Generatory liczb pseudolosowych o zadanym rozkładzie w jednym wymiarze

4 marca 2024

Na zajęciach skonstruujemy generator jednowymiarowy o funkcji gęstości prawdopodobieństwa

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

i dystrybuancie

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

Do generowania liczb pseudolosowych użyjemy shcematów dla: a) rozkładu złożonego, b) łańcucha Markowa, c) metody eliminacji

1 Algorytmy

1.1 Rozkład złożony

Dystrybuanta rozkładu jest wielomianem zatem możemy spróbować zapisać ją w postaci rozkładu złożonego

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

gdzie: g_i to dystrybuanta rozkładu dyskretnego, a $H_i(x)$ dystrybuanty rozkładów podlegających superpozycji. Zgodnie z definicją dystrybuanta musi być funkcją nieujemną, zatem musimy przekształcić $F(x)$ do postaci akceptowalnej

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

skąd odczytujemy

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

$$g_2 = \tfrac{1}{5}, \quad H_2 = 2x^2 - x^4 \tag{6}$$

Teraz należy znaleźć funkcje odwrotne H_1^{-1} i H_2^{-1} aby otrzymać liczbę losową $X = x$. H_1 jest liniowa więc od razu dostajemy

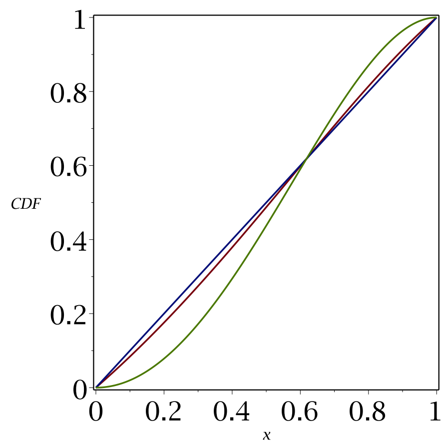
$$H_1(x) = x = U \sim U(0, 1) \tag{7}$$

Dla H_2^{-1}

$$H_2(x) = 2x^2 - x^4 = U \sim U(0, 1) \tag{8}$$

Mamy równanie 4 stopnia

$$x^4 - 2x^2 + U = 0 \tag{9}$$



Rysunek 1: Dystrybuanty: $F(x)$-niebieski, $H_1(x)$-czerwony, $H_2(x)$-zielony.

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

$$y^2 - 2y + U = 0, \quad y_1, y_2 = 1 \pm \sqrt{1-U} \tag{10}$$

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

$$x = \sqrt{1 - \sqrt{1-U}} \tag{11}$$

Algorytm dla rozkładu złożonego

$$U_1, U_2 \sim U(0,1) \tag{12}$$

$$X = \begin{cases} U_2, & \text{gdy } U_1 \leqslant g_1 \\ \sqrt{1-\sqrt{1-U_2}}, & \text{gdy } U_1 > g_1 \end{cases} \tag{13}$$

1.2 Łańcuch Markowa

W metodzie tej generujemy ciąg

$$\{X_0,\, X_1,\, X_2, \ldots\} \tag{14}$$

gdzie związek pomiędzy ostatnim elementem X_i a kolejnym X_{i+1} określamy na podstawie prawdopodobieństwa przejścia, które spełnia warunek **detailed balance**

$$T(X_{i+1}|X_i) = T(X_i|X_{i+1}) = \frac{1}{2\Delta}, \quad \Delta \in [0,1] \tag{15}$$

i prawdopodobeińśtwa akceptacji nowego stanu (liczby)

$$p_{acc} = \min\left\{\frac{T(X_i|X_{i+1})f(X_{i+1})}{T(X_{i+1}|X_i)f(X_i)}, 1\right\} \tag{16}$$

Algorytm Metropolisa generowania nowego elementu w łańcuchu

$$U_1, U_2 \sim U(0,1) \tag{17}$$

$$X_{i+1} = \begin{cases} x_{new} = X_i + (2U_1 - 1)\Delta, & \text{gdy } x_{new} \in [0,1] \wedge U_2 \leqslant p_{acc} \\ X_i, & \text{w przeciwnym przypadku} \end{cases} \tag{18}$$

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 $\mathcal{G}$. Algorytm generowania ciągu liczb metodą eliminacji

$$U_1 \sim U(0,1) \tag{19}$$

$$G_2 \sim \mathcal{G}, \qquad np.\, \mathcal{G} = 1.15 \cdot U(0,1) \tag{20}$$

$$\begin{cases} G_2 \leqslant f(U_1) \Longrightarrow X = U_1 \\ G_2 > f(U_1) \Longrightarrow \text{ losujemy nową parę } U_1, G_2 \end{cases} \tag{21}$$

2 Zadania do wykonania

1. Pisząc program w języku $C/C++$ do wygenerowania liczb o rozkładzie jednorodnym $U(0,1)$ można wykorzystać funkcję

```
double uniform(){
  return rand()/(double)RAND_MAX;
}
```

2. Wygenerować $N = 10^6$ liczb pseudolosowych o dystrybuancie $F(x)$ dla: (a) dla rozkładu złożonego, (b) łańcucha Markowa oraz (c) metody eliminacji.
   Metodą łańcucha Markowa wygenerować dwa ciągi liczb dla: (a) $\Delta = 0.5$ oraz (b) $\Delta = 0.05$.

3. Dla każdego ciągu liczb sporządzić histogram o liczbie podprzedziałów $k = 10$ i porównać go z funkcją $f(x)$ (uwaga: na rysunku wysokość każdego słupka należy przeskalować czynnikiem $1/k$)

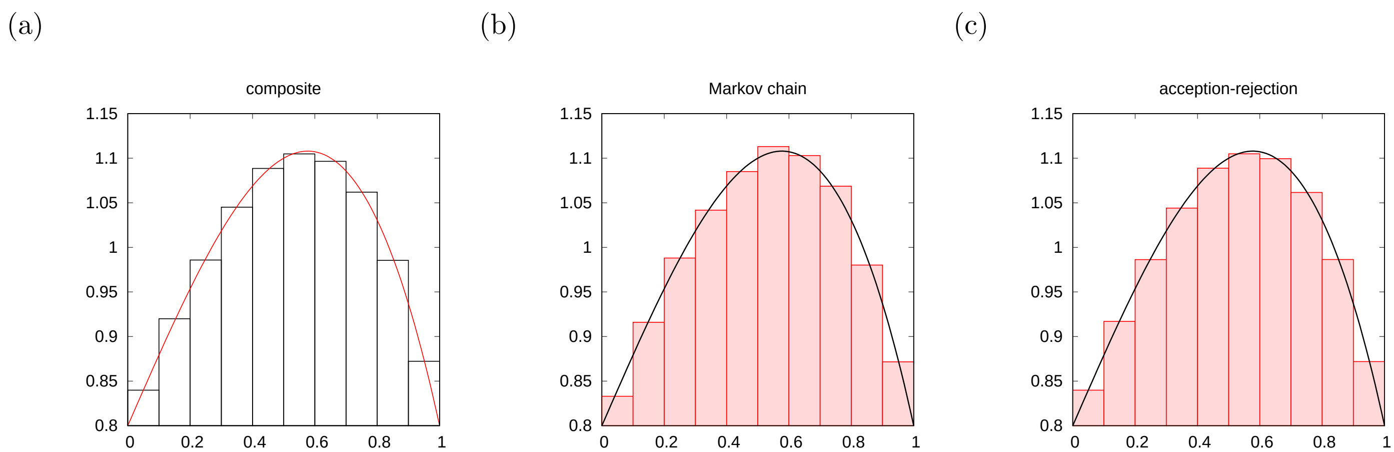
4. Dla każdej metody wykonać test χ^2 i porównać uzyskane wyniki z wartością graniczną rozkładu przyjmując poziom istotności równy $\alpha = 0.05$ - na podstawie porównania stwierdzić czy hipotezę H_0 tego, że uzyskany ciąg liczb pseudolosowych ma rozkład $F(x)$ należy odrzucić czy nie. Wartość statystyki testowej dla $k-1$ stopni swobody liczymy ze wzoru

$$\chi^2_{k-1} = \sum_{i=1}^{k} \frac{(n_i - p_i N)^2}{p_i N} \tag{22}$$

   gdzie: p_i to prawdopodobieństwo że zmienna losowa znajdzie się w przedziale i-tym (do wyznaczenia używamy dystrybuanty), n_i ilość liczb pseudolosowych, które znalazły się w i-tym przedziale, N to całkowita ilość wylosowanych liczb.

5. Przeprowadzić analizę różnic obu wygenerowanych łańcuchów Markowa.

3 przykładowe wyniki



Rysunek 2: Histogramy dla trzech metod generowania liczb pseudolosowych o dystrybuancie $F(x)$.