Análise de Algoritmos

Slides de Paulo Feofiloff

[com erros do coelho e agora também da cris]

Quicksort

CLRS 7

Partição

Problema: Rearranjar um dado vetor $A[p\,..\,d]$ e devolver um índice q tal que $p \leq q \leq d$ e

$$A[p\,..\,q-1] \leq A[q] < A[q+1\,..\,d]$$

Entra:

$$\begin{array}{c|c|c|c|c|c|c|c|c|c|c|} & p & & & & & & & & & d \\ \cline{2-11} A & 99 & 33 & 55 & 77 & 11 & 22 & 88 & 66 & 33 & 44 \\ \cline{2-11} \end{array}$$

Partição

Problema: Rearranjar um dado vetor $A[p\,.\,.\,d]$ e devolver um índice q tal que $p \leq q \leq d$ e

$$A[p\,.\,.\,q-1] \leq A[q] < A[q+1\,.\,.\,d]$$

Entra:

p | | | | | | | | | d

A	99	33	55	77	11	22	88	66	33	44

Sai:

p | | | | q | | | | | d

A	33	11	22	33	44	55	99	66	77	88

Particione

	p									d
A	99	33	55	77	11	22	88	66	33	44

Particione

	i	j									x
A		99	33	55	77	11	22	88	66	33	44

Particione

	i	j								x
A	99	33	55	77	11	22	88	66	33	44

Particione

i		j								x
A	99	33	55	77	11	22	88	66	33	44

	i		j							x
A	33	99	55	77	11	22	88	66	33	44

Particione

i		j								x
A	99	33	55	77	11	22	88	66	33	44

	i		j							x
A	33	99	55	77	11	22	88	66	33	44

	i			j						x
A	33	99	55	77	11	22	88	66	33	44

Particione

i		j								x
A	99	33	55	77	11	22	88	66	33	44

	i		j							x
A	33	99	55	77	11	22	88	66	33	44

	i				j					x
A	33	99	55	77	11	22	88	66	33	44

Particione

i		j								x
A	99	33	55	77	11	22	88	66	33	44

	i		j							x
A	33	99	55	77	11	22	88	66	33	44

		i				j				x
A	33	11	55	77	99	22	88	66	33	44

Particione

i		j								x
A	99	33	55	77	11	22	88	66	33	44

	i		j							x
A	33	99	55	77	11	22	88	66	33	44

		i				j				x
A	33	11	55	77	99	22	88	66	33	44

			i				j			x
A	33	11	22	77	99	55	88	66	33	44

Particione

i		j								x
A	99	33	55	77	11	22	88	66	33	44
	i		j							x
A	33	99	55	77	11	22	88	66	33	44
		i				j				x
A	33	11	55	77	99	22	88	66	33	44
			i				j			x
A	33	11	22	77	99	55	88	66	33	44
			i					j		x
A	33	11	22	77	99	55	88	66	33	44

Particione

i		j								x
A	99	33	55	77	11	22	88	66	33	44

	i		j							x
A	33	99	55	77	11	22	88	66	33	44

		i				j				x
A	33	11	55	77	99	22	88	66	33	44

			i				j			x
A	33	11	22	77	99	55	88	66	33	44

			i						j	x
A	33	11	22	77	99	55	88	66	33	44

Particione

i		j								x
A	99	33	55	77	11	22	88	66	33	44

	i		j							x
A	33	99	55	77	11	22	88	66	33	44

		i				j				x
A	33	11	55	77	99	22	88	66	33	44

			i				j			x
A	33	11	22	77	99	55	88	66	33	44

				i						j
A	33	11	22	33	99	55	88	66	77	44

Particione

i		j								x
A	99	33	55	77	11	22	88	66	33	44

	i		j							x
A	33	99	55	77	11	22	88	66	33	44

		i				j				x
A	33	11	55	77	99	22	88	66	33	44

			i				j			x
A	33	11	22	77	99	55	88	66	33	44

				i						j
A	33	11	22	33	99	55	88	66	77	44

	p				q					d
A	33	11	22	33	44	55	88	66	77	99

Particione

Rearranja $A[p\,..\,d]$ de modo que $p \le q \le d$ e
$A[p\,..\,q{-}1] \le A[q] < A[q{+}1\,..\,d]$

PARTICIONE (A, p, d)
1 $\quad x \leftarrow A[d]$ $\qquad \triangleright$ x é o "pivô"
2 $\quad i \leftarrow p{-}1$
3 $\quad$ **para** $j \leftarrow p$ **até** $d-1$ **faça**
4 $\qquad$ **se** $A[j] \le x$
5 $\qquad\quad$ **então** $i \leftarrow i+1$
6 $\qquad\qquad A[i] \leftrightarrow A[j]$
7 $\quad A[i{+}1] \leftrightarrow A[d]$
8 $\quad$ **devolva** $i+1$

Invariantes: no começo de cada iteração de 3–6,

(i0) $A[p\,..\,i] \le x$ $\qquad$ (i1) $A[i{+}1\,..\,j{-}1] > x$ $\qquad$ (i2) $A[d] = x$

Consumo de tempo

Quanto tempo consome em função de $n := d - p + 1$?

linha	consumo de todas as execuções da linha
1-2	$= 2\,\Theta(1)$
3	$= \Theta(n)$
4	$= \Theta(n)$
5-6	$= 2\,\mathrm{O}(n)$
7-8	$= 2\,\Theta(1)$
total	$= \Theta(2n+4) + \mathrm{O}(2n) \quad = \Theta(n)$

Conclusão:

O algoritmo PARTICIONE consome tempo $\Theta(n)$.

Quicksort

Rearranja $A[p \,.\,.\, d]$ em ordem crescente.

QUICKSORT (A, p, d)
    se $p < d$
        então $q \leftarrow$ PARTICIONE (A, p, d)
            QUICKSORT $(A, p, q-1)$
            QUICKSORT $(A, q+1, d)$

	p									d
A	99	33	55	77	11	22	88	66	33	44

Quicksort

Rearranja $A[p\,..\,d]$ em ordem crescente.

QUICKSORT (A, p, d)
1 **se** $p < d$
2 **então** $q \leftarrow$ PARTICIONE (A, p, d)

---

3 QUICKSORT $(A, p, q-1)$
4 QUICKSORT $(A, q+1, d)$

	p				q					d
A	33	11	22	33	44	55	88	66	77	99

No começo da linha 3,

$$A[p\,..\,q{-}1] \le A[q] \le A[q{+}1\,..\,d]$$

Quicksort

Rearranja $A[p\,..\,d]$ em ordem crescente.

QUICKSORT (A, p, d)
se $p < d$
  então $q \leftarrow$ PARTICIONE (A, p, d)
    QUICKSORT $(A, p, q-1)$
    QUICKSORT $(A, q+1, d)$

	p				q					d
A	11	22	33	33	44	55	88	66	77	99

Quicksort

Rearranja $A[p \mathinner{..} d]$ em ordem crescente.

QUICKSORT (A, p, d)
1 **se** $p < d$
2 **então** $q \leftarrow$ PARTICIONE (A, p, d)
3 QUICKSORT $(A, p, q-1)$
4 QUICKSORT $(A, q+1, d)$

	p				q					d
A	11	22	33	33	44	55	66	77	88	99

Quicksort

Rearranja $A[p\,.\,.\,d]$ em ordem crescente.

```
QUICKSORT (A, p, d)
1    se p < d
2        então q ← PARTICIONE (A, p, d)
3              QUICKSORT (A, p, q − 1)
4              QUICKSORT (A, q + 1, d)
```

No começo da linha 3,

$$A[p\,.\,.\,q-1] \leq A[q] \leq A[q+1\,.\,.\,d]$$

Consumo de tempo?

Quicksort

Rearranja $A[p\,.\,.\,d]$ em ordem crescente.

QUICKSORT (A, p, d)
1 **se** $p < d$
2 **então** $q \leftarrow$ PARTICIONE (A, p, d)
3 QUICKSORT $(A, p, q-1)$
4 QUICKSORT $(A, q+1, d)$

No começo da linha 3,

$$A[p\,.\,.\,q{-}1] \leq A[q] \leq A[q{+}1\,.\,.\,d]$$

Consumo de tempo?

$T(n) :=$ consumo de tempo no pior caso sendo

$$n := d - p + 1$$

Consumo de tempo

Quanto tempo consome em função de $n := d - p + 1$?

linha		consumo de todas as execuções da linha
1	$=$	?
2	$=$	?
3	$=$	?
4	$=$	?
total	$=$	????

Consumo de tempo

Quanto tempo consome em função de $n := d - p + 1$?

linha		consumo de todas as execuções da linha
1	$=$	$\Theta(1)$
2	$=$	$\Theta(n)$
3	$=$	$T(k)$
4	$=$	$T(n-k-1)$
total	$=$	$T(k) + T(n-k-1) + \Theta(n+1)$

$$0 \le k := q - p \le n - 1$$

Recorrência

$T(n) :=$ consumo de tempo máximo quando $n = d - p + 1$

$$T(0) = \Theta(1)$$

$$T(1) = \Theta(1)$$

$$T(n) = T(k) + T(n-k-1) + \Theta(n) \quad \text{para } n = 2, 3, 4, \ldots$$

Recorrência

$T(n) :=$ consumo de tempo máximo quando $n = d - p + 1$

$$T(0) = \Theta(1)$$
$$T(1) = \Theta(1)$$
$$T(n) = T(k) + T(n-k-1) + \Theta(n) \quad \text{para } n = 2, 3, 4, \ldots$$

Recorrência grosseira:

$$T(n) = T(0) + T(n-1) + \Theta(n)$$

$T(n)$ é $\Theta(???)$.

Recorrência

$T(n) :=$ consumo de tempo máximo quando $n = d - p + 1$

$$T(0) = \Theta(1)$$

$$T(1) = \Theta(1)$$

$$T(n) = T(k) + T(n-k-1) + \Theta(n) \quad \text{para } n = 2, 3, 4, \ldots$$

Recorrência grosseira:

$$T(n) = T(0) + T(n-1) + \Theta(n)$$

$T(n)$ é $\Theta(n^2)$.

Demonstração: ... Exercício!

Recorrência cuidadosa

$T(n) :=$ consumo de tempo máximo quando $n = d - p + 1$

$T(0) = \Theta(1)$

$T(1) = \Theta(1)$

$T(n) = \max_{0 \le k \le n-1} \{T(k) + T(n-k-1)\} + \Theta(n)$ para $n = 2, 3, 4, \ldots$

Versão simplificada

$$T(0) = 1$$

$$T(1) = 1$$

$$T(n) = \max_{0 \le k \le n-1} \{T(k) + T(n-k-1)\} + n \quad \text{para } n = 2, 3, 4, \ldots$$

n	0	1	2	3	4	5
$T(n)$	1	1	$2+2$	$5+3$	$9+4$	$14+5$

Versão simplificada

$$T(0) = 1$$

$$T(1) = 1$$

$$T(n) = \max_{0 \le k \le n-1} \{T(k) + T(n-k-1)\} + n \quad \text{para } n = 2, 3, 4, \ldots$$

n	0	1	2	3	4	5
$T(n)$	1	1	$2+2$	$5+3$	$9+4$	$14+5$

Vou mostrar que $T(n) \le n^2 + 1$ para $n \ge 0$.

Demonstração

Prova: Trivial para $n \leq 1$. Se $n \geq 2$ então

$$\begin{aligned}
T(n) &= \max_{0 \leq k \leq n-1} \left\{ T(k) + T(n-k-1) \right\} + n \\
&\overset{\text{hi}}{\leq} \max_{0 \leq k \leq n-1} \left\{ k^2 + 1 + (n-k-1)^2 + 1 \right\} + n \\
&= \cdots \\
&= n^2 - n + 3 \\
&\leq n^2 + 1\,.
\end{aligned}$$

Prove que $T(n) \geq \frac{1}{2}\, n^2$ para $n \geq 1$.

Algumas conclusões

$T(n)$ é $\Theta(n^2)$.

O consumo de tempo do QUICKSORT no pior caso é $\mathrm{O}(n^2)$.

O consumo de tempo do QUICKSORT é $\mathrm{O}(n^2)$.

Quicksort no melhor caso

$M(n)$:= consumo de tempo mínimo quando $n = d - p + 1$

$$M(0) = \Theta(1)$$

$$M(1) = \Theta(1)$$

$$M(n) = \min_{0 \le k \le n-1} \{M(k) + M(n-k-1)\} + \Theta(n)$$ para $n = 2, 3, 4, \ldots$

Quicksort no melhor caso

$M(n)$:= consumo de tempo mínimo quando $n = d - p + 1$

$$M(0) = \Theta(1)$$

$$M(1) = \Theta(1)$$

$$M(n) = \min_{0 \le k \le n-1} \{M(k) + M(n-k-1)\} + \Theta(n) \quad \text{para } n = 2, 3, 4, \ldots$$

Mostre que, para $n \ge 1$,

$$M(n) \ge (n+1)\lg(n+1).$$

Isto implica que no melhor caso o QUICKSORT é $\Omega(n \lg n)$.

Que é o mesmo que dizer que o QUICKSORT é $\Omega(n \lg n)$.

Mais algumas conclusões

$M(n)$ é $\Theta(n \lg n)$.

O consumo de tempo do QUICKSORT no melhor caso é $\Omega(n \log n)$.

Na verdade ...

O consumo de tempo do QUICKSORT no melhor caso é $\Theta(n \log n)$.

Quicksort aleatorizado

CLRS 7.4

Exercício

Considere a recorrência

$$T(1) = 1$$

$$T(n) = T(\lceil n/3 \rceil) + T(\lfloor 2n/3 \rfloor) + 5n$$

para $n = 2, 3, 4, \ldots$

Solução assintótica: $T(n)$ é $\mathrm{O}(???)$, $T(n)$ é $\Theta(???)$

Exercício

Considere a recorrência

$$T(1) = 1$$

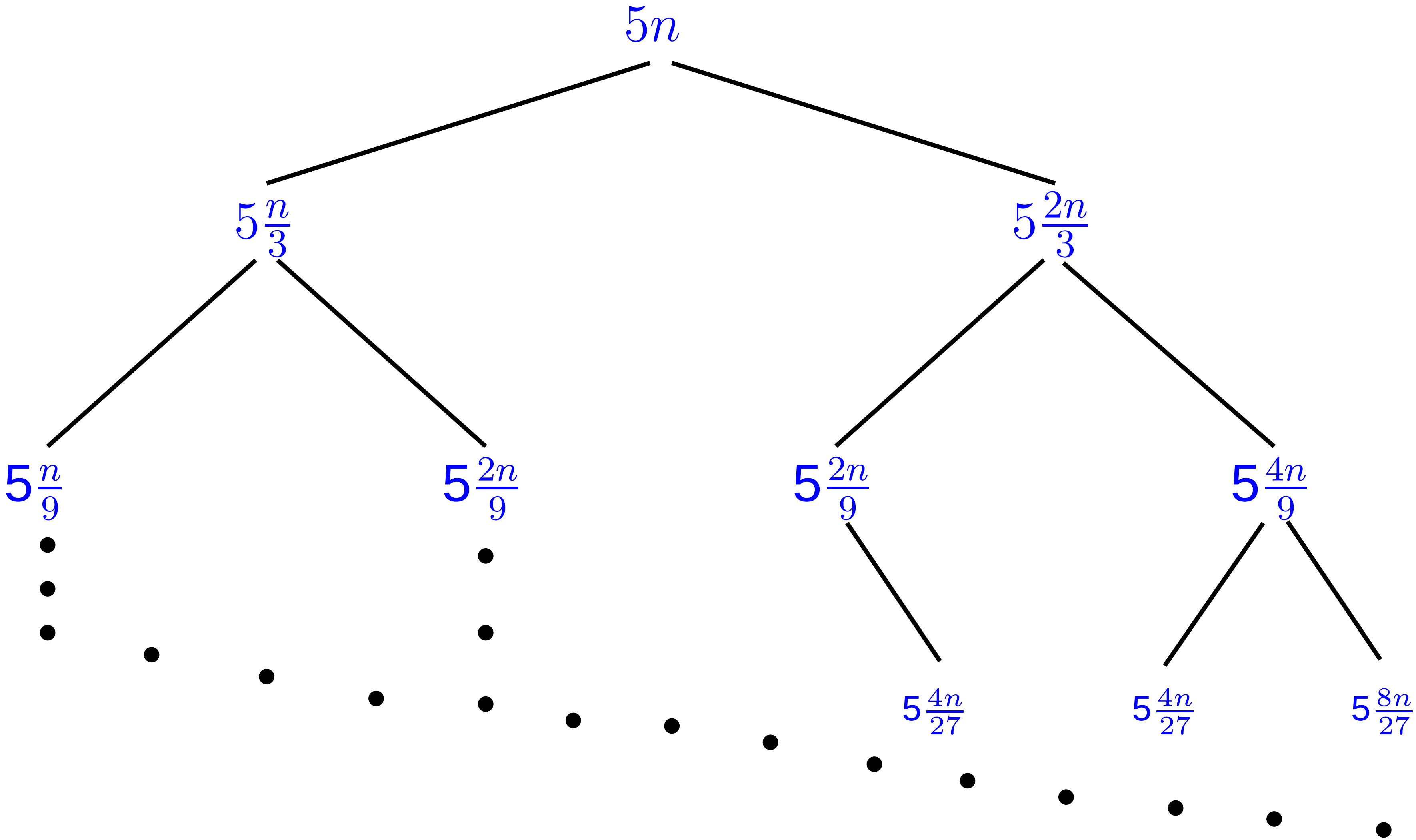
$$T(n) = T(\lceil n/3 \rceil) + T(\lfloor 2n/3 \rfloor) + 5n$$

para $n = 2, 3, 4, \ldots$

Solução assintótica: $T(n)$ é $\mathrm{O}(???)$, $T(n)$ é $\Theta(???)$

Vamos olhar a árvore da recorrência.

Árvore da recorrência



total de níveis $\leq \log_{3/2} n$

Árvore da recorrência

soma em cada horizontal $\leq 5n$

número de “níveis” $\leq \log_{3/2} n$

$T(n)$ = a soma de tudo

$$T(n) \;\leq\; 5n \log_{3/2} n \;+\; \underbrace{1 + \cdots + 1}_{\log_{3/2} n}$$

$T(n)$ é $\mathrm{O}(n \lg n)$.

De volta a recorrência

$$T(1) = 1$$

$$T(n) = T(\lceil n/3 \rceil) + T(\lfloor 2n/3 \rfloor) + 5n$$

para $n = 2, 3, 4, \ldots$

n	$T(n)$
1	1
2	$1 + 1 + 5 \cdot 2 = 12$
3	$1 + 12 + 5 \cdot 3 = 28$
4	$12 + 12 + 5 \cdot 4 = 44$

Vamos mostrar que $T(n) \leq 100\, n \lg n$ para $n = 2, 3, 4, 5, 6, \ldots$
Para $n = 2$ temos $T(2) = 12 < 100 \cdot 2 \cdot \lg 2$.
Para $n = 3$ temos $T(3) = 28 < 100 \cdot 3 \cdot \lg 3$.
Suponha agora que $n > 3$. Então

Continuação da prova

$$
\begin{aligned}
T(n) &= T(\lceil \tfrac{n}{3} \rceil) + T(\lfloor \tfrac{2n}{3} \rfloor) + 5n \\
&\overset{\text{hi}}{\leq} 100\lceil \tfrac{n}{3} \rceil \lg \lceil \tfrac{n}{3} \rceil + 100\lfloor \tfrac{2n}{3} \rfloor \lg \lfloor \tfrac{2n}{3} \rfloor + 5n \\
&\leq 100\frac{n+2}{3} \lceil \lg \frac{n}{3} \rceil + 100\frac{2n}{3} \lg \frac{2n}{3} + 5n \\
&< 100\frac{n+2}{3} (\lg \frac{n}{3} + 1) + 100\frac{2n}{3} \lg \frac{2n}{3} + 5n \\
&= 100\frac{n+2}{3} \lg \frac{2n}{3} + 100\frac{2n}{3} \lg \frac{2n}{3} + 5n \\
&= 100\frac{n}{3} \lg \frac{2n}{3} + 100\frac{2}{3} \lg \frac{2n}{3} + 100\frac{2n}{3} \lg \frac{2n}{3} + 5n
\end{aligned}
$$

Continuação da continuação da prova

$$
\begin{aligned}
&< 100n\lg\frac{2n}{3} + 67\lg\frac{2n}{3} + 5n \\
&= 100n\lg n + 100n\lg\frac{2}{3} + 67\lg n + 67\lg\frac{2}{3} + 5n \\
&< 100n\lg n + 100n(-0.58) + 67\lg n + 67(-0.58) + 5n \\
&< 100n\lg n - 58n + 67\lg n - 38 + 5n \\
&= 100n\lg n - 53n + 67\lg n - 38 \\
&< 100n\lg n
\end{aligned}
$$

Iiiéééééssss!

De volta ao caso médio

O consumo de tempo do QUICKSORT no caso médio é ???.

Partição $\frac{1}{10}$ para $\frac{9}{10}$:

$$R(n) = R(\lfloor \tfrac{n-1}{10} \rfloor) + R(\lceil \tfrac{9n-9}{10} \rceil) + \Theta(n)$$

De volta ao caso médio

O consumo de tempo do QUICKSORT no caso médio é $???$.

Partição $\frac{1}{10}$ para $\frac{9}{10}$:

$$R(n) = R(\lfloor \tfrac{n-1}{10} \rfloor) + R(\lceil \tfrac{9n-9}{10} \rceil) + \Theta(n)$$

Solução: $R(n)$ é $\Theta(n \lg n)$

De volta ao caso médio

O consumo de tempo do QUICKSORT no caso médio é $???$.

Partição $\frac{1}{10}$ para $\frac{9}{10}$:

$$R(n) = R(\lfloor \tfrac{n-1}{10} \rfloor) + R(\lceil \tfrac{9n-9}{10} \rceil) + \Theta(n)$$

Solução: $R(n)$ é $\Theta(n \lg n)$

Isso sugere que consumo médio é $\Theta(n \lg n)$.

Confirmação?

Particione

Rearranja $A[p\,..\,d]$ de modo que $p \le q \le d$ e
$A[p\,..\,q{-}1] \le A[q] < A[q{+}1\,..\,d]$

```
PARTICIONE (A, p, d)
1    x ← A[d]              ▷ x é o “pivô”
2    i ← p−1
3    para j ← p até d − 1 faça
4         se A[j] ≤ x
5              então i ← i + 1
6                    A[i] ↔ A[j]
7    A[i+1] ↔ A[d]
8    devolva i + 1
```

Invariantes: no começo de cada iteração de 3–6,

(i0) $A[p\,..\,i] \le x$ (i1) $A[i{+}1\,..\,j{-}1] > x$ (i2) $A[d] = x$

Exemplos

Número médio de execuções da linha 4 do PARTICIONE.

Suponha que $A[p \mathinner{..} r]$ é permutação de $1 \mathinner{..} n$.

$A[p \mathinner{..} r]$	execs
1,2	1
2,1	1
média	1

$A[p \mathinner{..} r]$	execs
1,2,3	2+1
2,1,3	2+1
1,3,2	2+0
3,1,2	2+0
2,3,1	2+1
3,2,1	2+1
média	16/6

Mais um exemplo

$A[p\,..\,r]$	execs
1,2,3,4	3+3
2,1,3,4	3+3
1,3,2,4	3+2
3,1,2,4	3+2
2,3,1,4	3+3
3,2,1,4	3+3
1,2,4,3	3+1
2,1,4,3	3+1
1,4,2,3	3+1
4,1,2,3	3+1
2,4,1,3	3+1
4,2,1,3	3+1

$A[p\,..\,r]$	execs
1,3,4,2	3+1
3,1,4,2	3+1
1,4,3,2	3+1
4,1,3,2	3+1
3,4,1,2	3+1
4,3,1,2	3+1
2,3,4,1	3+3
3,2,4,1	3+3
2,4,3,1	3+2
4,2,3,1	3+2
3,4,2,1	3+3
4,3,2,1	3+3
média	116/24

Quicksort aleatorizado

PARTICIONE-ALEA(A, p, d)
```
    i ← RANDOM (p, d)
    A[i] ↔ A[d]
    devolva PARTICIONE (A, p, d)
```

QUICKSORT-ALE (A, p, d)
```
    se p < d
        então q ← PARTICIONE-ALEA (A, p, d)
              QUICKSORT-ALE (A, p, q − 1)
              QUICKSORT-ALE (A, q + 1, d)
```

Análise do consumo medio?
Basta contar o número esperado de comparações na linha 4 do PARTICIONE

Exemplo

1	3	6	2	5	7	4

1	3	2	4	5	7	6

1	2	3	4	5	6	7

$$
\begin{array}{c|ccccccc|}
 & 1 & 2 & 3 & 4 & 5 & 6 & 7 \\
\hline
1 & & 1 & 0 & 1 & 0 & 0 & 0 \\
2 & 1 & & 1 & 1 & 0 & 0 & 0 \\
3 & 0 & 1 & & 1 & 0 & 0 & 0 \\
4 & 1 & 1 & 1 & & 1 & 1 & 1 \\
5 & 0 & 0 & 0 & 1 & & 1 & 0 \\
6 & 0 & 0 & 0 & 1 & 1 & & 1 \\
7 & 0 & 0 & 0 & 1 & 0 & 1 & \\
\hline
\end{array}
$$

Resumo de probabilidade

$(S, \Pr)$ espaço de probabilidade, $S =$ conjunto finito (eventos elementares)
$\Pr\{\} =$ (distribuição de probabilidades) função de S em $[0, 1]$ tal que

p1. $\Pr\{s\} \geq 0$;

p2. $\Pr\{S\} = 1$; e

p3. $A, B \subseteq S, A \cap B = \emptyset \Rightarrow \Pr\{A \cup B\} = \Pr\{A\} + \Pr\{B\}$.

Um evento é um subconjunto de S.

Uma variável aleatória é uma função númerica definida sobre os eventos elementares.

'$X = k$' é uma abreviação de $\{s \in S : X(s) = k\}$

$E[X] = \sum_{k \in X(S)} k \cdot \Pr\{X = k\} = \sum_{s \in S} X(s) \cdot \Pr\{s\}$

Caso médio

O consumo de tempo do QUICKSORT no caso médio é $???$.

Partição $\frac{1}{3}$ para $\frac{2}{3}$:

$$R(n) = R(\lfloor \tfrac{n-1}{3} \rfloor) + R(\lceil \tfrac{2n-2}{3} \rceil) + \Theta(n)$$

Solução: $R(n)$ é $\Theta(n \lg n)$. veja exercício a seguir

Consumo de tempo esperado

Suponha $A[p\,..\,r]$ permutação de $1\,..\,n$.

X_{ab} = número de comparações entre a e b
na linha 4 de PARTICIONE

Queremos calcular

$$
\begin{aligned}
X &= \text{total de comparações “}A[j] \le x\text{”} \\
&= \sum_{a=1}^{n-1} \sum_{b=a+1}^{n} X_{ab}
\end{aligned}
$$