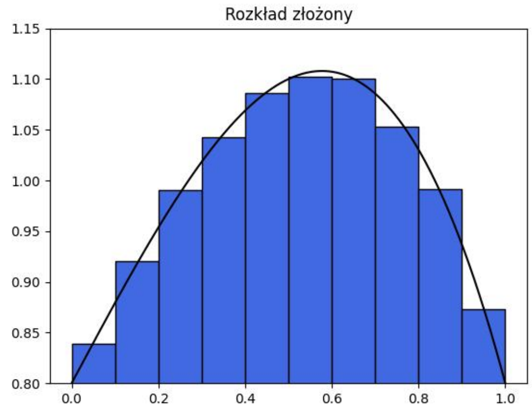
$$\sum_{i=1}^{k} \frac{(n_i - p_i N)^2}{p_i N}, \tag{5}$$

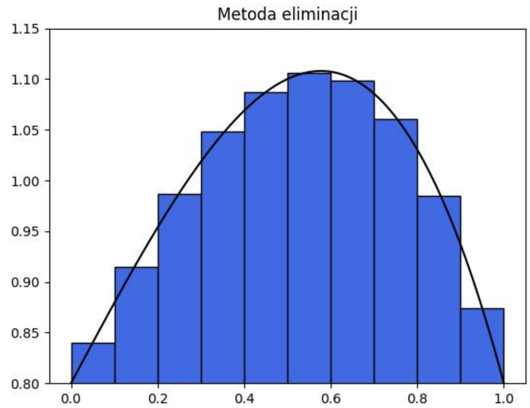
gdzie p_i to prawdopodobieństwo, że zmienna losowa znajdzie się w i-tym przedziale, n_i to ilość liczb pseudolosowych w i-tym przedziale, a N to całkowita liczba wylosowanych liczb.

Na podstawie porównania dla każdej metody stwierdzono, czy hipotezę H_0 tego, że uzyskany ciąg liczb pseudolosowych ma rozkład $F(x)$ należy odrzucić, czy nie.

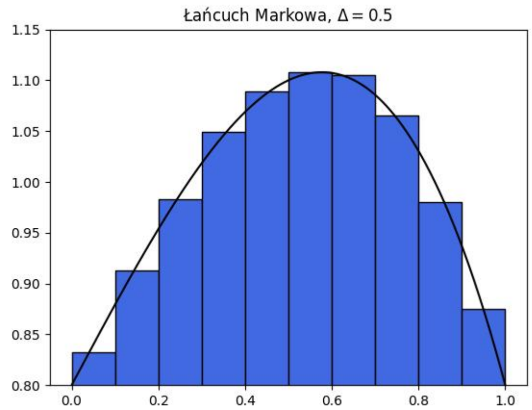
2 Wyniki

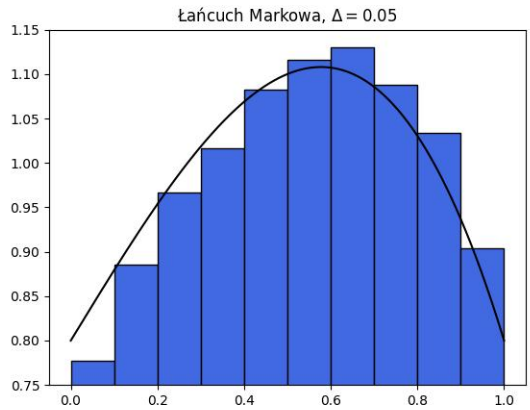
Wygenerowano $N = 10^6$ liczb pseudolosowych o dystrybuancie $F(x)$ dla każdej z metod. Dla każdego ciągu liczb sporządzono histogram o $k = 10$ podprzedziałów i porównano go z funkcją $f(x)$. Następnie wykonano testy χ^2.





Rys. 1: Histogramy o liczbie podprzedziałów $k = 10$ dla ciągów liczb pseudolosowych wygenerowanych za pomocą rozkładu złożonego (po lewej) oraz metody eliminacji (po prawej).





Rys. 2: Histogramy o liczbie podprzedziałów $k = 10$ dla ciągów liczb pseudolosowych wygenerowanych za pomocą łańcucha Markowa z parametrem $\Delta = 0.5$ (po lewej) oraz $\Delta = 0.05$ (po prawej).

Dla każdego z wygenerowanych ciągów obliczono wartość statystyki testowej χ^2 zgodnie ze wzorem (5).

Sprawdzimy teraz, czy hipotezę H_0 tego, że uzyskany ciąg liczb pseudolosowych ma rozkład $f(x)$ należy odrzucić, czy nie.

Rozkład złożony:
$3,325(\alpha = 0,05) < \chi^2_{k-1} = 12,52 < 19,023(\alpha = 0.95)$ - nie odrzucamy hipotezy H_0.

Metoda eliminacji:
$\chi^2_{k-1} = 2,949 < 3,325(\alpha = 0,05)$ - odrzucamy hipotezę H_0. Generowany rozkład jest zbyt bliski rozkładowi zadanemu.

Łańcuch Markowa ($\Delta = 0.5$):
$19,023(\alpha = 0.95) < \chi^2_{k-1} = 23,306$ - odrzucamy hipotezę H_0. Generowany rozkład za bardzo odbiega od rozkładu zadanego.

Łańcuch Markowa ($\Delta = 0.05$):
$19,023(\alpha = 0.95) < \chi^2_{k-1} = 1188,33$ - odrzucamy hipotezę H_0. Generowany rozkład za bardzo odbiega od rozkładu zadanego.

Widzimy zatem, że jedyną metodą, która przeszła test, jest metoda rozkładu złożonego. Metoda eliminacji okazała się zbyt dokładna. Z kolei ciągi liczb wygenerowane za pomocą łańcucha Markowa za bardzo odbiegały od zadanego rozkładu $f(x)$. Warto jednak zwrócić uwagę, że generator ten znacznie lepiej poradził sobie dla parametru $\Delta = 0.5$ niż $\Delta = 0.05$ co widać również na (**Rys.2**). Być może przy odpowiednim doborze parametru Δ ciąg wygenerowany za pomocą metody łańcucha Markowa przeszedłby test χ^2.