Analiza Algorytmów 2019/2020
(zadania na ćwiczenia)

Wybór lidera

1 — Wyznacz wartość oczekiwaną dla zmiennej losowej $X \sim Geo(p)$. (3p)

2 — Wyznacz wariancję dla zmiennej losowej $X \sim Geo(p)$. (3p)

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? (9p)

a. $f(p) = np(1-p)^{n-1}$,

b. $f(n) = np(1-p)^{n-1}$,

c. $f(k) = \binom{n}{k} p^k (1-p)^{n-k}$.

4 — Udowodnij: $(1+x)^r \geq 1 + rx$ dla $x \geq -1, r \geq 1$. (3p)

5 — Udowodnij: $(1+x)^r \leq 1 + rx$ dla $x \geq -1, r \in (0,1)$. (3p)

6 — Udowodnij: $1 + x \leq e^x$ dla $x \in \mathbb{R}$. (3p)

7 — Udowodnij: $\frac{x}{e^x} < \frac{1.5}{x^2}$ dla $x > 0$. (3p)

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 $i \geq 2$. (12p)

9 — Przedstaw definicję rodziny funkcji W Lamberta i wykres jej gałęzi rzeczywistych. (5p)

10 — Wykorzystaj funkcję W Lamberta by analitycznie wyznaczyć rzeczywiste rozwiązania równania $3^x = x^3$. Jak nazwa się w Mathematica funkcja W Lamberta? (5p)

11 — Zajrzyj do tej pracy i udowodnij, że dla $x \geq e$ (10p)

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

12 — Uzupełnienie dowodu z wykładu. Sprawdź, że jeśli $K \geq 1, f > 1, u \geq 2$ oraz (5p)

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

Analiza strumieni danych: przybliżone zliczanie

13 — Dla ciągłych i niezależnych zmiennych losowych $X_1, X_2, \ldots, X_n$ 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. (10p)

14 — Dla n niezależnych zmiennych losowych $U_1, U_2, \ldots, U_n$ 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 $k/(n+1)$. (5p)

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

15 — 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 $\mathbb{E}\left(\hat{n}_k\right) = n$ oraz, że dla $k \geq 3$ mamy (10p)

$$\mathbb{Var}\left(\hat{n}_k\right) = \frac{n(n-k+1)}{k-2} \,.$$

16 — (Nierówność Markowa) Niech X oznacza zmienną losową, która przyjmuje tylko nieujemne wartości. Wtedy dla wszystkich $a > 0$ (5p)

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

17 — (Nierówność Czebyszewa) Niech X oznacza zmienną losową o skończonej wartości oczekiwanej i skończonej, niezerowej wariancji. Pokaż, że dla dowolnego $a > 0$ zachodzi nierówność: (5p)

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

Wskazówka: wykorzystaj nierówność Markowa.

18 — (Nierówność Chernoffa dla sumy prób Bernoulliego) Niech $X_1, \ldots, X_n$ będą niezależnymi próbami Bernoulliego takimi, że $\mathbb{P}\left(X_i = 1\right) = p_i$. Niech $X = \sum_{i=1}^{n} X_i$ oraz $\mu = \mathbb{E}\left(X\right)$. Pokaż, że (10p)

a) dla dowolnego $\delta > 0$

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

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

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

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 $0 < \delta < 1$ (5p)

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

20 — Niech S_n będzie liczbą orłów w n rzutach symetryczną monetą. Pokaż, że (5p)

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
Initialization: C ← 1
Upon event:
if random( ) <= 2^(-C) then        ▷ random zwraca liczbę losową z zakresu [0, 1)
    C ← C + 1
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 $\mathbb{E}\left(2^{C_n}\right) = n + 2$ oraz $\mathbb{Var}\left(2^{C_n}\right) = \frac{1}{2}n(n+1)$. W oparciu o C_n zdefiniuj nieobciążony estymator wartości n i policz jego wariancję. (10p)

Analiza strumieni danych: przybliżone sumowanie

22 — Przypomnij sobie podstawowe informacje o rozkładzie wykładniczym. (9p)

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 przekształca 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 $S_1, S_2, \ldots, S_n$ będzie ciągiem niezależnych zmiennych losowych o rozkładzie wykładniczym oraz $S_i \sim Exp(\lambda_i)$ dla $\lambda_i > 0$. Niech $\Lambda = \lambda_1 + \lambda_2 + \ldots + \lambda_n$. Pokaż, że pierwsza statystyka pozycyjna $S_{\min} = \min\{S_1, S_2, \ldots, S_n\}$ ma rozkład wykładniczy z parametrem Λ: (5p)

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

24 — 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. (6p)

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

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

Pokaż, że zmienna

$$G_m = \sum_{i=1}^{m} S^{(i)}_{\text{min}}$$

ma rozkład gamma zdefiniowany gęstością

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

Wskazówka 1: wykorzystaj poprzednie zadanie i zasadę indukcji.
Wskazówka 2: $\Gamma(m) = (m-1)!$ dla $m \in \mathbb{N}_{>0}$. (10p)

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

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

mamy $\mathbb{E}\left(\bar{\Lambda}_m\right) = \Lambda$. (6p)

27 — 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: (6p)

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

28 — Zaproponuj algorytm, który umożliwi estymacje średniej wartości

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

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

Wskazówka: zauważ, że dla $\lambda_1 = \lambda_2 = \ldots \lambda_n = 1$ mamy $\mathbb{E}\left(\bar{\Lambda}_m\right) = n$. (10p)

Blockchain

29 — Pokaż, że dla oznaczeń z zadania **23** mamy:

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

Wyjaśnij jaki związek ma ta własność z analizą ataku na blockchain. (6p)

30 — Przeczytaj Definicje 4 w notatkach do wykładu. Pokaż, że rozkład wykładniczy

a) spełnia warunki tej definicji,

b) jest jedynym rozkładem spełniającym warunki tej definicji.

Wskazówka: zdefiniuj $G(x) = \mathbb{P}\left(X > x\right)$ i pokaż, że z definicji własności braku pamięci musi zachodzić warunek $G(x) = G(1)^x$. Następnie pokaż, że jedyną ciągłą funkcją, która spełnia ten warunek dla dowolnego $x > 0$ jest $G(x) = e^{-\lambda x}$ dla $\lambda > 0$. (10p)

31 — Przypomnij definicje i podstawowe własności rozkładu Poissona. Podaj przykłady jego zastosowań. Pokaż, że rozkład Poissona z parametrem μ jest rozkładem granicznym dla rozkładu dwumianowego $Bin(n, p_n)$ jeśli $\lim_{n\to\infty} np_n = \mu > 0$. (6p)

32 — 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. (4p)

33 — Udowodnij Lemat 3 z notatek do wykładu. (10p)
Wskazówka: rozważ problem ruiny gracza w sytuacji gdy gracz ma nieograniczone środki.

34 — Udowodnij Lemat 4 z notatek do wykładu. (10p)
Wskazówka: zajrzyj do tej pracy.

35 — (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. (10p)
Wskazówka: możesz zajrzeć do tej książki.

Samostabilizacja

36 — Udowodnij Lemat 2 z notatek do wykładu. (10p)

37 — Udowodnij Lemat 3 z notatek do wykładu. (10p)

38 — Udowodnij Lemat 4 z notatek do wykładu. (10p)

39 — Przedstaw samostabilizujący algorytm wyznaczania maksymalnego zbioru niezależnego opracowany w ramach zadania 13 na laboratorium. Dowiedź jego poprawności i zbieżności. Wyprowadź możliwie ścisłe organicznie górne na liczbę kroków do chwili osiągnięcia legalnej konfiguracji. (10p)

40 — (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. (10p)

41 — (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. (10p)

Funkcje tworzące

42 — Udowodnij Lemat 1 z notatek do wykładu. (10p)

43 — Wykorzystując funkcje tworzące rozwiąż następujące równania rekurencyjne. (12p)

a) $a_n = 2a_{n-1} + 3a_{n-2}$ oraz $a_0 = 0, a_1 = 1$,

b) $a_n = 5a_{n-1} - 8a_{n-2} + 4a_{n-3}$ oraz $a_0 = 1, a_1 = 2, a_2 = 4$,

c) $a_n = 3a_{n-1} - 3a_{n-2} + a_{n-3}$ oraz $a_0 = 0, a_1 = 0, a_2 = 1$,

d) $a_n = 2a_{n-1} + a_{n-2} - 2a_{n-3}$ oraz $a_0 = 0, a_1 = a_2 = 1$.

44 — Wyznacz funkcję tworzące dla następujących ciągów. (21p)

a) $a_n = \begin{cases} 1 \text{ dla } 2 \mid n, \\ 0 \text{ dla } 2 \nmid n, \end{cases}$

b) $a_n = \binom{n}{k}$,

c) $a_n = \begin{cases} \binom{n+3}{3} 3^n \text{ dla } 2 \mid n, \\ 0 \text{ dla } 2 \nmid n, \end{cases}$

d) $a_n = n2^{n+1}$,

e) $a_n = 1/n$ dla $n \geq 1$,

f) $a_n = nH_n$,

g) $a_n = n^3$ dla $n \geq 2$.

45 — Pokaż, że (8p)

$$[z^n](1 - \alpha z)^{-j} = \binom{n+j-1}{j-1} \alpha^n .$$

46 — Pokaż, że jeśli $A(z) = \sum_{n \geq 0} a_n z^n$ ma promień zbieżności $R > 0$, to (6p)

$$(\forall r \in (0, R)) \left(a_n = \mathcal{O}\left(r^{-n} \right) \right) .$$

47 — Pokaż, że funkcja tworząca dla ciągu liczb Catalana spełnia zależność (6p)

$$C(z) = 1 + zC(z)^2$$

oraz, że wobec tego

$$C(z) = \frac{1 - \sqrt{1-4z}}{2z} .$$

48 — Wykorzystując wynik poprzedniego ćwiczenia udowodnij następujące asymptotyczne przybliżenie dla liczb Catalana (10p)

$$c_n \sim \frac{4^n}{n^{3/2}\sqrt{\pi}} .$$

49 — Wykorzystując uogólnienie symbolu Newtona na liczby rzeczywiste pokaż, że $n$ta liczbę Catalana wyraża się następującym wzorem (10p)

$$c_n = \frac{1}{n+1}\binom{2n}{n} .$$

Powodzenia,
J.L.