AULA 10

Quicksort aleatorizado

CLRS 7.4

Relembrar . . .

Antes de mais nada, vamos relembrar rapidamente os melhores momentos da aula 9:

- PARTICIONE;
- consumo de tempo do PARTICIONE;
- QUICKSORT; e
- consumo de tempo do QUICKSORT.

Partição

Problema: Rearranjar um dado vetor $A[p \mathinner{.\,.} r]$ e devolver um índice q, $p \leq q \leq r$, tais que

$$A[p \mathinner{.\,.} q-1] \leq A[q] < A[q+1 \mathinner{.\,.} r]$$

Entra:

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

Sai:

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

Particione

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

Particione

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

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

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

Quicksort

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

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

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

Quicksort

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

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

---

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

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

No começo da linha 3,

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

Quicksort

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

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

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

Quicksort

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

```
QUICKSORT (A, p, r)
    se p < r
        então q ← PARTICIONE (A, p, r)
              QUICKSORT (A, p, q − 1)
              QUICKSORT (A, q + 1, r)
```

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

O consumo de tempo é proporcional ao número de execuções da linha 4 do PARTICIONE.

Resumo

O consumo de tempo do QUICKSORT é $O(n^2)$.

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

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

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

$$T(1) = 1$$

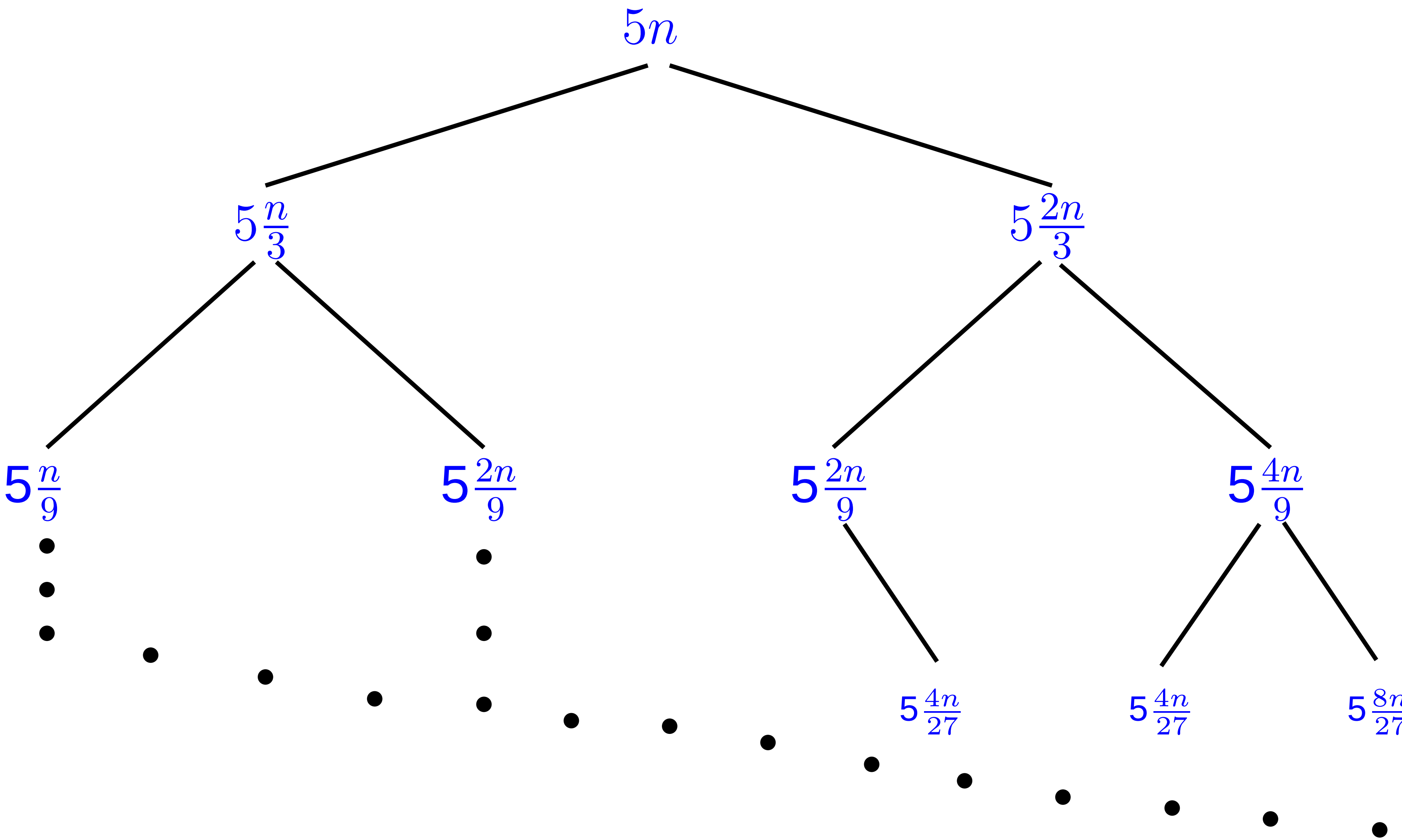
$$T(n) = T(\lceil n/3 \rceil) + T(\lfloor 2n/3 \rfloor) + 5n$$

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

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

Vamos olhar a árvore da recorrência.

Arvoré da recorrência



total de níveis $\leq \log_{3/2} n$

Arvoré 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

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

Continuação da continuação da prova

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

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?

Exemplos

Número médio de execuções da linha 4 do PARTICIONE.

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

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

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

Quicksort caso médio

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

```
QUICKSORT (A, p, r)
    se p < r
        então q ← PARTICIONE (A, p, r)
              QUICKSORT (A, p, q − 1)
              QUICKSORT (A, q + 1, r)
```

Análise do consumo médio?

Basta contar o número esperado de comparações na linha 4 do PARTICIONE

Suponha que $A[1\,.\,.\,n]$ é permutação aleatória uniforme de $1, \ldots, n$

Cada permutação tem probabilidade $1/n!$.

Consumo de tempo médio

O número esperado $C(n)$ de comparações na linha 4 do PARTICIONE satisfaz a recorrência:

$$C(0) = 0$$

$$C(n) = \frac{1}{n}\Big(\sum_{k=0}^{n-1} C(k) + C(n-k-1)\Big) + n - 1 \quad \text{para } n = 1, 2, 3, \ldots$$

Rescrevendo ...

$$C(0) = 0$$

$$C(n) = \frac{2}{n}\sum_{k=0}^{n-1} C(k) + n - 1 \quad \text{para } n = 1, 2, 3, \ldots$$

Consumo de tempo médio (cont.)

O número esperado $C(n)$ de comparações na linha 4 do PARTICIONE satisfaz a recorrência:

$$C(0) = 0$$

$$C(n) = \frac{2}{n}\sum_{k=0}^{n-1} C(k) + n - 1 \quad \text{para } n = 1, 2, 3, \ldots$$

Para nos livrarmos da divisão, multiplicamos ambos os lados por n

$$nC(n) = 2\sum_{k=0}^{n-1} C(k) + n^2 - n$$

Consumo de tempo médio (cont.)

Agora, para nos livramos do somatório, subtraímos a equação anterior de

$$(n-1)C(n-1) = 2\sum_{k=0}^{n-2} C(k) + (n-1)^2 - (n-1)$$

e obtemos

$$nC(n) - (n-1)C(n-1) = 2\,C(n-1) + 2n - 2.$$

Agora temos que resolver uma recorrência 'mais simples':

$$C(0) = 0$$

$$nC(n) = (n+1)C(n-1) + 2n - 2 \quad \text{para } n = 1, 2, 3, \ldots$$

Consumo de tempo médio (cont.)

Multiplicando, ambos os lados por $\frac{1}{n(n+1)}$ obtemos

$$\frac{1}{n+1}C(n) = \frac{1}{n}C(n-1) + \frac{2}{n+1} - \frac{2}{n(n+1)}\ .$$

Fazendo $S(n) := \frac{1}{n+1}C(n)$ chegamos a recorrência

$$S(0) = 0$$

$$S(n) = S(n-1) + \frac{2}{n+1} - \frac{2}{n(n+1)} \quad \text{para } n = 1, 2, 3, \ldots$$

que tem como solução

$$S(n) = 2\Big(H_{n+1} - 2 - \frac{1}{n+1}\Big)\ .$$

Consumo de tempo médio (cont.)

Concluímos que

$$C(n) = 2(n+1)H_{n+1} - 4(n+1) + 2 \quad \text{para } n = 0, 1, 2, \ldots$$

$$\ln(n+1) < H_{n+1} < 1 + \ln(n+1)$$

Verifiquemos o valor de $C(n)$ para valores pequenos

n	0	1	2	3	4
$C(n)$	0	1	1	$16/6$	$116/24$

Hmmmmm, parece que está certo.

Conclusões

$C(n)$ é $\Theta(n \log n)$.

O consumo de tempo esperado do algoritmo QUICKSORT é $\Theta(n \log n)$.

Quicksort aleatorizado

PARTICIONE-ALEA(A, p, r)
1 $\quad i \leftarrow$ RANDOM (p, r)
2 $\quad A[i] \leftrightarrow A[r]$
3 $\quad$ **devolva** PARTICIONE (A, p, r)

QUICKSORT-ALE (A, p, r)
1 $\quad$ **se** $p < r$
2 $\qquad$ **então** $q \leftarrow$ PARTICIONE-ALEA (A, p, r)
3 $\qquad\qquad$ QUICKSORT-ALE $(A, p, q-1)$
4 $\qquad\qquad$ QUICKSORT-ALE $(A, q+1, r)$

Análise do consumo médio?

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

	1	2	3	4	5	6	7
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	

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

$$X = \text{total de comparações "}A[j] \leq x\text{"}$$
$$= \sum_{a=1}^{n-1} \sum_{b=a+1}^{n} X_{ab}$$

Consumo de tempo esperado

Supondo $a < b$,

$$X_{ab} = \begin{cases} 1 & \text{se primeiro pivô em } \{a, \ldots, b\} \text{ é } a \text{ ou } b \\ 0 & \text{caso contrário} \end{cases}$$

Qual a probabilidade de X_{ab} valer 1?

Consumo de tempo esperado

Supondo $a < b$,

$$X_{ab} = \begin{cases} 1 & \text{se primeiro pivô em } \{a, \ldots, b\} \text{ é } a \text{ ou } b \\ 0 & \text{caso contrário} \end{cases}$$

Qual a probabilidade de X_{ab} valer 1?

$$\text{Pr}\,\{X_{ab}{=}1\} \;=\; \frac{1}{b-a+1} + \frac{1}{b-a+1} \;=\; \mathrm{E}[X_{ab}]$$

Consumo de tempo esperado

Supondo $a < b$,

$$X_{ab} = \begin{cases} 1 & \text{se primeiro pivô em } \{a, \ldots, b\} \text{ é } a \text{ ou } b \\ 0 & \text{caso contrário} \end{cases}$$

Qual a probabilidade de X_{ab} valer 1?

$$\Pr\{X_{ab}=1\} \;=\; \frac{1}{b-a+1} + \frac{1}{b-a+1} \;=\; \mathrm{E}[X_{ab}]$$

$$X = \sum_{a=1}^{n-1} \sum_{b=a+1}^{n} X_{ab}$$

$\mathrm{E}[X] =$????

Consumo de tempo esperado

$$
\begin{aligned}
\mathrm{E}[X] &= \sum_{a=1}^{n-1} \sum_{b=a+1}^{n} \mathrm{E}[X_{ab}] \\
&= \sum_{a=1}^{n-1} \sum_{b=a+1}^{n} \textsf{Pr}\,\{X_{ab}=1\} \\
&= \sum_{a=1}^{n-1} \sum_{b=a+1}^{n} \frac{2}{b-a+1} \\
&= \sum_{a=1}^{n-1} \sum_{k=1}^{n-a} \frac{2}{k+1} \\
&< \sum_{a=1}^{n-1} 2\left(\frac{1}{1}+\frac{1}{2}+\cdots+\frac{1}{n}\right) \\
&< 2n\left(\frac{1}{1}+\frac{1}{2}+\cdots+\frac{1}{n}\right) \; < \; 2n\,(1+\ln n)
\end{aligned}
$$

CLRS (A.7), p.1060

Conclusões

O consumo de tempo esperado do algoritmo QUICKSORT-ALE é $\mathrm{O}(n \lg n)$.

Do exercício 7.4-4 do CLRS temos que

O consumo de tempo esperado do algoritmo QUICKSORT-ALE é $\Theta(n \lg n)$.