Hledání maxima v permutaci (Hofri)

Úloha: Najít maximální člen v posloupnosti $A_1, .., A_n$ navzájem různých přirozených čísel (pro jednoduchost předp., že je to permutace čísel $1, .., n$)

Vstup: n .. kladné celé číslo

$A_1, \dots, A_n$ navzájem různá čísla

Výstup: $\max\limits_{1 \le i \le n} A_i$

Popis činnosti: posloupnost se prohlíží jednou zleva doprava, největší dosud nalezené číslo si pamatujeme jako maximum

Algoritmus:

	počet opakování instrukcí:
1. read n	1
2. $k := 1$	1
3. $m := A_k$	$M_n + 1$
4. $k := k+1$	n
5. if $k > n$	n
6. then goto 10	1
7. if $m > A_k$	n
8. then goto 4	$n - M_n$
9. else goto 3	M_n náhodná veličina
10. output m	1

Doba výpočtu: $T_n = \alpha + \beta n + \gamma M_n$

α, β, γ .. konstanty

M_n = počet maxim (lokálních) napočítaných při prohlížení posloupnosti zleva doprava, první člen napočítáme jako maximum (jiná terminologie: M_n je počet "rekordů")

zřejmě $M_n = 0$ když $A_1 = n$

$M_n = n-1$, když $A_k = k \quad \forall k = 1, \dots, n$

(ve statistice se veličina M_n používá jako základ pro test náhodnosti)

<u>rozdělení M_n</u>

předpokládáme rovnoměrné rozdělení vstupů, tj. $n!$ možných permutací čísel $1, \dots, n$

označíme

$$p_{ni} = P(M_n = i)$$

rozdělíme všechny permutace na 2 disjunktní skupiny:

1) permutace, kde $A_n = n$ (tj. maximum je na konci)

těch je $(n-1)!$

$P(A_n = n) = \frac{(n-1)!}{n!} = \frac{1}{n}$

v tomto případě $M_n = M_{n-1} + 1$

tedy $P(M_n = i \mid A_n = n) = P(M_{n-1} + 1 = i) = P(M_{n-1} = i-1) = p_{n-1, i-1}$

2) permutace, kde $A_n \neq n$

těch je $n! - (n-1)! = (n-1)(n-1)!$

$P(A_n \neq n) = \frac{(n-1)(n-1)!}{n!} = \frac{n-1}{n}$

v tomto případě $M_n = M_{n-1}$

takže $P(M_n = i \mid A_n \neq n) = P(M_{n-1} = i) = p_{n-1, i}$

platí: $(A_n = n) \cap (A_n \neq n) = \emptyset$

$(A_n = n) \cup (A_n \neq n) = I$ (jistý jev)

tedy: $(M_n = i) = (M_n = i \cap A_n = n) \cup (M_n = i \cap A_n \neq n)$

$$p_{ni} = P(M_n = i) = P(M_n = i \cap A_n = n) + P(M_n = i \cap A_n \neq n) =$$

$$= P(M_n = i \,/\, A_n = n)P(A_n = n) + P(M_n = i \,/\, A_n \neq n)P(A_n \neq n) =$$

(věta o celkové pravděpodobnosti)

$$= p_{n-1,i-1} \cdot \frac{1}{n} + p_{n-1,i} \frac{n-1}{n}$$

dostáváme rekurentní vztah:

$$p_{ni} = \frac{1}{n} p_{n-1,i-1} + \frac{n-1}{n} p_{n-1,i} \qquad \text{pro } n \geq 2,\ i \geq 1$$

s počátečními podmínkami:

$$p_{1,i} = \begin{cases} 1 & \text{pro } i = 0 \\ 0 & \text{pro } i \neq 0 \end{cases}$$

(tj. $n = 1 \Leftrightarrow$ max. je na začátku $\Leftrightarrow$ $\Leftrightarrow M_n = 0$)

$$p_{n,0} = \frac{1}{n}$$

(tj. $M_n = 0 \Leftrightarrow$ max. je na začátku $\Leftrightarrow$ $\Leftrightarrow P(A_1 = n) = \frac{1}{n}$)

přímo z těchto rovnic pravděpodobnosti počítat nebudeme

použijeme vytvořující funkci

$$G_n(x) = \sum_{i \geq 0} p_{ni} x^i \ .$$

$$p_{ni} = \frac{1}{n} p_{n-1,i-1} + \frac{n-1}{n} p_{n-1,i} \qquad / \cdot x^i, \ \sum_{i \geq 1}$$

$$\sum_{i \geq 1} p_{ni} x^i = \frac{x}{n} \sum_{i \geq 1} p_{n-1,i-1} x^{i-1} + \frac{n-1}{n} \sum_{i \geq 1} p_{n-1,i} x^i$$

$$\sum_{i\ge 0} p_{ni}\, x^i - \frac{1}{n} = \frac{x}{n} \sum_{i\ge 0} p_{n-1,i}\, x^i + \frac{n-1}{n}\left(\sum_{i\ge 0} p_{n-1,i}\, x^i - \frac{1}{n-1}\right)$$

$$G_n(x) - \frac{1}{n} = \frac{x}{n} G_{n-1}(x) + \frac{n-1}{n}\left(G_{n-1}(x) - \frac{1}{n-1}\right)$$

$$G_n(x) = \frac{x+n-1}{n} G_{n-1}(x)$$

$$G_1(x) = p_{10} + \cancel{p_{11} x} = 1$$

tedy

$$G_n(x) = \frac{x+n-1}{n} \cdot \frac{x+n-2}{n-1} \cdot \ldots \cdot \frac{x+1}{2} G_1(x) =$$

$$= \frac{1}{x} \cdot \frac{x+n-1}{n} \cdot \frac{x+n-2}{n-1} \cdot \ldots \cdot \frac{x+1}{2} \cdot \frac{x}{1} = \frac{1}{x}\binom{x+n-1}{n}$$

$$G_n(x) = \frac{1}{x}\binom{x+n-1}{n} = \frac{1}{x}(-1)^n \binom{-x}{n} = \frac{(-1)^n}{x\, n!}\, n! \binom{-x}{n}$$

připomeneme Stirling. čísla 1. druhu: $n!\binom{x}{n} = \sum_{k=0}^{n} (-1)^{n-k} \genfrac{[}{]}{0pt}{}{n}{k} x^k$

$$G_n(x) = \frac{(-1)^n}{x\, n!} \sum_{i=0}^{n} (-1)^{n-i} \genfrac{[}{]}{0pt}{}{n}{i} (-1)^i x^i = \frac{1}{n!} \sum_{i=0}^{n} \genfrac{[}{]}{0pt}{}{n}{i} x^{i-1}$$

platí: $\genfrac{[}{]}{0pt}{}{n}{0} = 0$, $\genfrac{[}{]}{0pt}{}{n}{m} = 0$ pro $m > n$

$$G_n(x) = \frac{1}{n!} \sum_{i=1}^{n} \genfrac{[}{]}{0pt}{}{n}{i} x^{i-1} = \frac{1}{n!} \sum_{i=0}^{n-1} \genfrac{[}{]}{0pt}{}{n}{i+1} x^i = \frac{1}{n!} \sum_{i\ge 0} \genfrac{[}{]}{0pt}{}{n}{i+1} x^i$$

$\underline{G_n(x) = \sum_{i\ge 0} \frac{\genfrac{[}{]}{0pt}{}{n}{i+1}}{n!} x^i}$ a tedy $\underline{p_{ni} = \frac{\genfrac{[}{]}{0pt}{}{n}{i+1}}{n!}}$

odtud dále plyne: $\genfrac{[}{]}{0pt}{}{n}{i+1}$ = počet permutací, pro které počet lokálních maxim počítaje zleva doprava je roven i

Vypočteme střední hodnotu M_n:

$$EM_n = \sum_{i=0}^{n-1} i\, p_{ni} = \sum_{i=0}^{n-1} i \frac{\left[{n \atop i+1}\right]}{n!}$$

budeme potřebovat následující identity: $\left[{n\atop 0}\right]=0$, $\left[{n\atop 1}\right]=(n-1)!$

$$\sum_{m=0}^{n} \left[{n\atop m}\right] = n!, \quad \sum_{k=0}^{n} \left[{n\atop k}\right]\binom{k}{m} = \left[{n+1\atop m+1}\right] = \sum_{k=0}^{n} \left[{k\atop m}\right]\frac{n!}{k!}$$

$$EM_n = \frac{1}{n!}\sum_{i=0}^{n-1} i\left[{n\atop i+1}\right] = \frac{1}{n!}\sum_{i=1}^{n}(i-1)\left[{n\atop i}\right] = \frac{1}{n!}\sum_{i=0}^{n} i\left[{n\atop i}\right] - \frac{1}{n!}\underbrace{\sum_{i=0}^{n}\left[{n\atop i}\right]}_{n!} =$$

$$= \frac{1}{n!}\sum_{i=0}^{n}\left[{n\atop i}\right]\binom{i}{1} - 1 = \frac{1}{n!}\left[{n+1\atop 2}\right] - 1 = \frac{1}{n!}\left[{n+1\atop 1+1}\right] - 1 =$$

$$= \frac{1}{n!}\sum_{k=0}^{n} \left[{k\atop 1}\right]\frac{n!}{k!} - 1 = \sum_{k=1}^{n} \frac{(k-1)!}{k!} - 1 = \sum_{k=1}^{n} \frac{1}{k} - 1 = H_n - 1$$

pro velká n: $H_n \sim \log n$

Střední hodnota $T_n = \alpha + \beta n + \gamma M_n$:

$$ET_n = \alpha + \beta n + \gamma EM_n = \alpha + \beta n + \gamma (H_n - 1) \sim$$

$$\sim \alpha + \beta n + \gamma(\log n - 1)$$

$$ET_n = O(n + \log n)$$

Střední hodnota M_n – jiný způsob:

budeme počítat $EM_n = G_n'(1)$ z rekurentního vztahu

$$G_n(x) = \frac{x+n-1}{n}\, G_{n-1}(x) , \quad kde \quad G_1(x) = 1$$

$$G_n'(x) = \frac{1}{n}\, G_{n-1}(x) + \frac{x+n-1}{n}\, G_{n-1}'(x)$$

z toho pro $x = 1$:

$$G_n'(1) = \frac{1}{n} + G_{n-1}'(1) \quad \text{protože} \quad G_{n-1}(1) = \sum p_{n-1,i}\, 1^i = 1$$

$$G_n'(1) = \frac{1}{n} + \frac{1}{n-1} + G_{n-2}'(1) = \frac{1}{n} + \frac{1}{n-1} + \frac{1}{n-2} + \dots + \frac{1}{2} + \underbrace{G_1'(1)}_{0} = \sum_{k=2}^{n} \frac{1}{k} = \sum_{k=1}^{n} \frac{1}{k} - 1 = H_n - 1$$

<u>rozptyl:</u>

$T_n = \alpha + \beta n + \gamma M_n$

$\operatorname{var} T_n = \gamma^2 \operatorname{var} M_n$

vyjdeme ze vztahu $G_n(x) = \underbrace{\frac{x+n-1}{n}}_{Q_n(x)} G_{n-1}(x)$

zjistíme, že $Q_n(x)$ je vytvořující fce náh. veličiny X_n s alternativním

rozdělením : $P(X_n = 0) = \frac{n-1}{n}$

$P(X_n = 1) = \frac{1}{n}$

$Q_n(x) = \frac{n-1}{n} x^0 + \frac{1}{n} x^1 = \frac{x+n-1}{n}$

platí $G_n(x) = Q_n(x) G_{n-1}(x)$,

tj. $M_n = X_n + M_{n-1}$, kde X_n, M_{n-1} jsou nezávislé náh. vel.

(veličiny X_n vlastně odpovídají jevu $A_n = n$, resp. $A_n \neq n$)

potom $\operatorname{var} M_n = \operatorname{var} X_n + \operatorname{var} M_{n-1}$ (samozřejmě také $EM_n = EX_n + EM_{n-1}$ a máme další způsob výpočtu EM_n)

$\operatorname{var} X_n = EX_n^2 - (EX_n)^2 = \frac{1}{n} - \frac{1}{n^2} = \frac{n-1}{n^2}$

$\operatorname{var} M_n = \frac{1}{n} - \frac{1}{n^2} + \operatorname{var} M_{n-1} = \frac{1}{n} - \frac{1}{n^2} + \frac{1}{n-1} - \frac{1}{(n-1)^2} + \operatorname{var} M_{n-2} = \dots$

$\dots = H_n - 1 - (H_n^{(2)} - 1) = H_n - H_n^{(2)}$

kde $H_n^{(2)} = \sum_{k=1}^{n} \frac{1}{k^2} \longrightarrow$ konst $\left(= \frac{\pi^2}{6}\right)$ pro $n \to \infty$

tedy asymptoticky opět $\operatorname{var} M_n \sim \log n$

$\operatorname{var} T_n = \gamma^2 \operatorname{var} M_n = \gamma^2 (H_n - H_n^{(2)}) \sim \log n$

směrodatná odchylka : $\sqrt{\operatorname{var} T_n} = \gamma \sqrt{\operatorname{var} M_n} \sim \sqrt{\log n}$

Poznámka k vyšším harmonickým číslům (Knuth)

nechť $H_n^{(r)} = \sum\limits_{k=1}^{n} \frac{1}{k^r}$

pro $r \geq 2$ tyto řady konvergují ke svým nekonečným součtům

$$H_n^{(r)} = \sum_{k=1}^{n} \frac{1}{k^r} \longrightarrow \sum_{k=1}^{\infty} \frac{1}{k^r} = H_\infty^{(r)} < \infty$$

(zdola)

veličina $H_\infty^{(r)}$ je v matematice známa jako Riemannova ζ-funkce

a pro r sudé jsou její hodnoty známé, např. $H_\infty^{(2)} = \frac{\pi^2}{6}$,

$H_\infty^{(4)} = \frac{\pi^4}{90}$, $H_\infty^{(6)} = \frac{\pi^6}{945}$

Poznámka k rozptylu.

rozptyl je velký, což značí, že veličina M_n není dobře

lokalizovaná kolem své střední hodnoty

nerovnosti Čebyševova typu nemají valný smysl, leda

pro konkrétní hodnoty n

Poznámka k algoritmu:

analogickým způsobem se hledá minimální prvek v permutaci.

analýza algoritmu je stejná

použití – třídění extrakcí maxima

Ještě jeden způsob (podle Mehlhorna):

M_i = počet přepsání proměnné m (tj. průběžného maxima) při hledání maxima v permutaci délky i

označím si $Y_i = EM_i$, tj. v celé permutaci $Y_n = EM_n$

Platí $P(a(1) = k) = \frac{1}{n}$ pro vš. $k = 1, \dots, n$

je-li $a(1) = n$, pak $Y_n = 0$

je-li $a(1) = k < n$, pak přepsání proměnné m nastane pouze pro členy posloupnosti rovné $k+1, k+2, \dots, n$

členy s hodnotou $< k$ to neovlivní, tak já je z té posloupnosti rovnou vyloučím, a bude:

$$Y_n = \sum_{k=1}^{n-1} (1 + Y_{n-k}) \frac{1}{n}$$

Y_{n-k} = oč. počet přepsání m při hledání maxima ~~v posl.~~ mezi prvky $k+1, k+2, \dots, n$

$$= \sum_{k=1}^{n-1} (1 + Y_k) \frac{1}{n}$$

$1 + Y_{n-k}$ = oč. počet přepsání m při hledání maxima mezi prvky $k, k+1, k+2, \dots, n$

řešení rekurze:

$$n Y_n = n - 1 + \sum_{k=1}^{n-1} Y_k$$

$$(n+1) Y_{n+1} = n + \sum_{k=1}^{n} Y_k$$

odečtu: $(n+1) Y_{n+1} - n Y_n = 1 + Y_n$

$$Y_{n+1} = \frac{1}{n+1} + Y_n = \frac{1}{n+1} + \frac{1}{n} + \dots + Y_1$$

počáteční podmínka: $Y_1 = 0$, odtud $Y_{n+1} = H_{n+1} - 1$

Tímto trikem se ale nedá spočítat rozptyl!

Třídění extrakcí maxima

Třídíme permutaci $\sigma_n = (a_1, a_2, \dots, a_n)$

postup: najdeme maximum, dáme ho na konec této posloupnosti
(tj. prohodíme ho s posledním prvkem)
vezmeme permutaci σ_{n-1}, najdeme maximum, dáme ho na konec této posloupnosti – tím máme na posledních dvou místech hodnoty $n-1, n$

atd.

postupně pro $k = n, n-1, \dots, 1$ hledáme vždy maximum v permutaci σ_k – k tomu potřebujeme čas T_k

v nejhorším případě $T_k \sim k + \max M_k = k + k - 1$

v průměrném případě $ET_k \sim k + EM_k \sim k + H_k - 1 \sim k + \log k$

doba výpočtu celého algoritmu $T = \sum_{k=1}^{n} T_k$

v nejhorším případě $T \sim \sum_{k=1}^{n} k + \sum_{k=1}^{n} (k-1) = \frac{n(n+1)}{2} + \frac{(n-1)n}{2} = n^2$

v průměrném případě $ET \sim \sum_{k=1}^{n} k + \sum_{k=1}^{n} \log k \leq \frac{n(n+1)}{2} + n \log(n) =$

$$= \frac{n^2}{2} + n \log n + \frac{n}{2}$$

Závěr: v obou případech je doba výpočtu $O(n^2)$, ale v průměrném případě je u n^2 menší konstanta

Pozn: symbol $O(n^2)$ znamená, že ex. konstanty M, N tak, že pro $n \geq N$ platí $|T_n| \leq M n^2$.

Cykly v permutaci

Mějme permutaci σ_n danou tabulkou

$$\sigma = \begin{pmatrix} 1 & 2 & \dots & n \\ a_1 & a_2 & \dots & a_n \end{pmatrix}, \quad \text{kde } a_j = \sigma(j).$$

Permutace je vlastně dána zobrazením σ.

Příklad:

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 7 & 4 & 5 & 3 & 2 & 6 & 1 \end{pmatrix}$$

platí $\sigma(1) = 7$, $\sigma(2) = 4$, atd.

Vezmeme prvek a_1 a tvoříme posloupnost $a_k = \sigma(a_{k-1})$. Nechť po p_1 krocích se v těle posloupnosti znovu objeví prvek a_1. Pak posloupnost $a_1, a_2, \dots, a_{p_1-1}$ se nazývá <u>cyklem generovaným a_1</u>. Prvek, který není v tomto cyklu, podobným způsobem generuje jiný, disjunktní cyklus.

V našem příkladě jsou cykly: $7, 1 \mid 4, 3, 5, 2 \mid 6$.

Takováto reprezentace permutace ovšem není jednoznačná – např. $6 \mid 1\,7 \mid 3\,5\,2\,4$ reprezentuje tutéž permutaci.

<u>Kanonická reprezentace</u>: Uvnitř každého cyklu provedeme rotaci tak, aby se maximální člen dostal na začátek a cykly uspořádáme rostoucím způsobem (podle 1. členu).

Dostaneme: $5\,2\,4\,3 \mid 6 \mid 7\,1$.

Tato reprezentace už je jednoznačná.

Jaký je počet cyklů v permutaci (náhodně vybrané)?
Seřadíme-li prvky v kanonické reprezentaci za sebou, dostaneme permutaci 5 2 4 3 6 7 1. Začátek každého cyklu je tam, kde se nachází průběžné maximum (musíme za průběžné maximum brát i první člen). Počet cyklů je tedy náhodná veličina $M_n + 1$ ze úlohy o hledání maxima.

platí: $P(\sigma \text{ má } k \text{ cyklů}) = P(M_n + 1 = k) = \dfrac{\left[{n \atop k}\right]}{n!}$ ⊛

$$\text{průměrný počet cyklů} = E(M_n + 1) = H_n \sim \log n$$

$$\text{rozptyl} = \operatorname{var} M_n = H_n - H_n^{(2)} \sim \log n$$

$$\text{průměrná délka cyklu} = \frac{n}{H_n}$$

⊛ neboli $\left[{n \atop k}\right]$ je počet permutací, které mají k cyklů