

Analiza Algorytmów 2021/2022 (zadania na ćwiczenia)

Wybór lidera - 14 III 2022

- 1 – ✗ Wyznacz wartość oczekiwaną dla zmiennej losowej $X \sim Geo(p)$.
- 2 – ✗ Wyznacz wariancję dla zmiennej losowej $X \sim Geo(p)$.
- 3 – ✗ Niech $p \in [0, 1]$ oraz $n \geq k \geq 1, n, k \in \mathbb{N}$. Dla jakich wartości argumentu funkcja f przyjmuje wartości maksymalne?
- $f(p) = np(1-p)^{n-1}$,
 - $f(n) = np(1-p)^{n-1}$,
 - $f(k) = \binom{n}{k} p^k (1-p)^{n-k}$.
- 4 – ✗ Udowodnij: $(1+x)^r \geq 1+rx$ dla $x \geq -1, r \geq 1$.
- 5 – ✗ Udowodnij: $(1+x)^r \leq 1+rx$ dla $x \geq -1, r \in (0, 1)$.
- 6 – ✗ Udowodnij: $1+x \leq e^x$ dla $x \in \mathbb{R}$.
- 7 – ✗ Udowodnij: $\frac{x}{e^x} < \frac{1.5}{x^2}$ dla $x > 0$.
- 8 – ✗ Niech $f_i(n) = n \frac{1}{2^i} \left(1 - \frac{1}{2^i}\right)^{n-1}$. Udowodnij, że funkcje $f_i(2^{i-1}), f_{i+1}(2^{i-1})$ są malejące oraz, że funkcja $f_{i-1}(2^i)$ jest rosnąca dla $i \geq 2$.
- 9 – ✗ Przedstaw definicję rodziny funkcji [W Lamberta](#) i wykres jej gałęzi rzeczywistych.
- 10 – ✗ Wykorzystaj funkcję W Lamberta by analitycznie [wyznaczyć](#) rzeczywiste rozwiązania równania $3^x = x^3$. Jak nazwa się w Mathematica funkcja W Lamberta?
- 11 – ✗ Zajrzyj do [tej pracy](#) i udowodnij, że dla $x \geq e$

$$\ln x - \ln \ln x < W_0(x) \leq \ln x - \frac{1}{2} \ln \ln x .$$

- 12 – ✗ Uzupełnienie dowodu z wykładu. Sprawdź, że jeśli $K \geq 1, f > 1, u \geq 2$ oraz

$$3e(K+1)u^{\frac{-1}{2(K+1)}} \geq 1 - \frac{1}{f} \quad \text{to} \quad K \geq \frac{\ln u}{2W_0\left(\frac{3e}{2} \frac{f}{f-1} \ln u\right)} - 1 .$$

Analiza strumieni danych: przybliżone zliczanie

13 – x Dla ciągłych i niezależnych zmiennych losowych X_1, X_2, \dots, X_n o tym samym rozkładzie zadanym funkcją gęstości $f(x)$ i dystrybuantą $F(x)$ pokaż, że k -ta statystyka pozycyjna $X_{k:n}$ ma rozkład opisany funkcją gęstości

$$f_k(x) = \frac{F^{k-1}(x) [1 - F(x)]^{n-k} f(x)}{B(k, n - k + 1)},$$

gdzie $B(\alpha, \beta)$ oznacza [funkcję beta](#). Wskazówka: możesz przejrzeć [te wykłady](#).

14 – x Dla n niezależnych zmiennych losowych U_1, U_2, \dots, U_n o rozkładzie jednostajnym: $U_i \sim \mathcal{U}(0, 1)$, pokaż, że k -ta statystyka pozycyjna ma rozkład $Beta(k, n - k + 1)$ i wartość oczekiwaną równą $k/(n + 1)$.

Wskazówka. W tym oraz w następnym zadaniu wykorzystaj różne reprezentacje funkcji beta: podaną w definicji oraz [przez silnię](#) dla argumentów będących liczbami naturalnymi

15 – x Niech $U_{k:n}$ oznacza k -tą statystykę pozycyjną dla n niezależnych zmiennych losowych o rozkładzie jednostajnym $\mathcal{U}(0, 1)$. Pokaż, że dla estymatora $\hat{n}_k = \frac{k-1}{U_{k:n}}$ oraz $k \geq 2$ mamy $\mathbb{E}(\hat{n}_k) = n$ oraz, że dla $k \geq 3$ mamy

$$\text{Var}(\hat{n}_k) = \frac{n(n - k + 1)}{k - 2}.$$

16 – x (Nierówność Markowa) Niech X oznacza zmienną losową, która przyjmuje tylko nieujemne wartości. Wtedy dla wszystkich $a > 0$

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}(X)}{a}.$$

17 – x (Nierówność Czebyszewa) Niech X oznacza zmienną losową o skończonej wartości oczekiwanej i skończonej, niezerowej wariancji. Pokaż, że dla dowolnego $a > 0$ zachodzi nierówność:

$$\mathbb{P}(|X - \mathbb{E}(X)| < a) > 1 - \frac{\text{Var}(X)}{a^2}.$$

Wskazówka: wykorzystaj nierówność Markowa.

18 – x (Nierówność Chernoffa dla sumy prób Bernoulliego) Niech X_1, \dots, X_n będą niezależnymi próbami Bernoulliego takimi, że $\mathbb{P}(X_i = 1) = p_i$. Niech $X = \sum_{i=1}^n X_i$ oraz $\mu = \mathbb{E}(X)$. Pokaż, że

a) dla dowolnego $\delta > 0$

$$\mathbb{P}(X \geq (1 + \delta)\mu) \leq \left(\frac{e^\delta}{(1 + \delta)^{(1 + \delta)}} \right)^\mu,$$

b) dla dowolnego $0 < \rho \leq 1$

$$\mathbb{P}(X \leq (1 - \rho)\mu) \leq \left(\frac{e^{-\rho}}{(1 - \rho)^{(1 - \rho)}} \right)^\mu.$$

Wskazówka: możesz zajrzeć do książki do rozdziału 4.2. w [tej książce](#).

19 – ✗ Wykorzystując oznaczenia i uzyskane w poprzednim zadaniu nierówności, pokaż, że dla dowolnego $0 < \delta < 1$

$$\mathbb{P}(|X - \mu| \geq \delta\mu) \leq 2e^{-\mu\delta^2/3}.$$

20 – ✗ Niech S_n będzie liczbą orłów w n rzutach symetryczną monetą. Pokaż, że

a) stosując nierówność Czebyszewa mamy

$$\mathbb{P}\left(\left|S_n - \frac{n}{2}\right| \geq \frac{n}{4}\right) \leq \frac{4}{n},$$

b) stosując nierówność Chernoffa z poprzedniego zadania mamy

$$\mathbb{P}\left(\left|S_n - \frac{n}{2}\right| \geq \frac{n}{4}\right) \leq 2e^{-n/24}.$$

21 – ✗ Rozważ następujący algorytm, z którego wywodzi się idea algorytmu HyperLogLog.

Probabilistic Counter

1: Initialization: $C \leftarrow 1$

Upon event:

2: **if** random() $\leq 2^{-C}$ **then** ▷ random zwraca liczbę losową z zakresu $[0, 1)$

3: $C \leftarrow C + 1$

4: **end if**

Innymi słowy, przy wystąpieniu zdarzenia rzucamy monetą C razy i jeśli za każdym razem otrzymujemy reszkę zwiększamy licznik C o jeden. W przeciwnym razie nie robimy nic. Niech C_n oznacza wartość przechowywaną w liczniku C po zaobserwowaniu n zdarzeń. Pokaż, że $\mathbb{E}(2^{C_n}) = n + 2$ oraz $\text{Var}(2^{C_n}) = \frac{1}{2}n(n + 1)$. W oparciu o C_n zdefiniuj nieobciążony estymator wartości n i policz jego wariancję.

Analiza strumieni danych: przybliżone sumowanie

22 – ✗ Przypomnij sobie podstawowe informacje o rozkładzie wykładniczym.

a) Przypomnij wzór na gęstość i dystrybuantę. Wyprowadź wzór na wartość oczekiwaną i wariancję.

b) Załóżmy, że dysponujesz generatorem zwracającym liczby z przedziału $[0, 1)$ zgodnie z rozkładem jednostajnym. Przedstaw procedurę, która będzie przekształcała zwracane wartości, w wartości zgodne z rozkładem wykładniczym z parametrem λ .

Wskazówka: zobacz strony 28 i 29 w tej [książce](#).

c) Przedstaw i udowodnij twierdzenie na którym opera się procedura z punktu b).

23 – ✗ Niech S_1, S_2, \dots, S_n będzie ciągiem niezależnych zmiennych losowych o rozkładzie wykładniczym oraz $S_i \sim \text{Exp}(\lambda_i)$ dla $\lambda_i > 0$. Niech $\Lambda = \lambda_1 + \lambda_2 + \dots + \lambda_n$. Pokaż, że pierwsza statystyka pozycyjna $S_{\min} = \min\{S_1, S_2, \dots, S_n\}$ ma rozkład wykładniczy z parametrem Λ :

$$S_{\min} \sim \text{Exp}(\Lambda).$$

24 – x Niech X i Y będą niezależnymi zmiennymi losowymi o funkcjach gęstości odpowiednio $f_X(x)$ oraz $f_Y(y)$. Dla $Z = X + Y$ pokaż, że

$$f_Z(z) = \int_{-\infty}^{+\infty} f_X(x)f_Y(z-x)dx.$$

Jaki związek ma to zadanie z następnym zadaniem?

Wskazówka: zobacz [splot rozkładów prawdopodobieństwa](#).

25 – x Przyjmijmy $\Lambda > 0, m \in \mathbb{N}_{>0}$ oraz niech $S_{\min}^{(1)}, S_{\min}^{(2)}, \dots, S_{\min}^{(m)}$ będą niezależnymi zmiennymi losowymi o takim samym rozkładzie wykładniczym

$$S_{\min}^{(i)} \sim \text{Exp}(\Lambda).$$

Pokaż, że zmienna

$$G_m = \sum_{i=1}^m S_{\min}^{(i)}$$

ma rozkład gamma zdefiniowany gęstością

$$g_m(x) = \Lambda \frac{(\Lambda x)^{m-1}}{\Gamma(m)} e^{-\Lambda x} \quad \text{dla } x > 0.$$

Wskazówka 1: wykorzystaj poprzednie zadanie i zasadę indukcji.

Wskazówka 2: $\Gamma(m) = (m-1)!$ dla $m \in \mathbb{N}_{>0}$.

26 – x Przy oznaczeniach z zadania 25 pokaż, że dla $m \geq 2$ i estymatora zdefiniowanego jako

$$\bar{\Lambda}_m = \frac{m-1}{\sum_{i=1}^m S_{\min}^{(i)}}$$

mamy $\mathbb{E}(\bar{\Lambda}_m) = \Lambda$.

27 – x Przy oznaczeniach z zadań 25 i 26 pokaż, że dla $m \geq 3$ błąd standardowy estymatora $\bar{\Lambda}_m$ zależy jedynie od parametru m i wyraża się wzorem:

$$\text{SE}(\bar{\Lambda}_m) = \frac{1}{\sqrt{m-2}}.$$

28 – x Zaproponuj algorytm, który umożliwi estymację średniej wartości

$$Av = \frac{\lambda_1 + \lambda_2 + \dots + \lambda_n}{n}$$

dla wszystkich unikalnych elementów strumienia \mathfrak{M} . Wyznacz obciążenie zaproponowanego estymatora.

Wskazówka: zauważ, że dla $\lambda_1 = \lambda_2 = \dots = \lambda_n = 1$ mamy $\mathbb{E}(\bar{\Lambda}_m) = n$.

Blockchain

29 – \times Pokaż, że dla oznaczeń z zadania **23** mamy:

$$\mathbb{P}(S_{\min} = S_i) = \frac{\lambda_i}{\Lambda}.$$

30 – \times Udowodnij Lemat 3 z [notatek do wykładu](#).

31 – \times Rzucamy monetą do czasu wyrzucenia r reszek. Przyjmijmy, że reszka i orzeł wypadają z prawdopodobieństwem odpowiednio p i q . Wyprowadź rozkład zmiennej losowej X opisującej liczbę wyrzuconych orłów. Jak nazywa się ten rozkład? Wyprowadź wzór na wartość oczekiwaną i wariancję.

32 – \times Przypomnij definicje i podstawowe własności rozkładu Poissona. Pokaż, że rozkład Poissona z parametrem μ jest rozkładem granicznym dla rozkładu dwumianowego $\text{Bin}(n, p_n)$ jeśli $\lim_{n \rightarrow \infty} np_n = \mu > 0$. Możesz zajrzeć do [tej książki](#).

33 – \times (Problem kolekcjonera kuponów) Mamy n urn do których losowo (jednostajnie) wrzucamy kule. Niech X będzie liczbą kul wrzuconych do chwili, kiedy w każdej urnie jest co najmniej jedna kula. Wykorzystując aproksymację liczby kul w danej urnie przez rozkład Poissona pokaż, że dla dużych wartości n mamy

$$\Pr[X > n \ln n + cn] \approx 1 - e^{-e^{-c}},$$

a następnie wyznacz najmniejszą wartość c taką, by dla dużych wartości n wartość X znajdowała się w przedziale $[n \ln n - cn, n \ln n + cn]$ w 99% przypadków.

Wskazówka: możesz zajrzeć do [tej książki](#).

Samostabilizacja

34 – Przedstaw samostabilizujący algorytm wyznaczania maksymalnego zbioru niezależnego opracowany w ramach zadania 11 na laboratorium. Dowiedz jego poprawności i zbieżności. Wyprowadź możliwie ściśle organicznie górne na liczbę kroków do chwili osiągnięcia legalnej konfiguracji.

35 – (Konserwatywny MM) Zaproponuj modyfikację algorytmu Maximal Matching przedstawionego w notatkach do wykładu, która umożliwi jego działanie przy dodatkowym założeniu, że każdy proces jest typu A albo B i dwa procesy tego samego typu nie mogą tworzyć pary. Ponadto możesz przyjąć, że typ danego procesu jest z góry zadany i nigdy nie może zostać zmieniony. Uzasadnij poprawność i zbieżność algorytmu.

36 – (Liberalny MM) Zaproponuj modyfikację algorytmu Maximal Matching przedstawionego w notatkach do wykładu, która umożliwi działanie przy dodatkowym założeniu, że liczba krawędzi incydentnych z danym wierzchołkiem może być większa niż jeden (tzw. b-matching). Uzasadnij poprawność algorytmu. Możesz się oprzeć np. na [tej pracy](#).

Łańcuchy Markowa

37 – Rozważmy łańcuch Markowa o stanach $\{1, 2, 3\}$ opisany macierzą przejścia

$$P = \begin{pmatrix} 1/2 & 1/3 & 1/6 \\ 3/4 & 0 & 1/4 \\ 0 & 1 & 0 \end{pmatrix}.$$

- Czy łańcuch jest nieredukowalny? Czy jest nieokresowy? Uzasadnij.
- Łańcuch startuje w stanie 1. Wyznacz prawdopodobieństwo, że po dwóch krokach jest w stanie 3.
- Wyznacz rozkład stacjonarny łańcucha oraz macierz $M = \lim_{n \rightarrow \infty} P^n$.

38 – Rozważmy łańcuch Markowa o stanach $\{1, 2, 3\}$ opisany macierzą przejścia

$$P = \begin{pmatrix} 1 & 0 & 0 \\ 1/4 & 1/2 & 1/4 \\ 0 & 0 & 1 \end{pmatrix}.$$

Sprawdź ile rozkładów stacjonarnych istnieje dla tego łańcucha oraz wyznacz macierz $K = \lim_{n \rightarrow \infty} P^n$. Jakiej własności nie ma macierz K , którą miała macierz M z poprzedniego zadania?

39 – Zaproponuj model dla prostej kolejki z ograniczonym rozmiarem: jeśli pojawi się więcej niż ustalona liczba n klientów, to kolejni klienci są odrzucani i rozmiar kolejki się nie zwiększa. Wyznacz rozkład stacjonarny dla łańcucha Markowa związanego z tym modelem.

40 – N ponumerowanych molekuła gazu przemieszcza się losowo między dwoma zamkniętymi pojemnikami przez mały otwór. Ruch molekuł można zamodelować następująco: w każdym kroku losujemy jednostajnie jedną z molekuł i przenosimy do innego pojemnika, niż ten w którym aktualnie się znajduje. Rozważmy łańcuch Markowa, o zbiorze stanów $S = \{0, 1, \dots, N\}$ reprezentującym liczbę molekuł w pierwszym pojemniku.

- Dla powyższego łańcucha wyznacz prawdopodobieństwa przejścia oraz rozkład stacjonarny.
- Niech $N = 100$. Uruchamiamy proces z pierwszym pojemnikiem pustym. Jeśli zmiana stanu następuje raz na jedną nanosekundę, średnio jak długo trzeba poczekać zanim pojemnik znów będzie pusty?

Powodzenia,
J.L.