# AULA 11

# Análise probabilística

CLRS 5.1, C.2, C.3

# Máximo

Problema: Encontrar o elemento máximo de um vetor $A[1\,..\,n]$ de números inteiros positivos distintos.

```
MAX (A, n)
1    max ← 0
2    para i ← 1 até n faça
3         se A[i] > max
4              então max ← A[i]
5    devolva max
```

Quantas vezes linha 4 é executada?

# Máximo

Problema: Encontrar o elemento máximo de um vetor $A[1\,..\,n]$ de números inteiros positivos distintos.

```
MAX (A, n)
1    max ← 0
2    para i ← 1 até n faça
3         se A[i] > max
4              então max ← A[i]
5    devolva max
```

Quantas vezes linha 4 é executada?
Melhor caso, pior caso, caso médio?

# Máximo

Problema: Encontrar o elemento máximo de um vetor $A[1\,..\,n]$ de números inteiros positivos distintos.

```
MAX (A, n)
1    max ← 0
2    para i ← 1 até n faça
3         se A[i] > max
4              então max ← A[i]
5    devolva max
```

Quantas vezes linha 4 é executada?
Melhor caso, pior caso, caso médio?

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

Cada permutação tem probabilidade $1/n!$.

# Um pouco 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. $R, T \subseteq S, R \cap T = \emptyset \Rightarrow \Pr\{R \cup T\} = \Pr\{R\} + \Pr\{T\}$.

$\Pr\{U\}$ é abreviação de $\displaystyle\sum_{u \in U} \Pr\{u\}$.

# Um pouco 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. $R, T \subseteq S, R \cap T = \emptyset \Rightarrow \Pr\{R \cup T\} = \Pr\{R\} + \Pr\{T\}$.

$$\Pr\{U\} \text{ é abreviação de } \sum_{u \in U} \Pr\{u\}.$$

No problema do máximo:

- $S$ é o conjunto das permutações dos números em $A[1\,..\,n]$;
- na distribuição uniforme, para cada $s \in S$, $\Pr\{s\} = 1/n!$.

# Mais um pouco de probabilidade

Um evento é um subconjunto de $S$.

# Mais um pouco de probabilidade

Um evento é um subconjunto de $S$.

No problema do máximo, eventos são subconjuntos de permutações de $A[1\,..\,n]$.

Exemplo.

$$U := \{\text{permutações de } A[1\,..\,n] \text{ em que } A[n] \text{ é máximo}\}$$

é um evento de $S$.

# Mais um pouco de probabilidade

Um evento é um subconjunto de $S$.

No problema do máximo, eventos são subconjuntos de permutações de $A[1\,..\,n]$.

Exemplo.

$$U := \{\text{permutações de } A[1\,..\,n] \text{ em que } A[n] \text{ é máximo}\}$$

é um evento de $S$.

Se $\Pr\{\}$ é distribuição uniforme, então

$$\Pr\{U\} = ???.$$

# Mais um pouco de probabilidade

Um evento é um subconjunto de $S$.

No problema do máximo, eventos são subconjuntos de permutações de $A[1\,..\,n]$.

Exemplo.

$$U := \{\text{permutações de } A[1\,..\,n] \text{ em que } A[n] \text{ é máximo}\}$$

é um evento de $S$.

Se $\Pr\{\}$ é distribuição uniforme, então

$$\Pr\{U\} = 1/n.$$

# Mais um pouco de probabilidade

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

# Mais um pouco de probabilidade

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

Exemplo de variável aleatória

$$X(A) := \text{número de execuções da linha 4 em MAX}(A, n)$$

# Mais um pouco de probabilidade

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

Exemplo de variável aleatória

$$X(A) := \text{número de execuções da linha 4 em MAX}(A, n)$$

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

Esperança $\mathrm{E}[X]$ de uma variável aleatória $X$

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

# Mais um pouco de probabilidade

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

Exemplo de variável aleatória

$$X(A) := \text{número de execuções da linha 4 em MAX}(A, n)$$

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

Esperança $\mathrm{E}[X]$ de uma variável aleatória $X$

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

Linearidade da esperança: $\mathrm{E}[\alpha X + Y] = \alpha \mathrm{E}[X] + \mathrm{E}[Y]$

# De volta ao máximo

Problema: Encontrar o elemento máximo de um vetor $A[1\,..\,n]$ de números inteiros distintos.

```
MAX (A, n)
1    max ← 0
2    para i ← 1 até n faça
3         se A[i] > max
4              então max ← A[i]
5    devolva max
```

Quantas vezes linha 4 é executada no caso médio?

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

Cada permutação tem probabilidade $1/n!$.

# Exemplos

| $A[1\,..\,2]$ | linha 4 |
|---|---|
| 1,2 | 2 |
| 2,1 | 1 |
| $\mathrm{E}[X]$ | 3/2 |

| $A[1\,..\,3]$ | linha 4 |
|---|---|
| 1,2,3 | 3 |
| 1,3,2 | 2 |
| 2,1,3 | 2 |
| 2,3,1 | 2 |
| 3,1,2 | 1 |
| 3,2,1 | 1 |
| $\mathrm{E}[X]$ | 11/6 |

# Mais um exemplo

| $A[1\,..\,4]$ | linha 4 |
|---|---|
| 1,2,3,4 | 4 |
| 1,2,4,3 | 3 |
| 1,3,2,4 | 3 |
| 1,3,4,2 | 3 |
| 1,4,2,3 | 2 |
| 1,4,3,2 | 2 |
| 2,1,3,4 | 3 |
| 2,1,4,3 | 2 |
| 2,3,1,4 | 3 |
| 2,3,4,1 | 3 |
| 2,4,1,3 | 2 |
| 2,4,3,1 | 2 |

| $A[1\,..\,4]$ | linha 14 |
|---|---|
| 3,1,2,4 | 2 |
| 3,1,4,2 | 2 |
| 3,2,1,4 | 2 |
| 3,2,4,1 | 2 |
| 3,4,1,2 | 2 |
| 3,4,2,1 | 2 |
| 4,1,2,3 | 1 |
| 4,1,3,2 | 1 |
| 4,2,1,3 | 1 |
| 4,2,3,1 | 1 |
| 4,3,1,2 | 1 |
| 4,3,2,1 | 1 |

$\mathrm{E}[X]$ 50/24

# Variáveis aleatórias

$X$ = número total de execuções da linha 4

# Variáveis aleatórias

$$X = \text{número total de execuções da linha 4}$$

$$X_i = \begin{cases} 1 \text{ se “}max \leftarrow A[i]\text{” é executado} \\ 0 \text{ caso contrário} \end{cases}$$

$$\begin{aligned} X &= \text{número total de execuções da linha 4} \\ &= X_1 + \cdots + X_n \end{aligned}$$

# Variáveis aleatórias

$$X = \text{número total de execuções da linha 4}$$

$$X_i = \begin{cases} 1 \text{ se “}max \leftarrow A[i]\text{” é executado} \\ 0 \text{ caso contrário} \end{cases}$$

$$\begin{aligned} X &= \text{número total de execuções da linha 4} \\ &= X_1 + \cdots + X_n \end{aligned}$$

Esperanças:

$$\begin{aligned} \mathrm{E}[X_i] &= \text{probabilidade de que } A[i] \text{ seja} \\ &\quad\ \text{máximo em } A[1\,..\,i] \\ &= 1/i \end{aligned}$$

# Esperança

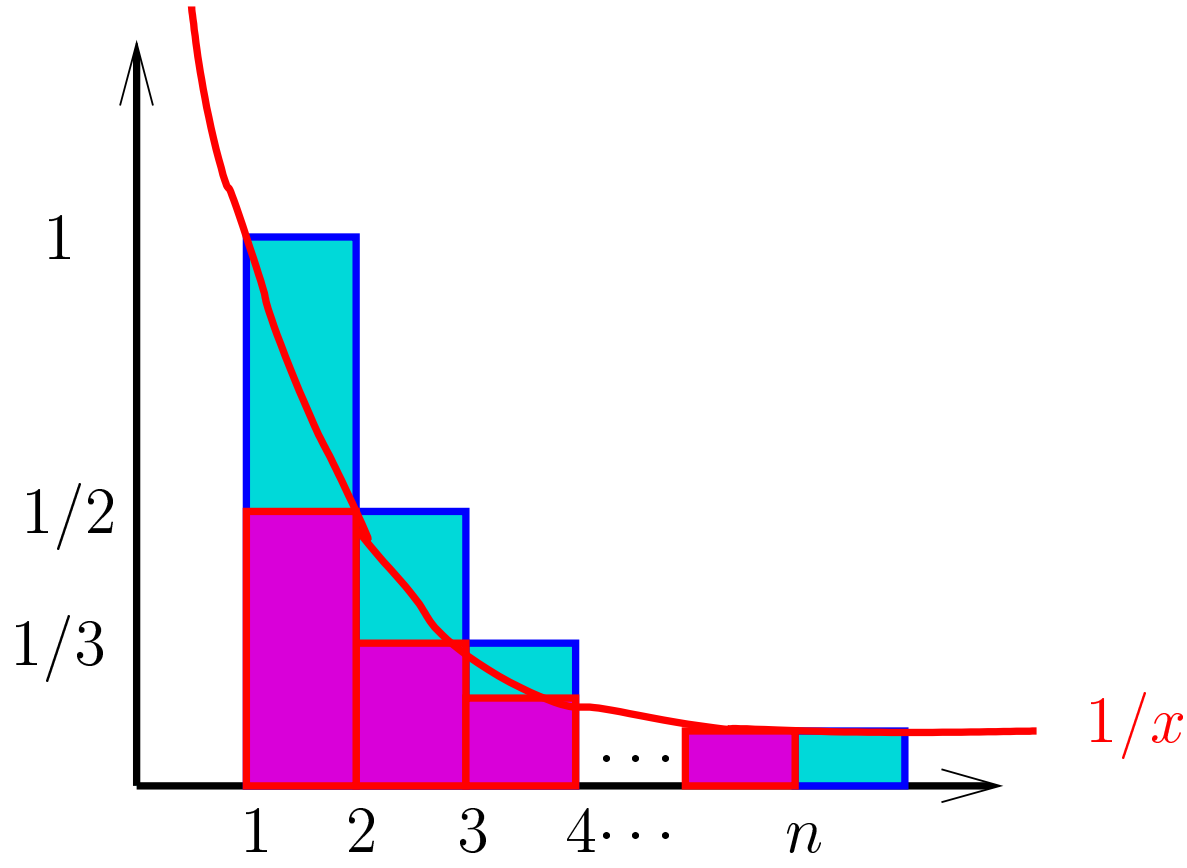
$$
\begin{aligned}
\mathrm{E}[X] &= \mathrm{E}[X_1 + \cdots + X_n] \\
&= \mathrm{E}[X_1] + \cdots + \mathrm{E}[X_n] \\
&= 1/1 + \cdots + 1/n \\
&< 1 + \ln n \\
&= \Theta(\lg n)
\end{aligned}
$$

$$2.92 < \frac{1}{1} + \cdots + \frac{1}{10} < 2.93 < 3.30 < 1 + \ln 10$$

$$5.18 < \frac{1}{1} + \cdots + \frac{1}{100} < 5.19 < 6.60 < 1 + \ln 100$$

$$9.78 < \frac{1}{1} + \cdots + \frac{1}{10000} < 9.79 < 10.21 < 1 + \ln 10000$$

# Série harmônica




$$\ln n = \int_1^n \frac{dx}{x} < H_n = 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \cdots + \frac{1}{n} < 1 + \int_1^n \frac{dx}{x} = 1 + \ln n$$

# Experimentos

Para cada valor de $n = 252, 512, 1024, \ldots$ foram geradas 10, 100 ou 200 amostras de seqüencias de inteiros através do trecho de código

```
for (i = 0; i < n; i++){
  v[i]=(int)((double)INT_MAX*rand()/(RAND_MAX+1
}
```

onde `rand()` é a função geradora de números (pseudo-)aleatórios da biblioteca do `C`.

A coluna $\mathrm{E}[\hat{X}]$ nas tabelas a seguir mostra o número médio de vezes que a linha 4 do algoritmo MAX foi executada para cada valor de $n$ e cada amostra de seqüências.

# Experimentos (10)

| $n$ | $E[\hat{X}]$ | $1 + \ln n$ |
|---:|---:|---:|
| 256 | 7.20 | 6.55 |
| 512 | 6.90 | 7.24 |
| 1024 | 7.30 | 7.93 |
| 2048 | 7.10 | 8.62 |
| 4096 | 10.20 | 9.32 |
| 8192 | 9.00 | 10.01 |
| 16384 | 10.80 | 10.70 |
| 32768 | 11.00 | 11.40 |
| 65536 | 12.50 | 12.09 |
| 131072 | 12.60 | 12.78 |
| 262144 | 13.20 | 13.48 |
| 524288 | 13.20 | 14.17 |
| 1048576 | 12.80 | 14.86 |
| 2097152 | 13.90 | 15.56 |
| 4194304 | 14.90 | 16.25 |
| 8388608 | 17.90 | 16.94 |

# Experimentos (100)

| $n$ | $E[\hat{X}]$ | $1 + \ln n$ |
|---:|---:|---:|
| 256 | 5.92 | 6.55 |
| 512 | 6.98 | 7.24 |
| 1024 | 7.55 | 7.93 |
| 2048 | 8.39 | 8.62 |
| 4096 | 8.97 | 9.32 |
| 8192 | 9.26 | 10.01 |
| 16384 | 10.44 | 10.70 |
| 32768 | 11.32 | 11.40 |
| 65536 | 11.66 | 12.09 |
| 131072 | 12.38 | 12.78 |
| 262144 | 13.17 | 13.48 |
| 524288 | 13.56 | 14.17 |
| 1048576 | 14.54 | 14.86 |
| 2097152 | 15.10 | 15.56 |
| 4194304 | 15.61 | 16.25 |
| 8388608 | 16.56 | 16.94 |

# Experimentos (200)

| $n$ | $E[\hat{X}]$ | $1 + \ln n$ |
|---:|---:|---:|
| 256 | 6.12 | 6.55 |
| 512 | 6.86 | 7.24 |
| 1024 | 7.38 | 7.93 |
| 2048 | 7.96 | 8.62 |
| 4096 | 8.87 | 9.32 |
| 8192 | 9.41 | 10.01 |
| 16384 | 10.28 | 10.70 |
| 32768 | 10.92 | 11.40 |
| 65536 | 11.31 | 12.09 |
| 131072 | 12.37 | 12.78 |
| 262144 | 12.92 | 13.48 |
| 524288 | 13.98 | 14.17 |
| 1048576 | 14.19 | 14.86 |
| 2097152 | 15.62 | 15.56 |
| 4194304 | 15.74 | 16.25 |
| 8388608 | 17.06 | 16.94 |

# 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 $\{s \in S : X(s) = k\}$

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

# Algoritmos aleatorizados ou probabilístico

CLRS 5.2, 5.3, C.2, C.3

# Algoritmos aleatorizados ou probabilístico

Um algoritmo é probabilístico ou aleatorizado se o seu comportamento não é determinado apenas pela entrada, mas também depende dos valores produzidos por um gerador de números aleatórios.

RANDOM$(a, b)$: devolve um inteiro $i$, $a \leq i \leq b$.

# Algoritmos aleatorizados ou probabilístico

Um algoritmo é probabilístico ou aleatorizado se o seu comportamento não é determinado apenas pela entrada, mas também depende dos valores produzidos por um gerador de números aleatórios.

RANDOM$(a, b)$: devolve um inteiro $i$, $a \leq i \leq b$.

Todos os inteiros devolvido por RANDOM são independentes dos inteiros previamente devolvidos com probabilidade uniforme, em outras palavras:

$$\Pr\{\text{RANDOM}(a, b) = i\} = ???.$$

# Algoritmos aleatorizados ou probabilístico

Um algoritmo é probabilístico ou aleatorizado se o seu comportamento não é determinado apenas pela entrada, mas também depende dos valores produzidos por um gerador de números aleatórios.

RANDOM$(a, b)$: devolve um inteiro $i$, $a \leq i \leq b$.

Todos os inteiros devolvido por RANDOM são independentes dos inteiros previamente devolvidos com probabilidade uniforme, em outras palavras:

$$\Pr\{\text{RANDOM}(a, b) = i\} = \frac{1}{b - a + 1}.$$

Existem implementações aproximadas de RANDOM. Veja D.E. Knuth, *Seminumerical algorithms*, volume 2, *The Art of Computer Programming*.

# Permutação aleatória uniforme

Problema: Obter uma permutação aleatória uniforme de um dado vetor $A[1\,..\,n]$ de números inteiros positivos distintos.

# Permutação *in-place*

Devolve uma permutação aleatória uniforme de $A[1\,..\,n]$.

```
PERMUTE-IN-PLACE(A, n)
1    para i ← 1 até n − 1 faça
2            j ← RANDOM(i, n)
3            A[j] ↔ A[i]
```

Não é óbvio que isso produz permutação aleatória uniforme de $A[1\,..\,n]$.

Consumo de tempo: $\Theta(n)$, supondo que RANDOM é $\Theta(1)$.

# Uniformidade

Considere uma permutação $B[1\,..\,n]$ de $A[1\,..\,n]$.

Qual é a probabilidade de PERMUTE-IN-PLACE$(A, n)$ produzir $B[1\,..\,n]$?

# Uniformidade

Considere uma permutação $B[1\,..\,n]$ de $A[1\,..\,n]$.

Qual é a probabilidade de PERMUTE-IN-PLACE$(A, n)$ produzir $B[1\,..\,n]$?

Relação invariante: Na linha 1 vale que

(i0) $\Pr\{B[1\,..\,i-1] = A[1\,..\,i-1]\} = (n-i+1)!/n!$.

# Uniformidade

Considere uma permutação $B[1\,..\,n]$ de $A[1\,..\,n]$.

Qual é a probabilidade de PERMUTE-IN-PLACE$(A, n)$ produzir $B[1\,..\,n]$?

Relação invariante: Na linha 1 vale que

(i0) $\Pr\{B[1\,..\,i-1] = A[1\,..\,i-1]\} = (n-i+1)!/n!$.

Na última iteração $i = n$ e da invariante (i0) temos que

$$\Pr\{B[1\,..\,n] = A[1\,..\,n]\} = (n-i+1)!/n! = 1/n!.$$

# Invariante

Considere uma iteração que não seja a última.
Da invariante (i0) sabemos que

$$\Pr\{B[1\,..\,i-1] = A[1\,..\,i-1]\} = (n-i+1)!/n!$$

Após a execução da linha 2 temos que

$$\Pr\{B[i] = A[j]\} = \frac{1}{n-i+1}\ .$$

Logo, após a execução da linha 3 temos que

$$\Pr\{B[1\,..\,i] = A[1\,..\,i]\} = \frac{(n-i+1)!}{n!} \times \frac{1}{n-i+1} = \frac{(n-i)!}{n!}.$$

Portanto, (i0) vale no início da próxima iteração.

# Permutação por ordenação

Devolve uma permutação aleatória uniforme de $A[1\,..\,n]$.

PERMUTE-POR-ORDENAÇÃO $(A, n)$
1 para $i \leftarrow 1$ até $n$ faça
2 $\quad P[i] \leftarrow$ RANDOM$(1, n^3)$
3 ordene $A[1\,..\,n]$ com chaves $P[1\,..\,n]$

No começo da linha 3, com grande probabilidade, $P[1\,..\,n]$ não tem elementos repetidos (Exercício 5.3-5 do CLRS).

Linha 3 faz permutação $A[j_1\,..\,j_n]$ de $A[1\,..\,n]$ tal que

$$P[j_1] \leq \cdots \leq P[j_n].$$

Consumo de tempo: $\Theta(n \lg n)$ se RANDOM for $\Theta(1)$

# Max aleatorizado

Devolve o elemento máximo de um vetor $A[1\,..\,n]$.

```
MAX-ALEATORIZADO(A, n)
0   PERMUTE(A, n)
1   max ← 0
2   para i ← 1 até n faça
3         se A[i] > max
4               então max ← A[i]
5   devolva max
```

Linha 0 faz uma permutação aleatória uniforme dos elementos de $A$.

Qual é a entrada que dá o pior caso?

# Exercícios

**Exercício 15.A** [CLRS 5.3-3]
Que acontece se trocarmos "RANDOM$(i, n)$" por "RANDOM$(1, n)$" no algoritmo PERMUTE-IN-PLACE? O algoritmo continua produzindo permutação aleatória uniforme?

**Exercício 15.B** [CLRS C.3-2]
Um vetor $A[1 \mathinner{..} n]$ contém $n$ números distintos aleatoriamente ordenados. Suponha que cada permutação dos $n$ números é igualmente provável. Qual é a esperança do índice do maior elemento do vetor? Qual é a esperança do índice do menor elemento do vetor?