

Monte Carlo e Las Vegas

- Dois tipos de algoritmos que diferem no local de sua incerteza

→ Monte Carlo : Na correteude.

PODE DAR **SOLUÇÕES ERRADAS**,
MAS GASTA **TEMPO DETERMINÍSTICO**

→ Las Vegas : No tempo de execução

DAÍ APENAS **SOLUÇÕES CORRETAS**
MAS GASTA **TEMPO ALEATÓRIO**

Monte Carlo

- DECISÃO DOIS TIPOS DE INSTÂNCIA : SIM E NÃO

- ERRO UNILATERAL

1. BASEADO NO SIM : NUNCA ERRA QUANDO DIZ SIM

2. BASEADO NO NÃO : NUNCA ERRA QUANDO DIZ NÃO

- ERRO BILATERAL : PODE ERRAR EM AMBAS RESPOSTAS

EX: IDENTIDADE DE POLINÔMIOS :

DADOS $F(x) = (x - a_1)(x - a_2) \dots (x - a_d)$

E $G(x) = b_d x^d + b_{d-1} x^{d-1} + \dots + b_0$

$a_i, b_i \in \mathbb{R}$

DECIDIR SE $F = G$

~> PODEMOS EXPANDIR F E COMPARAR OS COEFICIENTES : $O(d^d)$

~> PODEMOS TESTAR AS RAÍZES DE F EM G $O(d^2)$

EX: IDENTIDADE DE POLINÔMIOS:

DADOS $F(x) = (x - a_1)(x - a_2) \dots (x - a_d)$
e $G(x) = b_d x^d + b_{d-1} x^{d-1} + \dots + b_0$

DECIDIR SE $F = G$

~> Podemos expandir F e comparar os coeficientes: $O(d^d)$

~> Podemos testar as raízes de F em G : $O(d^2)$

~> Sortear um inteiro r entre 1 e $100d$ e testar se $F(r) = G(r)$ $O(d)$

Se $F(r) \neq G(r)$ então r é um certificado de que $F \neq G$.

Do contrário, retornamos SIM

~> O polinômio $H(x) = F(x) - G(x)$ possui no máximo d raízes em $[1, 100d]$

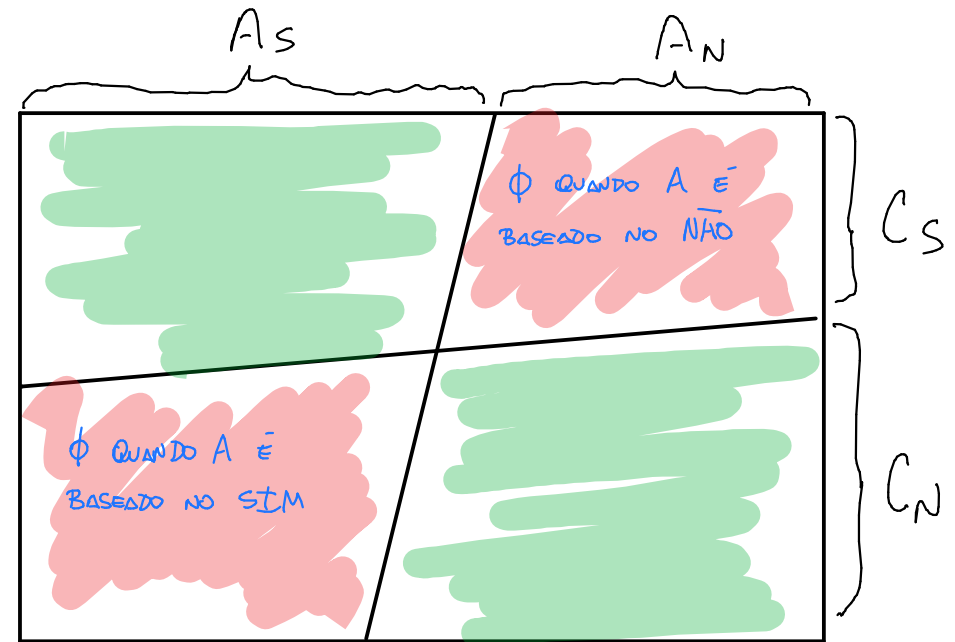
A probabilidade de r ser uma dessas raízes é $d/100d = 1/100$

Custo $O(d)$

PROBABILIDADES MONTE CARLO UNILATERAL

EVENTOS

- 1 A_S - O ALGORITMO RESPONDE SIM
- 2 A_N - O ALGORITMO RESPONDE NÃO
- 3 C_S - A RESPOSTA CORRETA É SIM
- 4 C_N - A RESPOSTA CORRETA É NÃO



GOSTARIAMOS DE ESTUDAR AS PROBABILIDADES

$$Pr(A_S | C_S), Pr(A_N | C_S), Pr(A_S | C_N), Pr(A_N | C_N)$$

- SE O ALG É BASEADO EM SIM, ENTÃO $A_S \subseteq C_S$

$$Pr(C_S | A_S) = 1$$

$$Pr(C_N | A_S) = 0$$

- SE O ALG É BASEADO EM NÃO, ENTÃO $A_N \subseteq C_N$

$$Pr(C_N | A_N) = 1$$

$$Pr(C_S | A_N) = 0$$

EX: ID. DE POLINÔMIOS.

BASEADO NO $\widetilde{\text{NÃO}}$, LOGO $A_N \subseteq C_N$

→ UM ALG BASEADO NO $\widetilde{\text{NÃO}}$ RESPONDE CORRETAMENTE SIM

PARA INSTÂNCIAS EM C_S . OU SEJA $C_S \subseteq A_S$.

→ DEVEMOS NOS PREOCUPAR COM A PROB. DE ENCONTRAR UM CERTIFICADO.

- CONSIDERE UM ALG. BASEADO NO $\tilde{NÃO}$ E QUE ERRO COM PROB NO MÁX ϵ_1 .

$$\rightarrow \Pr(A_S | C_N) \leq \epsilon_1$$

↓
INSTÂNCIAS
EM C_N

→ EXECUTE-O t VEZES $\leadsto A^t$

→ Qual a prob. global ϵ_t DE ERRO DE A_t ?

Ω
ESP.
DE UMA
EXECUÇÃO

$\Omega \times \Omega \times \dots \times \Omega$
ESP DE PROB. DE
VÁRIAS EXECUÇÕES

→ CERTAMENTE UMA INSTÂNCIA EM C_N

PARA A QUAL RESPONDEMOS SIM t VEZES $(A^1, A^2, \dots, A^t)$

$$\begin{aligned} \text{Logo, } \epsilon_t &= \Pr\left(C_N \cap \bigcap_{i=1}^t A_S^i\right) = \Pr(C_N) \cdot \Pr\left(\bigcap_{i=1}^t A_S^i \mid C_N\right) \\ &= \Pr(C_N) \cdot \Pr(A_S^t \mid C_N) \cdot \Pr\left(\bigcap_{i=1}^{t-1} A_S^i \mid C_N, A_S^t\right) \\ &= \Pr(C_N) \prod_{k=1}^t \Pr(A_S^k \mid C_N, \bigcap_{i=1}^{k-1} A_S^i) \\ &\stackrel{\text{IND DAS EXECUÇÕES}}{=} \underbrace{\Pr(C_N)}_{\leq 1} \prod_{k=1}^t \underbrace{\Pr(A_S^k \mid C_N)}_{\epsilon_1} \leq \epsilon_1^t \end{aligned}$$

LAS VEGAS

- O TEMPO É UMA VAR ALÉATORIA
- JULGAMOS PELA ESPERANÇA, VARIÂNCIA, ...

EX: BUSCA EM LISTA COM REPETIÇÃO.

$$\Omega = \{H, TH, TTH, T^3H, \dots\}$$

- LISTA COM n ENTRADAS
- NÚMEROS DE 0 A 9
- CADA NÚMERO APARECE $1/10$ DAS VEZES.

~> VARRER A LISTA DO COMEÇO AO FIM $\Theta(n)$

~> ÍMPARES E DEPOIS PARES $\Theta(n)$

~> TESTAR ALÉAT. UMA POSIÇÃO. REPETIR ATÉ ACHAR.

~> DISTRIB. GEOMÉTRICA COM PROB. DE SUCESSO $p = 1/10$

=> ESPERANÇA : $1/p = 10 = O(1)$.

TRANSFORMAR LAS VEGAS (QUE BUSCA CERT. P/ SIM) EM MONTE CARLO:

RODAR O LV UM NÚMERO LIMITADO DE VEZES (t). SE NÃO ENCONTRAR, RESPONDE NÃO.

PROB. DE ERRO DO MC = PROB. DO LV PRECISAR DE MAIS TEMPO

SEJA X A VAR. DE TEMPO DO LV.

SE $t = k\mu$, ONDE $\mu = E(X) = 1/p$, TEMOS

$$\epsilon_{MC} \leq \Pr(X \geq k\mu) \stackrel{\text{MARKOV}}{\leq} \frac{E(X)}{k\mu} = 1/k$$

TRANSFORMAR MC EM LV

$$\frac{1}{100} \quad \frac{99}{100}$$

~* MENOS EFICAZ. O LV PODE RODAR ? / SEMPRE

$$\frac{100}{99} < 2$$

EX: SE A É BASEADO NO NÃO E REPETIMOS ELE ATÉ ENCONTRAR CERTIFICADO

O QUE ACONTECE COM INSTÂNCIAS SIM?

~* EM ALGUNS CASOS PODEMOS FAZER COM QUE AS EXECUÇÕES DE MC **NÃO** SEJAM IND. E AI O ALG. PODE PARAR DEPOIS DE FAZER TODAS AS ESCOLHAS.

~* ISSO SERIA UM CERTIFICADO

~* SE HÁ DOIS ALGS Alg_S E Alg_N BASEADOS NO SIM E NÃO RESP. EXECUTAMOS ALTERNADAMENTE

→ A PROB. DE SUCESSO EM CADA EXEC É $\geq p = \min\{p_S, p_N\}$

→ LOGO, O TEMPO ESPERADO É $1/p$.

CLASSES DE COMPLEXIDADE

- RP (RANDOMIZED POLYNOMIAL TIME): Probs PARA OS QUAIS EXISTEM ALGORITMOS QUE RESPONDEM **SIM** COM PROB $\geq \frac{1}{2}$ SE $I \in C_S$,
E **NÃO** COM PROB. $\leq \frac{1}{2}$ SE $I \in C_N$

→ Probs. PARA OS QUAIS EXISTE Alg DE MONTE CARLO BASEADO NO SIM.

- Co-RP: Probs. t.q. $\exists$ Alg MONTE CARLO BASEADO NO NÃO,

- ZPP: (ZERO-ERROR PROBABILISTIC POLYNOMIAL TIME)
Probs. PARA OS QUAIS EXISTE Alg. DE LAS VEGAS DE
TEMPO ESPERADO POLYNOMIAL.

CLASSES DE COMPLEXIDADE

- BPP (Bounded-error Probabilistic Polynomial Time)
Probs. Para as quais existe Alg de Monte Carlo bilateral onde tanto a prob de resp. SIM com IEC_S, quanto responder NÃO com IEC_N são $> 3/4$.
- PP (Probabilistic Polynomial Time)
Probs. Para as quais existe Alg de Monte Carlo bilateral onde tanto a prob de resp. SIM com IEC_S, quanto responder NÃO com IEC_N são $> 1/2$.

OBS: $BPP \subseteq PP$. A diferença é o número de repetições do Alg para que o erro seja $< \epsilon$.

~* BPP = Polinomial

PP = Exponencial

PARADIGMAS COMBINATÓRIOS

1. Bolas e Latas

- Há m bolas idênticas
- Há n latas
- As bolas são associadas às latas aleatoriamente (com prob $1/n$)
- Uma lata pode receber mais de uma bola
- Quantas latas ficam vazias?
- Qual o número esperado de latas com mais do que k bolas?
- Quantas bolas precisamos? / que existam mais latas com bolas do que vazias?
- Quantas bolas há na lata mais cheia?

EX: m PESSOAS, HÁ DUAS PESSOAS QUE FAZEM ANIVERSÁRIO NO MESMO DIA?

SEJA A_i O EVENTO NO QUAL A BOLA i CAIR NUMA LATA DESOCUPADA.

A PROBABILIDADE DE TERMOS DUAS BOLAS NA MESMA LATA É

$$p = \Pr\left(\bigwedge A_i\right) = \prod_{i=1}^m \Pr\left(A_i \mid \bigwedge_{j < i} A_j\right)$$

NOTE QUE $\Pr\left(A_i \mid \bigwedge_{j < i} A_j\right)$ É A PROBABILIDADE DE i CAIR EM UMA LATA, DADO QUE HÁ $i-1$ LATAS OCUPADAS, OU SEJA $\frac{m-i+1}{m}$ LOGO

$$p = \prod_{i=1}^m \left(\frac{m-i+1}{m}\right) = \frac{\binom{m}{m}}{m^m} < \frac{1}{2}$$

- Qual o valor de m para que seja mais provável haver duas pessoas com mesmo aniversário?

~ DEVEMOS RESOLVER A EQUAÇÃO $p < \frac{1}{2}$: $m \approx \sqrt{2m \ln m} = O(\sqrt{m \log m})$

COLECCIONADOR DE CUPONS

- Há um conjunto de cupons $D = \{d_1, \dots, d_m\}$
- Gostaríamos de ter um de cada

→ Seq. DE VAR. IND $X_1, \dots$ CADA UMA INDICA O ÍNDICE QUE SAIU.

~ CADA ÍNDICE TEM PROB. $1/n$ DE SAIR.

DEFINA $W_{m,k}$ = # DE SORTEIOS REALIZADOS ATÉ QUE SE TENHA
OBTIDO k ITENS DISTINTOS, DOS m EXISTENTES

→ ESTAMOS INTERESSADOS EM $E[W_{m,n}]$.

~ Parece com bolas x latas

~ PARECE COM BOLAS X LATAS
 ~ SORTEIOS ~ ITENS EXISTENTES
 n LATAS, INFINITAS BOLAS

DEFINA $W_{m,k} = \#$ DE SORTEIOS REALIZADOS ATÉ QUE SE TENHA
OBTIDO k ITENS DISTINTOS, DOS m EXISTENTES

DEFINA $Z_i =$ QUANTIDADE DE SORTEIOS PARA QUE O $\#$ DE CUPONS
DISTINTOS VÁ DE $i-1$ P/ i .

~> QUANDO O COLECIONADOR POSSUI $i-1$ CUPONS DISTINTOS

A PROB. DE SORTEARMOS UM CUPOM DIFERENTE É $p_i = \frac{m - (i-1)}{m}$

~> LOGO, CADA Z_i É UMA V.A. GEOMÉTRICA COM PROB. DE SUCESSO p_i

$$\Rightarrow E(Z_i) = \frac{1}{p_i} = \frac{m}{m - (i-1)}$$

$$\Rightarrow W_{m,k} = Z_1 + \dots + Z_k \Rightarrow E(W_{m,k}) = \sum_{i=1}^k E(Z_i)$$

$$E[W_{m,k}] = \sum_{i=1}^k \frac{m}{m - i + 1} \Rightarrow E[W_{m,m}] = \sum_{i=1}^m \frac{m}{m - i + 1} = m \sum_{i=1}^m \frac{1}{i}$$
$$= m \ln m + \Theta(m).$$

EX: MENSAGEM EM PACOTES CLIENTE $\rightarrow$ SERVIDOR

- CAMINHOS LONGOS (MUITOS ROTEADORES)
- UM PACOTE PODE GUARDAR A INFO DE NO MAX UM ROTEADOR
 $\Rightarrow$ OU OCUPARIA MUITO ESPAÇO/BANDA
- SE HÁ m ROTEADORES, CADA PACOTE GUARDA A INFO DE UM ROTEADOR COM PROB $1/m$.

$\Rightarrow$ O NÚMERO DE PACOTES QUE PRECISAM SER RECEBIDOS PARA TERMOS INFO DE TODOS OS ROTEADORES É

$$m \log m + \theta(m)$$

- E SE O PACOTE NÃO SABE QUANTOS ROTEADORES TEM NO CAM.

COMO FAZER PARA QUE CADA ROTEADOR APAREÇA COM PROB. $1/m$?

ESTRATÉGIA:

- 1) GUARDA INFO DO PRIMEIRO ROTEADOR
 - 2) NO k -ÉSIMO ROTEADOR, TROCA A INFO GUARDADA PELA INFO DO ROTEADOR ATUAL COM PROB $1/k$.
- Para que a info do k -ésimo roteador chegue no servidor, precisamos trocar no passo z (Prob. $1/k$) e NÃO trocar depois (Prob. $\prod_{i=k+1}^m \frac{i-1}{i}$)

$$\Rightarrow \Pr(\text{GUARDAR a } k\text{-INFO}) = \frac{1}{k} \cdot \frac{k}{k+1} \cdot \frac{k+1}{k+2} \cdots \frac{m-1}{m} = \frac{1}{m}$$

Análise Probabilística

- Algoritmos determinísticos executam da mesma forma para a mesma instância. Por isso os analisamos pelo **Pior caso**.
- Na prática pode ser que o **caso típico** seja muito melhor que o pior caso. Ou seja, que a probabilidade de ocorrer o pior caso seja muito baixa.
- Analisar o alg levando em consideração a **distribuição das instâncias** é chamado de **análise probabilística**.
 - De algs determinísticos e randomizados!

EX: QUICKSORT. Pior caso : $O(n^2)$
caso médio : $O(n \log n)$

OBS: Podemos randomizar o QUICKSORT!

QUICKSORT-RANDOMIZADO (S)

1. REORDERE S DE FORMA ALEATÓRIA
2. QUICKSORT(S)

- Diferentes execuções de algs. randomizados possuem tempos diferentes p/ a mesma instância. Por isso analisamos o caso médio.

BUCKET SORT

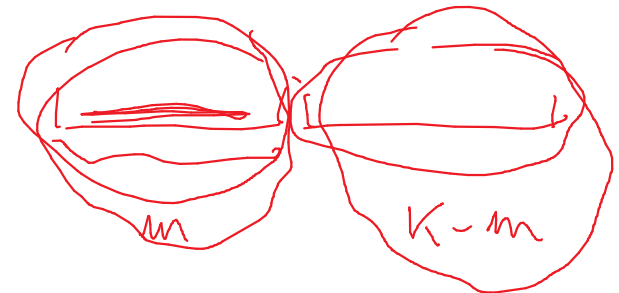
- Roda em tempo **linear!**
- Qual o modelo para isso?

→ Uma entrada com $n = 2^m$ NÚMEROS

→ Escolhido de forma aleatória e uniforme no intervalo $[0, 2^k]$
com $k > m$.

- Há $n = 2^m$ ^{lata} BUCKETS rotulados de 0 a $n-1$ em binário
- Cada número é colocado no bucket que representa seus primeiros m dígitos binários.

→ Feito em tempo $O(m)$



- Seja X_J o # de elts no bucket J .

→ Bolas e latas } X_J é V.A. Binom. $B(n, 1/n)$

↓
Entrada ↓
 Buckets

- Ordenamos cada bucket individualmente (usando, por ex, o Quicksort)

→ Tempo quadrático em X_J : $c X_J^2$

- O valor esperado para o tempo total $X = \sum c X_J^2$ é

$$E[X] \stackrel{\text{lin. esp.}}{=} E\left[\sum_J c X_J^2\right] \stackrel{+}{=} c \cdot n \cdot E[X_1^2].$$

→ Lembremos que como X_1 é V.A. Bin. $B(n, p)$, temos

$$E[X_1^2] = n(n-1)p^2 + np$$

$$\begin{cases} X = Y_1 + \dots + Y_n \\ X^2 = \sum_{i=1}^n Y_i^2 + \sum_{i \neq j} Y_i Y_j \\ E(X_i^2) = n E(Y_i) + n(n-1) p^2 \end{cases}$$

→ Neste caso, $p = 1/n$ e, portanto, $E[X_1] = 2 - 1/n < 2$.

⇒ Bucket Sort roda em tempo esperado $O(n)$

O MÉTODO PROBABILÍSTICO

- PROVAS DE EXISTÊNCIA

- É POSSÍVEL DE-RANDOMIZAR ALGS. RANDOMIZADOS.

EX: É POSSÍVEL COLORIR AS ARESTAS DO K_{40} COM DUAS CORES SEM CLIQUE MONOCROMÁTICA DE TAMANHO 8?

- HÁ 780 ARESTAS. LOGO 2^{780} COLORIÇÕES

→ SEJAM $C_1, \dots, C_{\binom{40}{8}}$ AS CLIQUES DE TAMANHO 8 EM $G = K_{40}$.

→ SEJA M_i O EVENTO "TODAS AS ARESTAS DE C_i SÃO AZUIS OU TODAS AS ARESTAS DE C_i SÃO VERMELHAS"

$$\rightarrow \Pr(M_i) = \frac{1}{2^{\binom{8}{2}-1}} = 2^{-27}$$

QUEREMOS CALCULAR $\Pr(\bigcap \bar{M}_i) = 1 - \Pr(\bigcup M_i)$

Pelo LIMITE DA UNIÃO, TEMOS

$$\Pr(\bigcup M_i) \leq \sum \Pr(M_i) = \frac{\binom{40}{8}}{2^{27}} < 0.573$$

$$\text{Logo, } \Pr(\bigcap \bar{M}_i) > 0.427 > 0$$

EX: TODO GRAFO CONTÉM UM GRAFO
BIP. COM METADE DAS ARESTAS.