Algorytmika
Lista zadań

Jacek Cichoń
WIT, PWr, 2023/24 (semestr letni)

1 Podstawowy model obliczeń

Zadanie 1

Pokaż, że następujący problem:

> Mamy dwa programy P i Q obliczające funkcje z liczb naturalnych w liczby naturalne. Czy dla każdego naturalnego n P(n)=Q(n)?

jest nierozstrzygalny.
Wskazówka: Zredukuj problem $(\forall n)(P(n) = 0)$ do powyższego problemu.

Zadanie 2

Zbiór $A \subseteq \mathbb{N}^k$ nazywamy rekurencyjnie przeliczalnym (RE) jeśli istnieje całkowita funkcja obliczalna $f : \mathbb{N} \to \mathbb{N}^k$ taka, że $A = \{f(n) : n \in \mathbb{N}\}$.

1. Pokaż, że jeśli $A \subseteq \mathbb{N}$ jest taki, że $A \in RE$ oraz $\mathbb{N} \setminus A \in RE$ to A jest zbiorem rekurencyjnym.
2. Pokaż, że jeśli A jest rekurencyjny, to A jest rekurencyjnie przeliczalny.
3. Uzasadnij tezę $STOP \in RE$.

Zadanie 3

Dla ciągów $x, y \in \Sigma^*$ określamy $x \sqsubseteq y \longleftrightarrow (\exists z \in \Sigma^*)(y = xz)$ oraz $x \sqsupseteq y \longleftrightarrow (\exists z \in \Sigma^*)(y = zx)$.

1. Pokaż, że $\sqsubseteq$ jest częściowym porządkiem na Σ^*.
2. Wyraź relację $\sqsupseteq$ za pomocą relacji $\sqsubseteq$ oraz funkcji `reverse` odwracania ciągów.

Zadanie 4

Ustalmy skończony alfabet Σ oraz wzorzec $P[1..m] \in \Sigma^*$. Niech p_n oznacza prawdopodobieństwo tego, że losowy ciąg długości n elementów Σ jest zgodny (w jakimś miejscu) ze wzorcem P. Pokaż, że $\lim_{n\to\infty} p_n = 1$.

Zadanie 5

Ustalmy skończony alfabet Σ. Ustalmy ciąg $A[1..k]$. Niech p_n oznacza prawdopodobieństwo tego, że losowy ciąg X długości n elementów Σ^* zawiera podciąg A, czyli, że istnieją $1 \leqslant j_1 < j_2 < \ldots < j_k \leqslant n$ takie, że $X[i_j] = A[j]$ dla wszystkich $j = 1, \ldots, k$.

1. Wyznacz dokładny wzór na p_n.
2. Wyznacz asymptotykę ciągu $(p_n)_n$

Zadanie 6

Załóżmy, że wzorzec P składa się z różnych znaków. Pokaż, że problem dopasowania wzorca P do ciągu długości n można rozwiązać w czasie $\mathrm{O}(n)$ (niezależnym od długości wzorca).

Ćwiczenie 1

Napisz w języku C program, który nie ma argumentów wejściowych i generuje swoją własną kopię.

Ćwiczenie 2

Napisz w języku `C` funkcje które sprawdzają, czy dany łańcuch jest prefixem bądź postfixem drugiego. Jakie funkcje języka `Python` służą do tego celu?

Ćwiczenie 3

Napisz w języku `C` oraz `Python` funkcję która dla danych dwóch łańcuchów `x[1:n]` i `y[1:m]` zwraca największą liczbę k taką, że `x[1:k] = y[m-k+1:m]`.

Ćwiczenie 4

Zaimplementuj możliwie optymalnie (skorzystaj z funkcji `memcmp`) naiwny algorytm zgodności wzorca z tekstem w języku `C`.

Ćwiczenie 5

Mamy dany wzorzec P.

1. Napisz procedurę, która wyznacza prefixowy automat skończony służący do wykrywania obecności wzorca P w dowolnym łańcuchu.
2. Wykorzystaj powyższą procedurę do napisania procedury służącej do wykrywania obecności wzorca P w dowolnym łańcuchu.

Ćwiczenie 6

Zaimplementuj metodę Hornera wyznaczania wartości wielomianu i dokładnie zbadaj jej złożoność obliczeniową.

Ćwiczenie 7

Zaimplementuj algorytm Rabina-Karpa wykrywania obecności wzorca w tekście.

Zadanie 7

Ustalmy skończony alfabet Σ. Niech $lcs(x,y)$ oznacza długość najdłuższego wspólnego podciągu ciągów $x, y \in \Sigma^*$.

1. Niech $scs(x,y)$ oznacza długość najkrótszego ciągu którego pociągami są $x, y \in \Sigma^*$. Pokaż, że
$$scs(x,y) = |x| + |y| - lcs(x,y)\ .$$
2. Niech $d(x,y)$ oznacza odległość edycyjną między ciągami $x, y \in \Sigma^*$, dla dopuszczalnych operacji dodawania oraz usuwania pojedyńczego znaku w ciągu (obie operację mają wagę 1). Pokaż, że
$$d(x,y) = |x| + |y| - lcs(x,y)\ .$$

Ćwiczenie 8

Jaka jest wartość oczekiwana $\mathrm{E}\left[lcs(x,y)\right]$, gdzie x, y są losowymi $\{0,1\}$ - ciągami długości 5 (rozważamy rozkład jednostajny)?

Zadanie 8

Załóżmy, że $(a_n)_{n\geqslant 0}$ jest ciągiem dodatnich liczb rzeczywistych takich, że
$$(\forall n, m \in \mathbb{N})(a_{n+m} \leqslant a_n \cdot a_m)\ .$$
Pokaż, że istnieje granica $\lim_n (a_n)^{\frac{1}{n}}$.

Zadanie 9

Ścieżką w siatce $\mathbb{Z}^2$ długości n nazywamy skończony ciąg punktów $(P_k)_{k=0}^{n}$ zbioru $\mathbb{Z}^2$ taki, że $P_{i+1} \in \{P_i + (\pm 1, 0), P_i + (0, \pm 1)\}$ dla każdego $i = 0, \ldots, n-1$. Niech L_n oznacza liczbę nieprzecinających się ścieżek w $\mathbb{Z} \times \mathbb{Z}$ zaczynających się w punkcie $(0,0)$ długości n.

1. Pokaż, że istnieje granica $\gamma_2 = \lim_n (L_n)^{\frac{1}{n}}$.
2. Znajdź jakieś rozsądne ograniczenia na liczbę γ

Zadanie 10

Oto twierdzenie Ramseya:

Dla dowolnej funkcji $F : [\mathbb{N}]^2 \to \{0, 1\}$ istnieje nieskończony zbiór $A \subseteq \mathbb{N}$ taki, że $|F[[A]^2]| = 1$.

Załóżmy, że częściowy porządek $(X, \preceq)$ ma następującą własność: dla dowolnego ciągu nieskończonego ciągu $(a_n)_n$ elementów X istnieją $i < j$ takie, że $i < j$ oraz $a_i \preceq a_j$. Pokaż, że wtedy dla dowolnego ciągu nieskończonego ciągu $(a_n)_n$ elementów X istnieje nieskończony podciąg $(a_{n_i})_i$ taki, że $a_{n_i} \preceq a_{n_j}$ jeśli $i < j$. Skorzystać możesz z twierdzenia Ramseya.

Ćwiczenie 9

Napisz algorytm, który dla danego ciągu $x \in \{A, C, G, T\}^*$ znajduje wszystkie podłańcuchy o następującej postaci

$$y = "ATG" uF \ ,$$

takie, że $F \in \{"TAA", "TAG", "TGA"\}$, $|u| \geqslant 30$ oraz żaden z łańuchów ze zbioru

$$\{"ATG", "TAA", "TAG", "TGA"\}$$

nie jest podłańcuchem ciągu u. Algorytm ma działać w czasie $\mathrm{O}(|x|)$.

Zadanie 11

Załóżmy, że $\preceq$ jest dobrym porządkiem zbioru Σ. Ustalmy liczbę $n \in \mathbb{N}$. Pokaż, że porządek leksykograficzny $\preceq_{lex}$ obcięty do zbioru

$$\Sigma^{\leqslant n} = \{\sigma \in \Sigma^* : |\sigma| \leqslant n\}$$

jest dobrym porządkiem. Wskazówka: Skorzystaj z twierdzenia Ramseya.

Ćwiczenie 10

Zaimplementuj algorytm wyznaczania najdłuższego wspólnego podciągu dla trzech ciągów.

2 Model podstawowy + kostka losowa

Zadanie 12

Mamy fałszywą monetę, która zwraca orła z prawdopodobieństwem p oraz reszkę z prawdopodobieństwem $q = 1 - p$. Wiemy tylko, że $0 < p < 1$. Jak możesz użyć tej monety do wygerowania uczciwej monety (tzn. takiej aby prawdopodobieństwo orła i reszki było równe $\frac{1}{2}$?
Wskazówka: Użyj (co najmniej) dwóch rzutów fałszywą monetą.

Zadanie 13

Załóżmy, że mamy metodę generowania liczb losowych ze zbioru $\{0, 1, 2, 3, 4\}$ zgodną z jednostajnym rozkładem. Użyj tej metody do zbudowania generatora losowych liczb ze zbioru $\{0, 1, 2, 3, 4, 5, 6\}$ zgodnego z rozkładem jednostajnym.
Wskazówka: Wygeneruj najpierw losową liczbę ze zbioru $\{, 1, \ldots, 24\}$ i zastosuj metodę odrzucania.

Zadanie 14

Rozważamy metodę Monte-Carlo do obliczenia pola koła $\{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leqslant 1\}$.

1. Jaka jest wariancja tej metody?
2. Skorzystaj z nierówności Czebyszewa do oszacowania liczby iteracji do osiągnięcia dokładność 0.01 z prawdopodobieństwem 0.99?
3. Skorzystaj z następującego wariantu nierówności Chernoffa

Jeśli $X_1, \ldots, X_n$ są niezależne takimi, że $0 \leqslant X_i \leqslant 1$ dla każdego $i = 1, \ldots, n$, $\mu = E(X_1 + \ldots + X_n)$ oraz $\varepsilon > 0$, to

$$Pr(|X - \mu| > \varepsilon\mu) \leqslant 2\exp\left(-\frac{\varepsilon^2}{2+\varepsilon}\mu\right)$$

do lepszego oszacowania liczby n potrzebnych prób.

Ćwiczenie 11

Zastosuj metodę Monte-Carlo do wyznaczenia aproksymacji $\int_0^\pi \sin(x)dx$.

Ćwiczenie 12

Zaimplementuj zrandomizowany min-cut Kargera. Jeżeli będziesz go implementował w języku Python, to do modelowania grafów możesz posłużyć się słownikami, np.

```
graph = {
  'a' : ['b','c']
  'b' : ['a','d'],
  ...
}
```

Przetestuj jego działanie na kilku prostych grafach.

Zadanie 15

Mamy fałszywą monetę, która zwraca orła z prawdopodobieństwem p oraz reszkę z prawdopodobieństwem $q = 1 - p$. Wiemy tylko, że $0 < p < 1$. Jak możesz użyć tej monety do wygerowania uczciwej monety (tzn. takiej aby prawdopodobieństwo orła i reszki było równe $\frac{1}{2}$?
Wskazówka: Użyj dwóch rzutów fałszywą monetą.

Zadanie 16

Niech k będzie ciałem. Niech $f \in k[x_1, \ldots, x_n]$ będzie wielomianem zmiennych $x_1, \ldots, x_n$. Załóżmy, że każda zmienna występuje w tym wielomianie w potędze $\leqslant d$.

1. Niech $A \subseteq k$ ma moc $|A|$ większą od d. Pokaż, że jeśli $f \neq 0$, to

   $$|\{a \in A^n : f(a) = 0\}| \leqslant d \cdot |A|^{n-1} .$$

   Wskazówka: Zastosuj indukcję matematyczną. Potraktuj $f \in k[x_1, \ldots, x_{n+1}]$ jako element pierścienia $k(x_1, \ldots, x_n)[x_{n+1}]$, gdzie $k(x_1, \ldots, x_n)$ oznacza ciało funkcji wymiernych zmiennych $x_1, \ldots, x_n$.

2. Załóżmy, że $|A| = 2 \cdot d$ oraz $f \neq 0$. Niech $\zeta = (a_1, \ldots, a_n)$ będzie losowym punktem zbioru A^n (losowanym zgodnie z rozkładem jednostajnym). Pokaż, że $\Pr[f(\zeta) = 0] \leqslant \frac{1}{2}$.
3. Załóżmy, że $|A| = 2 \cdot d$. Załóżmy, że $\zeta_1, \ldots, \zeta_{40}$ jest ciągiem niezależnych losowych punktów ze zbioru A^n. Pokaż, że jeśli $f \neq 0$ to $\Pr[f(\zeta_1) = f(\zeta_2) = \ldots = f(\zeta_{40}) = 0] \leqslant 10^{-12}$.

Zadanie 17

Załóżmy, że X jest zmienną losową o rozkładzie jednostajny w odcinku $[0, 1]$. Niech $Y = \max(X, 1-X)$. Wyznacz $\mathrm{E}[Y]$ oraz $\mathrm{var}[Y]$.

Zadanie 18

Załóżmy, że X jest zmienną losową o rozkładzie jednostajny w odcinku $[0, 1]$. Niech $Y = 1 - X$. Oblicz $COV(X, Y)$.

Zadanie 19(Problem Buffona)

Być może tego nie robiliście: na płaszczyźnie rysujemy równoległe linie $L = \{(k, x) : k \in \mathbb{Z} \wedge x \in \mathbb{R}\}$ i na płaszczyznę rzucamy odcinek T długości $l \leqslant 1$; jakie jest prawdopodobieństwo tego, że $P \cap L \neq \emptyset$?

Ćwiczenie 13

Zaimplementuj metodę "stratified sampling" dla przybliżania wartości całki $\int_a^b f dx$ i przetestuj jej skuteczność dla wyznaczania $\int_0^1 \sqrt{1-x^2}dx$. Pamiętaj o zastosowaniu możliwie dobrego generatora liczb pseudo-losowych. Zastosuj następnie metodę przeciwstawnych elementów (antithetic variaty method) do wyznaczenia całki $\int_0^1 \sqrt{1-x^2}dx$.

Ćwiczenie 14

Napisz pseudo-kod algorytmu który otrzymujemy w wyniku derandomizacji wykorzystującej wartość oczekiwaną zrandomizowanego algorytmu służącego do wyznaczenia cięcia w grafie (V, E) rozmiaru $\geqslant E/2$.

Zadanie 20

Rozważmy następującą losową wersję algorytmu sortowania przez scalanie:

1. Bierzemy nieposortowaną listę składającą się z n elementów i dzielimy ją na dwie listy, rzucając niezależną monetą dla każdego elementu, aby zdecydować, na której liście ją umieścić.
2. Następnie sortujemy obie listy i łączymy powstałe posortowane listy.
3. Procedura scalania polega na wielokrotnym porównywaniu najmniejszego elementu na każdej z dwóch list i usuwaniu znalezionego mniejszego elementu, aż jedna z list będzie pusta.

Oblicz oczekiwaną liczbę porównań potrzebnych do przeprowadzenia ostatecznego scalania. Nie musisz brać pod uwagę kosztu sortowania.

Ćwiczenie 15

Rozważmy następujący BARDZO KIEPSKI algorytm sużący wygenerowania ciągu długości n składającego się z samych zer:

```
for i = 1 to n {
  X[i] = randomBit
}
return X
```

Prawdopodobieństwo sukcesu tego algorytmu wynosi $\frac{1}{2^n}$, co jest raczej kiepskim wynikiem. Przeprowadź derandomizację tego algorytmu metodą prawdopodobieństw warunkowych, czyli rozważ ciąg funkcji

$$c(x_1, \ldots, x_k) = \Pr[Sukces|X_1 = x_1, \ldots, X_k = x_k] \ .$$

Ćwiczenie 16

Power of Two Choices. Rozważamy dwa procesy losowej alokacji $m = 10^6$ kul do $n = 10^3$ kubełków. Oprogramuj te procesy, przeprowadź symulacje i wygeneruj histogramy alokacji kul w kubełkach.

1. W pierwszym procesie każdą kulę umieszczamy w losowo (rozważamy rozkład jednostajny), niezależnie wybranych kubełkach.
2. W drugim procesie stosujemy metodę Dwóch Wyborów.

Ćwiczenie 17

Rozważamy następujący algorytm przybliżonego zliczania liczby stacji w środowisku bezprzewodowym:

1. każda stacja $v \in V$ generuje niezależnie liczbą losową $\zeta_v \in [0, 1]$;
2. zbieramy (nie ważne jak) listę $L = \{\zeta_v : v \in V\}$;
3. sortujemy listę L i otrzymujemy ciąg $x_1 \leqslant x_2 \leqslant \ldots \leqslant x_n$;
4. jeśli $n \leqslant 400$ to zwracamy n

5. jeśli $n > 400$ to zwracamy liczbę $399/x_{400}$

Zbadaj eksperymentalnie dokładność tego algorytmu dla liczby stacji $n \leqslant 10000$

Zadanie 21

Statystyki porządkowe. Ustalmy liczby naturaln $1 \leqslant k \leqslant n$. Rozważamy następującą zmienną losową $X_{n:k}$: generujemy niezależnie n liczb losowych z odcinka $[0,1]$; porządkujemy otrzymaną listę; zwracamy k-ty element tej listy.

1. Pokaż, że zmienna losowa $X_{n:k}$ ma gęstość

$$f_{n:k}(x) = \binom{n}{1}\binom{n-1}{k-1}x^{k-1}(1-x)^{n-k}$$

   Wskazówka: Ustal $x \in (0,1)$ oraz małą liczbę δ. Oszacuj $Pr[X_{n:k} \in [x, x+\delta)]$ i przejdź z δ do 0.
2. Wyznacz $\mathrm{E}[X_{n:k}]$.
3. Niech $Z_{n,k} = (k-1)/X_{n:k}$. Pokaż, że jeśli $k \geqslant 2$, to $\mathrm{E}[Z_{n:k}] = n$.
4. Wyznacz wariancję zmiennej losowej $X_{n:k}$ dla $k \geqslant 3$.
5. Co możesz powiedzieć o zmiennych losowych $Z_{n:1}$ i $Z_{n:2}$?

3 Streaming

Ćwiczenie 18

Oprogramuj procedurę generowania ciągu chwil w których algorytm R Vittera aktualizuje wartość licznika. Przetestuj ją generując ciągi długości 50.

c.d.n.
Powodzenia,
Jacek Cichoń