SELEÇÃO EM TEMPO LINEAR ESPERADO

Análise de Algoritmos

Prof. André Vignatti

A seleção do i-ésimo menor elemento do vetor A pode ser feita em tempo esperado $\Theta(n)$.

- Usa Randomized-Partition do QuickSort aleatorizado

```
Select(A, p, r, i)
---
if p = r then
    return A[p]
q ← Randomized-Partition(A, p, r)
k ← q − p + 1
if i = k then
    return A[q]
else if i < k then
    return Select(A, p, q − 1, i)
else
    return Select(A, q + 1, r, i − k)
```

Tempo de Pior Caso

$\Theta(n^2)$ quando a recursão sempre diminui o vetor em apenas 1 elemento (temos que ser muito azarados).

Tempo Esperado

Randomized-Partition retorna com mesma probabilidade qualquer elemento como pivô. Assim

- $A[p..q]$ tem 1 elemento com probabilidade $1/n$
- $A[p..q]$ tem 2 elementos com probabilidade $1/n$
- $\vdots$
- $A[p..q]$ tem n elementos com probabilidade $1/n$

Ou seja,

$$\boxed{A[p..q] \text{ tem } k \text{ elementos com probabilidade } 1/n}$$

Seja X_k v.a. tal que

$$X_k = \begin{cases} 1, & \text{se } A[p..q] \text{ tem } k \text{ elementos} \\ 0, & \text{caso contrário} \end{cases}$$

Observe que $\mathrm{E}[X_k] = \Pr(X_k = 1) = 1/n$

Em uma chamada, note que $X_k = 1$ somente para um único k, e $X_k = 0$ para os outros. Assim,

$$\begin{aligned} T(n) = \quad & X_1 \cdot [\text{tempo de execução se } A[p..q] \text{ tem 1 elemento}] \\ & + X_2 \cdot [\text{tempo de execuçãose } A[p..q] \text{ tem 2 elementos}] \\ & \vdots \\ & + X_n \cdot [\text{tempo de execução se } A[p..q] \text{ tem } n \text{ elementos}] \end{aligned}$$

Ou,

$$T(n) = \sum_{k=1}^{n} X_k \cdot [\text{tempo de execução se } A[p..q] \text{ tem } k \text{ elementos}]$$

Não sabemos quantificar o "tempo de execução se $A[p..q]$ tem k elementos", pois pode terminar imediatamente com a resposta correta, pode entrar na recursão em $A[p, q-1]$, ou pode entrar na recursão em $A[q+1, r]$. Então, para facilitar, vamos obter um **limitante superior**.

Ideia:

- supor que $T(n)$ é monotonicamente crescente
- supor que o i-ésimo menor elemento sempre está no sub-vetor maior
- o tempo do Randomized-Partition é $\Theta(n)$

Desta forma,

$$
\begin{aligned}
T(n) &= \sum_{k=1}^{n} X_k \cdot [\text{tempo de execução se } A[p..q] \text{ tem } k \text{ elementos}] \\
&\leq \sum_{k=1}^{n} X_k \cdot \Big(T\big(\max(k-1, n-k)\big) + \Theta(n) \Big). \\
&= \Theta(n) + \sum_{k=1}^{n} X_k \cdot T\big(\max(k-1, n-k)\big).
\end{aligned}
$$

Calculando a esperança

$$\begin{aligned}
\mathrm{E}\big[T(n)\big] &\leq \mathrm{E}\left[\Theta(n) + \sum_{k=1}^{n} X_k \cdot T\big(\max(k-1, n-k)\big)\right] \\
&= \mathrm{E}\big[\Theta(n)\big] + \mathrm{E}\left[\sum_{k=1}^{n} X_k \cdot T\big(\max(k-1, n-k)\big)\right] && \text{(linearidade da esperança)} \\
&= \Theta(n) + \sum_{k=1}^{n} \mathrm{E}\Big[X_k \cdot T\big(\max(k-1, n-k)\big)\Big] && \text{(linearidade da esperança)} \\
&= \Theta(n) + \sum_{k=1}^{n} \mathrm{E}\big[X_k\big] \cdot \mathrm{E}\Big[T\big(\max(k-1, n-k)\big)\Big] && \text{(v.a.'s independentes)} \\
&= \Theta(n) + \sum_{k=1}^{n} \frac{1}{n} \cdot \mathrm{E}\Big[T\big(\max(k-1, n-k)\big)\Big] \\
&= \Theta(n) + \frac{1}{n} \sum_{k=1}^{n} \mathrm{E}\Big[T\big(\max(k-1, n-k)\big)\Big]
\end{aligned}$$

Observe que

$$\max(k-1, n-k) = \begin{cases} k-1, & \text{se } k > \lceil n/2 \rceil, \\ n-k, & \text{se } k \leq \lceil n/2 \rceil. \end{cases}$$

- Se n é par, cada termo de $T(\lceil n/2 \rceil)$ até $T(n-1)$ aparece exatamente duas vezes no somatório.

- Se n é ímpar, esses termos aparecem duas vezes e $T(\lfloor n/2 \rfloor)$ aparece uma vez.

De qualquer forma,

$$\mathrm{E}\big[T(n)\big] \leq \Theta(n) + \frac{2}{n} \sum_{k=\lfloor n/2 \rfloor}^{n-1} \mathrm{E}[T(k)]$$

$\boxed{\textbf{Teorema. } \mathrm{E}\big[T(n)\big] = O(n).}$

Demonstração. (por indução) Iremos assumir que

- $T(n) \leq cn$ para alguma constante c
- $T(n) = O(1)$ para $n <$ alguma constante que iremos definir depois
- o termo $\Theta(n)$ é limitado superiormente por an, onde a é constante

Assim,

$$\begin{aligned}
\mathrm{E}\big[T(n)\big] &\leq \Theta(n) + \frac{2}{n} \sum_{k=\lfloor n/2 \rfloor}^{n-1} \mathrm{E}[T(k)] \\
&\leq an + \frac{2}{n} \sum_{k=\lfloor n/2 \rfloor}^{n-1} ck \\
&= an + \frac{2c}{n} \left(\sum_{k=1}^{n-1} k - \sum_{k=1}^{\lfloor n/2 \rfloor - 1} k \right) \\
&= an + \frac{2c}{n} \left(\frac{n(n-1)}{2)} - \frac{\lfloor n/2 \rfloor (\lfloor n/2 \rfloor - 1)}{2} \right) \\
&\leq an + \frac{2c}{n} \left(\frac{n(n-1)}{2)} - \frac{(n/2-1)(n/2-2)}{2} \right)
\end{aligned}$$

$$
\begin{aligned}
&= an + \frac{2c}{n}\left(\frac{n^2 - n}{2)} - \frac{n^2/4 - 3n/2 + 2}{2}\right) \\
&= an + \frac{c}{n}\left(\frac{3n^2}{4} + \frac{n}{2} - 2\right) \\
&= an + c\left(\frac{3n}{4} + \frac{1}{2} - \frac{2}{n}\right) \\
&\leq an + \frac{3cn}{4} + \frac{c}{2} \\
&= cn - \left(\frac{cn}{4} - \frac{c}{2} - an\right).
\end{aligned}
$$

Para completar a prova, escolhemos c tal que

$$\begin{aligned}
cn/4 - c/2 - an &\geq 0 \\
cn/4 - an &\geq c/2 \\
n(c/4 - a) &\geq c/2 \\
n &\geq \frac{c/2}{c/4 - a} \\
n &\geq \frac{2c}{c - 4a}.
\end{aligned}$$

Assim, supondo que $T(n) = O(1)$ para $n < 2c/(c - 4a)$, temos que $\mathrm{E}[T(n)] = O(n)$. $\square$