Generatory liczb pseudolosowych o zadanym rozkładzie w jednym wymiarze

Metody Monte Carlo w Fizyce

Julia Potempa (411073)

Akademia Górniczo-Hutnicza
im. Stanisława Staszica w Krakowie

19. marca 2025r.

1 Wstęp

1.1 Generator liczb pseudolosowych

Generator liczb pseudolosowych jest algorytmem matematycznym, który generuje liczby przypominające liczby losowe. Nazywa się je pseudolosowymi, przez to, że w rzeczywistości znając algorytm jesteśmy w stanie przewidzieć, jaka liczba zostanie "wylosowana". Dodatkowo takie generatory mają swój okres, to znaczy, że losowane w ten sposób liczby po pewnym czasie będą się powtarzały. Inna cechą takiego generatora jest ziarnistość, czyli próbkowanie przestrzeni dopuszczalnych rozwiązań tylko w wybranych punktach.

1.2 Wybrane metody implementacji

1.2.1 Rozkład złożony

Rozkład złożony wykorzystuje postać wielomianową dystrybuanty, którą można dzięki temu zapisać jako sumę:

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

gdzie: g_i to dystrybuanta rozkładu dyskretnego oraz $H_i(x)$ to dystrybuanta podlegająca superpozycji.
Dzięki takiemu rozwinięciu możemy odnaleźć funkcje odwrotne $H^{-1}(x)$ do funkcji $H(x)$, czyli odpowiednio wartości naszych zmiennych losowych w funkcji H.

1.2.2 Łańcuch Markowa

Metoda łańcucha Markowa polega na generowaniu następnego elementu na podstawie wartości poprzedniego. Ogólnie, jeżeli rozkład warunkowy X_{i+1} zależy wyłącznie od zmiennej X_i to taki proces stochastyczny posiada własność Markowa.
Związek pomiędzy poprzednim i następnym wyrazem łańcucha określa warunek równowagi (z ang. *detailed balance*):

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

gdzie dobieramy wartość Δ, natomiast przebieg algorytmu reguluje odpowiednio określone prawdopodobieństwo akceptacji p_{acc}.

1.2.3 Metoda eliminacji

W tej metodzie definiujemy dodatkową funkcję $g(x)$ o innym rozkładzie $\mathcal{G}$, która ogranicza naszą funkcję gęstości prawdopodobieństwa. W kolejnych krokach porównujemy wartość wylosowaną dla rozkładu $\mathcal{G}$ z wartością funkcji gęstości prawdopodobieństwa w punkcie określonym rozkładem $U(0,1)$. Jeżeli przekroczy ona wybraną przez nas funkcję $g(x)$, to zostaje ona odrzucona.

2 Problematyka

Ćwiczenie polegało na zaimplementowaniu w języku `C++` trzech metod generowania rozkładu zadanego funkcją gęstości prawdopodobieństwa:

$$f(x) = \frac{4}{5}\left(1 + x - x^3\right) \tag{1}$$

i dystrybuantą:

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

Dla metody rozkładu złożonego po doprowadzeniu dystrybuanty do postaci nieujemnej otrzymujemy wartości $g_1 = \frac{4}{5}$, $g_2 = \frac{1}{5}$ oraz funkcje $H_1(x) = x$, $H_2(x) = 2x^2 - x^4$. Dla nich obliczamy funkcje odwrotne $x = U$ oraz $x = \sqrt{1 - \sqrt{1-U}}$, dla $U \sim U(0,1)$. Algorytm zatem ma strukturę:

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

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

Dla łańcucha Markowa postać prawdopodobieństwa akceptacji kolejnej wartości przyjmuje formę:

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

Co określa nam ostatecznie postać algorytmu, w którym zastosowano wartości $\Delta = 0,5$ oraz $\Delta = 0,05$:

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

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

Ostatnią metodą jest metoda eliminacji, w której funkcja gęstości prawdopodobieństwa jest ograniczona od góry inną funkcją $g(x)$ o innym rozkładzie $\mathcal{G}$.

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

$$G_2 \sim \mathcal{G} = 1,15 \cdot U(0,1)$$

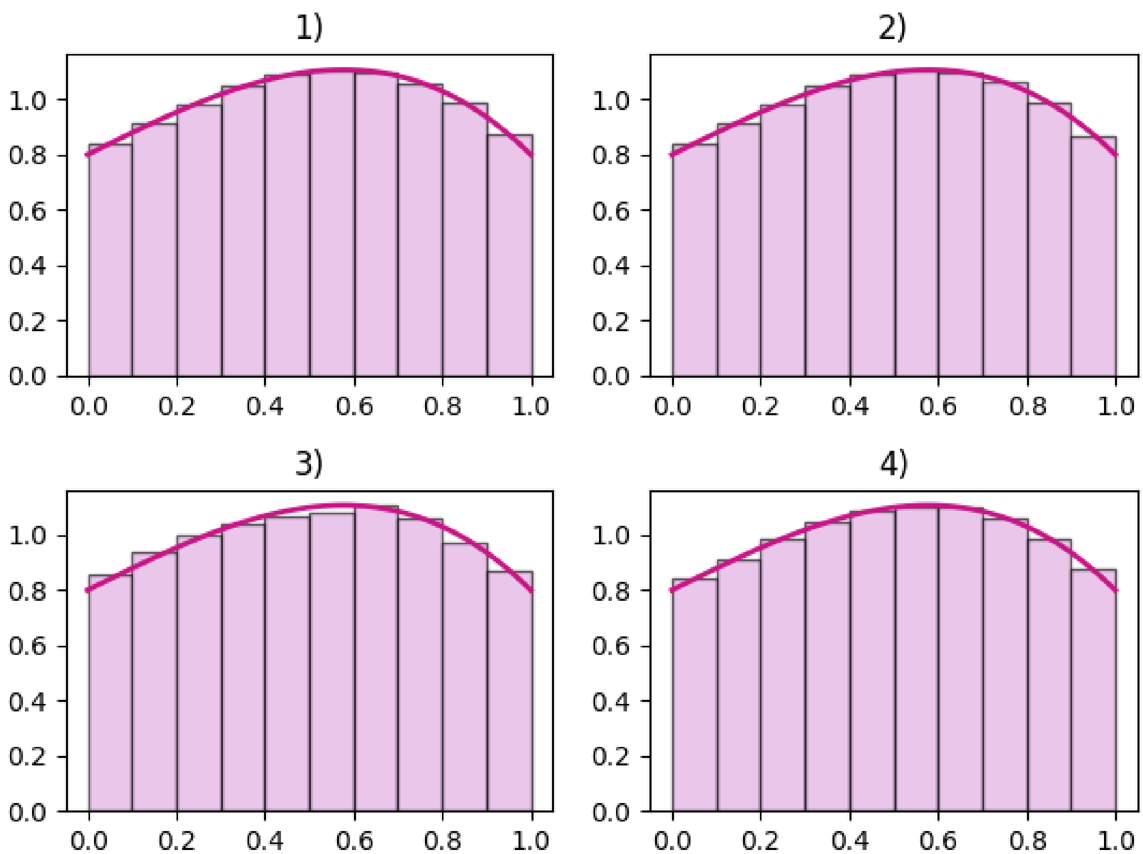
$$\begin{cases} X = U_1 & \text{gdy} \quad G_2 \leqslant f(U_1) \\ \text{losowanie nowej pary } (G_2, U_1), & \text{gdy} \quad G_2 > f(U_1) \end{cases}$$

Dodatkowo dla wszystkich zastosowanych metod został przeprowadzany test χ^2 o poziomie istotności $\alpha = 0,05$ dla $k-1$ stopni swobody, gdzie $k = 10$. Dla takich parametrów $\chi^2_{\text{kryt}} = 16,919$. Prawdopodobieństwo p_i jest obliczane na podstawie dystrybuanty:

$$p_i = F\left((i+1)\cdot\frac{1}{k}\right) - F\left(i\cdot\frac{1}{k}\right)$$

Wszystkie wykresy zostały narysowane za pomocą biblioteki `matplotlib` w języku `Python`, w którym też przeprowadzono wspomniany test χ^2.

3 Wyniki



Rysunek 1: Histogram ze zliczeniami wartości w danych przedziałach oraz krzywa funkcji gęstości prawdopodobieństwa, gdzie 1) rozkład złożony, łańcuchy Markowa dla 2) $\Delta = 0,5$, 3) $\Delta = 0,05$ oraz 4) metoda eliminacji.

Na otrzymanych wykresach widać, że wszystkie z wykorzystanych metod dość dobrze dopasowały się do zadanego rozkładu, natomiast statystyczna zgodność została przetestowana za pomocą testu χ^2, którego wyniki widnieją w tabeli 1.

Tabela 1: Tabela przedstawiająca wartość tablicową dla poziomu istotności $\alpha = 0,05$ oraz otrzymane wyniki testu χ^2 dla zastosowanych metod.

χ^2_{crit}	Rozkład złożony	Łańcuch Markowa dla $\Delta = 0,5$	Łańcuch Markowa dla $\Delta = 0,05$	Metoda eliminacji
16,919	6,234	7,060	251,427	4,548

Otrzymane wartości wykazały, że dla rozkładu złożonego, metody eliminacji oraz łańcucha Markowa dla $\Delta = 0,5$ hipoteza nie została odrzucona, co wskazuje na to, że były one najlepiej dopasowane do zadanej dystrybuanty, natomiast łańcuch Markowa o $\Delta - 0,05$ miał odchyły wykraczające poza ramy akceptowalności.

4 Podsumowanie

W tym ćwiczeniu zostały zastosowane trzy metody modyfikacji generatora liczb pseudolosowych: rozkład złożony, łańcuch Markowa oraz metoda eliminacji celem uzyskania rozkładu zdefiniowanego przez odpowiednią funkcję gęstości prawdopodobieństwa $f(x)$ oraz dystrybuantę $F(x)$. Uzyskane tymi metodami rozkłady przyjęły postać wizualnie bardzo dobrze dopasowaną, natomiast test χ^2 wykazał niezgodność dla metody łańcucha Markowa dla $\Delta = 0,05$. Metoda rozkładu złożonego, eliminacji oraz łańcuch Markowa z $\Delta = 0,5$ były zgodne na zadanym poziomie istotności $\alpha = 0,05$.