Rachunek Prawdopodobieństwa

ZADANIE DOMOWE 3

Termin wysyłania (MS Teams): **23 grudnia 2025 r. godz. 23:59**

Promocja świąteczno-noworoczna: spóźnienia będą naliczane dopiero od 01.01.2026 r.

Rozwiązania zadań 1 – 3 przedstaw w formie krótkiego sprawozdania.

- Przedstaw wyniki przeprowadzonych eksperymentów – zaprezentuj wykresy, zwięźle omów uzyskane rezultaty i wyciągnij wnioski.
- Na podstawie wykresów krótko scharakteryzuj koncentrację wyników uzyskanych w poszczególnych powtórzeniach wokół wartości średniej dla badanych wielkości.
- Na podstawie wykresów postaw hipotezy odnośnie asymptotyki wartości średnich badanych wielkości i uzasadnij ich wybór (w razie potrzeby w niektórych wykresach możesz zastosować skalę logarytmiczną).
- W przypadku zadań z ★ („teoretycznych"), przedstaw wymagane rozumowania i obliczenia oraz zwięźle omów wyniki i odpowiedz na pytania.

Za rozwiązania zadań 1 – 3 można uzyskać łącznie **7 punktów** („standardowe" 5 pkt + 2 pkt dodatkowe za rozwiązania zadań z ★).

Zadanie 1. [2 pkt] *The Power of Two Choices / ballanced allocation*, patrz np. rozdział 17 w [MU17]

Rozszerz symulacje z zadania domowego 2 o eksperymentalne wyznaczanie maksymalnej liczby kul w urnie po wrzuceniu n kul do n urn (*maximum load*) w przypadku, gdy:

(a) dla każdej kuli wybieramy niezależnie i jednostajnie losowo jedną z n urn, w której umieszczamy kulę,

(b) dla każdej kuli wybieramy niezależnie i jednostajnie losowo z powtórzeniami d urn i umieszczamy kulę w najmniej zapełnionej z wybranych urn („remisy" rozstrzygamy w dowolny sposób).

Przeprowadź podobne eksperymenty jak w zadaniu domowym 2 celem wyznaczenia $L_n^{(d)}$ – maksymalnego zapełnienia pojedynczej urny po wrzuceniu n kul – dla $d = 1$ (punkt (a)) oraz $d = 2$ (punkt (b)). Przedstaw uzyskane wyniki jak w zadaniu domowym 2.

Podobnie jak w zadaniu domowym 2, wykonaj dodatkowo wykresy $\frac{l_n^{(1)}}{f_1(n)}$ oraz $\frac{l_n^{(2)}}{f_2(n)}$, gdzie $l_n^{(d)}$ oznacza średnią wartość $L_n^{(d)}$ dla danego n oraz $d \in \{1, 2\}$, a funkcje f_1 i f_2 zadane są wzorami $f_1(n) = \frac{\ln n}{\ln \ln n}$ oraz $f_2(n) = \frac{\ln \ln n}{\ln 2}$ (por. rozdział 17 w [MU17]).

W celu uzyskania większej czytelności wyników możesz zwiększyć zakres n wykonując symulacje np. dla $n \in \{10\,000, 20\,000, \ldots, 1\,000\,000\}$ lub zastosować na wykresach skalę logarytmiczną.

Zadanie 2. [3 pkt = 2 + dodatkowy 1 pkt] *Sortowanie przez wstawianie losowych danych*

Uwaga. Podobne zadanie – dla większej liczby algorytmów sortujących – może się pojawić na laboratorium do kursu *Algorytmy i struktury danych* (4 semestr).

(a) Zaimplementuj oraz przetestuj działanie algorytmu sortowania przez wstawianie InsertionSort (patrz algorytm 1; jego szczegółowe omówienie można znaleźć np. w rozdziale 2.1 w [CLRS22]). W tym celu dla każdego $n \in \{100, 200, \ldots, 10\,000\}$ wykonaj po $k = 50$ niezależnych powtórzeń:

   i) generowania tablicy $A[1..n]$ będącej losową permutacją zbioru liczb $\{1, \ldots, n\}$ (wszystkie permutacje powinny być jednakowo prawdopodobne; podobnie jak poprzednio, zadbaj o dobry generator liczb pseudolosowych),

   ii) sortowania wygenerowanej tablicy A,

   iii) zapisywania statystyk obejmujących rozmiar danych n, liczbę wykonanych porównań między kluczami (elementami tablicy A) oraz liczbę przestawień kluczy.

(b) Po zakończeniu eksperymentów, korzystając z zebranych danych, przedstaw na wykresach:

   i) liczbę wykonanych porównań w poszczególnych powtórzeniach oraz średnią liczbę porównań $cmp(n)$ jako funkcję n,

   ii) liczbę przestawień w poszczególnych powtórzeniach oraz średnią liczbę przestawień kluczy $s(n)$ jako funkcję n,

   iii) iloraz $\frac{cmp(n)}{n}$ oraz $\frac{cmp(n)}{n^2}$ jako funkcje n,

   iv) iloraz $\frac{s(n)}{n}$ oraz $\frac{s(n)}{n^2}$ jako funkcje n.

★ (c) Niech S_n będzie liczbą przestawień w algorytmie InsertionSort podczas sortowania permutacji η zbioru liczb $\{1, \ldots, n\}$ wybranej jednostajnie losowo. Przeprowadź analizę teoretyczną (zgodnie z poniższymi wskazówkami), której celem jest wyznaczenie wzorów na $\mathbf{E}(S_n)$ oraz $\mathbf{var}(S_n)$ i porównanie ich z wynikami przeprowadzonych eksperymentów.

   i) Zauważ, że S_n jest równe liczbie inwersji w sortowanej permutacji η. Niech $\pi_\eta \in \Pi_n = \{0, \ldots, n-1\} \times \{0, \ldots, n-2\} \times \cdots \times \{0, 1\} \times \{0\}$ będzie wektorem inwersji permutacji η, tj. $\pi_\eta(i)$, $i \in \{1, \ldots, n\}$, zlicza liczbę elementów większych od i „na lewo od i" w permutacji η. Zauważ, że istnieje bijekcja między zbiorami $\mathbb{S}_n$ permutacji zbioru $\{1, \ldots, n\}$ a zbiorem wektorów inwersji Π_n. Wywnioskuj stąd, że jednostajnie losowy wybór permutacji $\eta \in \mathbb{S}_n$ jest równoważny jednostajnie losowemu wyborowi wektora $\pi \in \Pi_n$.

   ii) Niech zmienne losowe X_i, $i \in \{1, \ldots, n\}$, oznaczają i-tą współrzędną wektora inwersji π wybranego jednostajnie losowo z Π_n, tj. $X_i = \pi(i)$ (rzut wektora π na oś i). Zauważ, że zmienne losowe X_i są niezależne i mają rozkłady jednostajne, tj. $X_i \sim U(\{0, \ldots, n-i\})$.

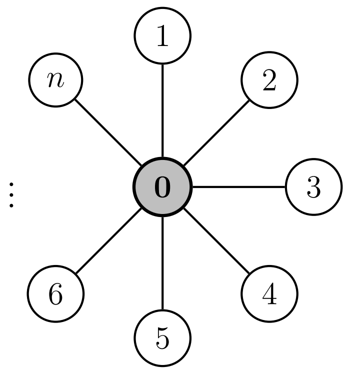
   iii) Wywnioskuj stąd, że $S_n = \sum_{i=1}^{n} X_i$. Wykorzystaj ten fakt do wyznaczenia wzorów na $\mathbf{E}(S_n)$ oraz $\mathbf{var}(S_n)$, a następnie oblicz σ_{S_n} oraz $\frac{\sigma_{S_n}}{\mathbf{E}(S_n)}$. Porównaj otrzymane wzory z wynikami przeprowadzonych eksperymentów.

Algorytm 1 Algorytm sortowania przez wstawianie (por. rozdział 2.1 w [CLRS22]).

procedure InsertionSort($A[1..n]$)
  for $j = 2$ **to** n **do**
    $key \leftarrow A[j]$
    // Wstaw $A[j]$ w posortowany podciąg $A[1..j-1]$.
    $i \leftarrow j - 1$
    while $i > 0$ **and** $A[i] > key$ **do**
      $A[i+1] \leftarrow A[i]$
      $i \leftarrow i - 1$
    end while
    $A[i+1] \leftarrow key$
  end for
end procedure

Zadanie 3. [2 pkt = 1 + dodatkowy 1 pkt] *Uproszczony model komunikacji z zakłóceniami*

Rozważmy uproszczony model komunikacji w sieci o topologii gwiazdy, składającej się z $n+1$ węzłów $\{0, 1, \ldots, n\}$ (patrz rys. 1). Początkowo centralna stacja (węzeł 0 oznaczony na rysunku kolorem szarym) posiada informację, którą chce rozesłać do wszystkich pozostałych węzłów. W tym celu w pojedynczej rundzie rozgłasza wiadomość, którą – z uwagi na zakłócenia – każda stacja $i > 0$ odbiera niezależnie z prawdopodobieństwem p. W kolejnych rundach stacja 0 powtarza proces nadawania tej wiadomości[1]. Niech T_n oznacza pierwszą rundę, po której wszystkie węzły w sieci posiadają informację rozsyłaną przez stację 0.



Rysunek 1: Graf rozważanej sieci o $n+1$ węzłach (stacjach).

(a) Zaprojektuj i przeprowadź symulacje (podobne do tych z poprzednich zadań), których celem jest eksperymentalne zbadanie minimalnej liczby rund T_n – w zależności od wartości n – potrzebnej do rozesłania informacji ze stacji 0 do wszystkich pozostałych stacji w przypadku, gdy $p = 0{,}5$ oraz $p = 0{,}1$.

(b) Po wykonaniu eksperymentów przedstaw na wykresach (po jednym dla każdej testowanej wartości p) liczbę potrzebnych rund w poszczególnych powtórzeniach oraz średnią liczbę rund $t(n)$ jako funkcję n.

★ (c) Niech $T_{n,i}$, $i \in \{1, \ldots, n\}$, oznacza rundę, w której stacja i po raz pierwszy odebrała wiadomość. Zidentyfikuj rozkład zmiennych losowych $T_{n,i}$ oraz wyraź zmienną losową T_n jako pewną funkcję i.i.d.[2] zmiennych losowych $T_{n,i}$, $i \in \{1, \ldots, n\}$.

[1] Rozważamy uproszczony model, więc nie uwzględniamy żadnego mechanizmu potwierdzeń odbioru itp.

[2] Przypomnijmy, że skrót i.i.d. oznacza niezależne zmienne losowe o tym samym rozkładzie.

★ (d) Poszukaj w literaturze informacji na temat asymptotyki wartości oczekiwanej zmiennej losowej T_n o tak określonym rozkładzie. Podaj znalezioną formułę (wystarczy dominujący składnik zależny od n i p) wraz z odpowiednim cytowaniem (musi być to książka albo artykuł naukowy).

★ (e) Porównaj znaleziony wzór z punktu (d) z otrzymanymi wynikami przeprowadzonych eksperymentów (podobnie jak w poprzednich zadaniach, gdzie eksperymentalnie określaliśmy asymptotykę badanych wielkości).

Rozwiązanie obejmujące implementację (kody źródłowe) oraz plik pdf ze sprawozdaniem należy przesłać przez platformę MS Teams. Nie należy dołączać żadnych zbędnych plików.

Literatura

[CLRS22] Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, and Clifford Stein. *Introduction to Algorithms*. The MIT Press, 4th edition, 2022.

[MU17] Michael Mitzenmacher and Eli Upfal. *Probability and Computing. Randomization and Probabilistic Techniques in Algorithms and Data Analysis*. Cambridge University Press, 2nd edition, 2017.