Metody Probabilistyczne i Statystyka

ZADANIE DOMOWE 3

Termin wysyłania (MS Teams): **15 stycznia 2023 godz. 23:59**

Za rozwiązanie zadań z tej listy można uzyskać łącznie **10 pkt**.

Zadanie 1. [3 pkt.] (*Testy NIST*, patrz Zadanie 1. z Listy 9.)

Poddaj testom NIST:

(a) output „słabego” generatora liczb losowych (np. generatora typu *linear congruential generator, LCG*) z wybranego języka programowania (np. C, Java, Python, ...),

(b) output „przyzwoitego” generatora liczb losowych (np. generatora *Mersenne Twister*) z wybranego języka programowania (np. C, Java, Python, ...),

(c) output zwrócony przez funkcję hashującą SHA-1 dla własnego nazwiska: `SHA-1(własne_nazwisko)`. Do obliczenia hasha `SHA-1()` użyj wybranego serwisu online, np. `https://emn178.github.io/online-tools/sha1.html`.

Do przeprowadzenia testów NIST użyj wybranego serwisu online, np. strony Zsolta Molnara: `https://mzsoltmolnar.github.io/random-bitstream-tester/`.

Zwięźle omów wyniki przeprowadzonych testów oraz płynące z nich wnioski.

Zadanie 2. [3 pkt.] (*Błądzenie losowe na liczbach całkowitych*, patrz Zadanie 1. z Listy 10.)

Zdefiniujmy zmienną losową $S_N = \sum_{n=1}^{N} X_n$, gdzie zmienne losowe X_n, $1 \leqslant n \leqslant N$, są niezależne i każda przyjmuje wartości 1 oraz -1 z prawdopodobieństwem 0.5. Dla $N = 0$ przyjmijmy $S_0 = 0$.

(a) Wyznacz numerycznie dystrybuantę zmiennej losowej S_N dla $N \in 5, 10, 15, 20, 25, 30$ (wygeneruj odpowiednie histogramy).

(b) Porównaj wyznaczone dystrybuanty z dystrybuantą rozkładu normalnego, która miałaby aproksymować S_N.

(c) Powtórz punkty (a) i (b) dla $N = 100$.

Zwięźle omów uzyskane wyniki oraz płynące z nich wnioski.

W rozwiązaniu zadania skorzystaj z wybranego narzędzia / pakietu obliczeń matematycznych (np. Matlab, Wolfram *Mathematica*, ...). W Matlabie do obliczania dystrybuanty rozkładu normalnego możesz użyć funkcji `normcdf`. W Mathematice do wyznaczania dystrybuanty służy funkcja `CDF`.

Zadanie 3. [4 pkt.] (*Błądzenie losowe na $\mathbb{Z}$ – rozkład „czasu nad osią OX”*)

Niech $X_1, X_2, \ldots$ będzie ciągiem niezależnych zmiennych losowych o rozkładzie jak w Zadaniu 2. i niech $S_N = \sum_{n=1}^{N} X_n$ dla $N \in \mathbb{N}$. Ciąg zmiennych losowych (proces losowy) $(S_N)_{N \in \mathbb{N}}$ nazywamy prostym błądzeniem losowym na liczbach całkowitych (startujemy w punkcie 0 i w każdym momencie czasu $N \geqslant 1$ niezależnie z jednakowym prawdopodobieństwem idziemy jeden krok w górę ($X_N = 1$) albo jeden krok w dół ($X_N = -1$); przykładowa trajektoria takiego błądzenia losowego dla pierwszych 13 kroków przedstawiona jest na rys. 1).

Celem tego zadania jest eksperymentalne zbadanie rozkładu frakcji czasu, którą rozważany proces „spędza nad osią OX". Formalnie, niech $D_n = \mathbb{1}(S_n > 0 \vee S_{n-1} > 0), n = 1, 2, \ldots$ ($D_n = 1$, gdy błądzenie losowe w chwili n albo znajdowało się nad osią OX, albo „zeszło w dół do 0 "; w przeciwnym przypadku $D_n = 0$). Niech ponadto $L_N = \sum_{n=1}^{N} D_n$ (L_n zlicza, w ilu momentach czasu od 1 do N błądzenie losowe znajdowało się nad osią OX). Oznaczmy $P_N = L_N/N$ („frakcja czasu" – dzielimy L_N przez łączną liczbę kroków N; $P_N \in [0, 1]$).

Dla każdego $N \in \{100, 1\,000, 10\,000\}$ wykonaj następujące kroki.

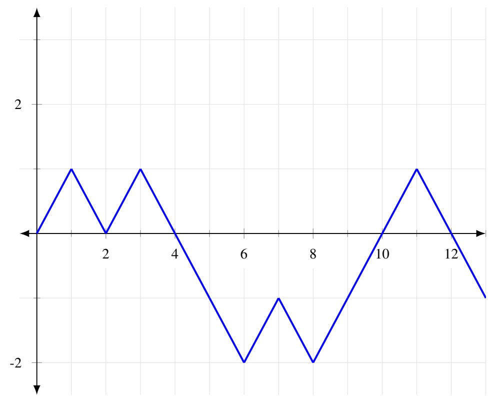
(a) Wygeneruj niezależnie $k = 5\,000$ realizacji procesu błądzenia losowego (S_N) (zasymuluj pierwsze N kroków) – oznaczmy te realizacje przez $(S_N)^{(1)}, \ldots, (S_N)^{(k)}$.

(b) Dla każdej realizacji błądzenia losowego $(S_N)^{(i)}$, $1 \leqslant i \leqslant k$, wyznacz wartość $P_N^{(i)}$ – frakcja czasu, którą błądzenie losowe „spędziło nad osią OX".

(c) Dla tak wyznaczonych wartości $(P_N^{(1)}, \ldots, P_N^{(k)})$ wygeneruj histogram z 20 „kubełkami" (ang. *bins*) równej szerokości (podziel przedział $[0, 1]$ na 20 rozłącznych podprzedziałów równej szerokości). Postaraj się znormalizować pionową oś histogramu tak, aby zamiast liczby wartości w poszczególnych kubełkach wyznaczona była estymacja funkcji gęstości prawdopodobieństwa. W tym celu np. w Matlabie w funkcji `histogram` możesz użyć opcji `'Normalization'` z wartością `'pdf'` (patrz dokumentacja; przykład: `histogram(X,'Normalization','pdf')`), a w Mathematice w funkcji `Histogram` możesz jako wartość argumentu `hspec` podać `"PDF"` (patrz dokumentacja).

(d) Porównaj uzyskany histogram z wykresem gęstości rozkładu arcusa sinusa z Zadania 7. z Listy 7.

Zwięźle omów uzyskane wyniki oraz płynące z nich wnioski.



Rysunek 1: Przykładowa realizacja procesu prostego błądzenia losowego na liczbach całkowitych – pierwsze 13 kroków.

Rozwiązanie zadania obejmujące

- implementacje symulacji (kod źródłowy w wybranym języku programowania) oraz
- pdf z wykresami, zwięzłym opisem wyników, wnioskami i odpowiedziami do zadań.

należy przesłać na platformę MS Teams. Nie należy dołączać żadnych zbędnych plików.