Monte Carlo e Las Vegas

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

  → **Monte Carlo**: Na corretude.

  Pode dar soluções erradas,
  mas gasta tempo determinístico

  → **Las Vegas**: No tempo de execução

  Dá apenas soluções corretos
  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 $F(x) = (x - a_1)(x - a_2) \cdots (x - a_d)$ e $G(x) = b_d x^d + b_{d-1} x^{d-1} + \cdots + b_0$ $\Rightarrow$ $a_i, b_i \in \mathbb{R}$

Decidir se $F = G$

$\rightsquigarrow$ Podemos expandir F e comparar os coeficientes: $O(2^d)$

$\rightsquigarrow$ Podemos testar as raízes de F em G

EX: IDENTIDADE DE POLINÔMIOS:

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

DECIDIR SE $F = G$

↝ PODEMOS EXPANDIR F E COMPARAR OS COEFICIENTES : $O(2^d)$

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

↝ SORTEAR UM INTEIRO r ENTRE 1 E $100d$ E TESTAR SE $F(r) = G(r)$

SE $F(r) \neq G(r)$ ENTÃO r É UM CERTIFICADO DE QUE $F \neq G$.

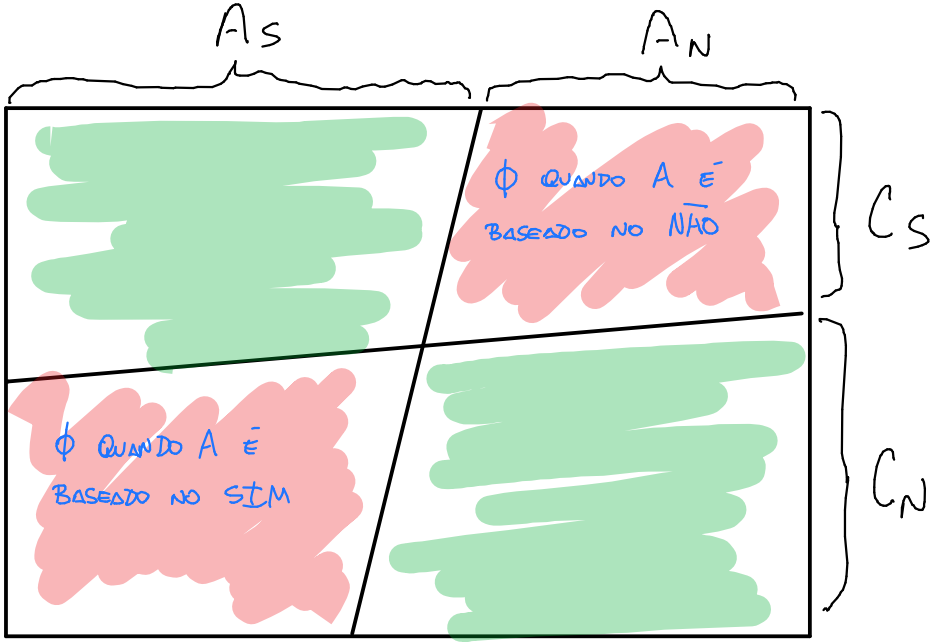
DO CONTRÁRIO, RETORNAMOS SIM

→ O POLINÔMIO $H(x) = F(x) - G(x)$ (POLINOMIO DE GRAU d) 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

Gostaríamos de estudar as probabilidades

$Pr(A_S \mid C_S)$, $Pr(A_N \mid C_S)$, $Pr(A_S \mid C_N)$, $Pr(A_N \mid C_N)$

$Pr(A_N \mid C_S) = 0$

- Se o A/G é baseado em SIM, então $A_S \subseteq C_S$

$$Pr(C_S \mid A_S) = 1 \qquad Pr(C_N \mid A_S) = 0$$

- Se o A/G é baseado em NÃO, então $A_N \subseteq C_N$

$$Pr(C_N \mid A_N) = 1 \qquad Pr(C_S \mid A_N) = 0$$

$\neq$

EX: ID. DE POLINÔMIOS.

BASEADO NO NÃO. LOGO $A_N \subseteq C_N$

→ UM A/G BASEADO NO 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 NÃO E QUE ERRA (→ INSTÂNCIAS EM C_N) COM PROB NO MÁX ε_1.

→ $Pr(A_S \mid C_N) \leq \varepsilon_1$

→ EXECUTE-O t VEZES ⇝ A^t

→ QUAL A PROB. GLOBAL ε_t DE ERRO DE A_t?

Ω	$\Omega \times \Omega \times \dots \times \Omega$
ESP. DE UMA EXECUÇÃO	ESP DE PROB. DE VÁRIAS EXECUÇÕES

⇝ CERTAMENTE UMA INSTÂNCIA EM C_N PARA A QUAL RESPONDEMOS SIM t VEZES $(A^1, A^2, \dots, A^t)$

LOGO, $\varepsilon_t = Pr\left(C_N \cap \bigcap_{i=1}^{t} A_S^i\right) = Pr(C_N) \cdot Pr\left(\bigcap_{i=1}^{t} A_S^i \mid C_N\right)$

$= Pr(C_N) \cdot Pr(A^t \mid C_N) \cdot Pr\left(\bigcap_{i=1}^{t-1} A_S^i \mid C_N, A_S^t\right)$

$= Pr(C_N) \prod_{K=1}^{t} Pr\left(A_S^K \mid C_N, \bigcap_{i=1}^{K-1} A_S^i\right)$

$\overset{\text{IND DAS EXECUÇÕES}}{=} Pr(C_N) \prod_{K=1}^{t} Pr(A_S^K \mid C_N)$

$Pr(C_N) \leq 1$; $Pr(A_S^K \mid C_N) \leq \varepsilon_1$

$\leq \varepsilon_1^t$

LAS VEGAS

- O TEMPO É UMA VAR ALEATÓRIA

- JULGAMOS PELA ESPERANÇA, VARIÂNCIA, ...

EX: BUSCA EM LISTA COM REPETIÇÃO.

$\Omega = \{H, TH, TTH, T^3H, (\ldots)\}$

- 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 ALEAT. UMA POSIÇÃO. REPETIR ATÉ ACHAR.

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

$\Rightarrow$ 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

$$\varepsilon_{MC} \leq Pr\left(X \geq \overset{t}{k\mu}\right) \overset{\text{Markov}}{\leq} \frac{E(X)}{k\mu} = 1/k$$

Transformar MC em LV

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

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

↝ Menos eficaz. O LV pode rodar ?/ sempre

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 aí 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 $\geq p = \text{Min}\{p_S, p_N\}$

→ 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. 1 se $I \in C_N$

  ⇝ Probs. para os quais existe Alg de Monte Carlo baseado no SIM.

- Co-RP: Probs. t.q. ∃ Alg Monte Carlo baseado no NÃO.

- ZPP: (Zero-error Probabilistic Polinomial time)
  Probs. para os quais existe Alg. de Las Vegas de tempo esperado polinomial.

Classes de Complexidade

- BPP (Bounded-error Probabilistic Polinomial time)
  Probs. para os quais existe Alg de Monte Carlo bilateral onde tanto a prob de resp. SIM com $I \in C_S$, quanto responder NÃO com $I \in C_N$ são $\geq 3/4$.

- PP (Probabilistic Polinomial time)
  Probs. para os quais existe Alg de Monte Carlo bilateral onde tanto a prob de resp. SIM com $I \in C_S$, quanto responder NÃO com $I \in C_N$ são $> 1/2$.

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

↝ 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 PARA 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

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

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

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

- QUAL O VALOR DE m PARA QUE SEJA MAIS PROVÁVEL HAVER DUAS PESSOAS COM MESMO ANIVERSÁRIO?

↝ DEVEMOS RESOLVER A EQUAÇÃO $p < \frac{1}{2}$: $m \approx \sqrt{2m \ln n^n} = O\left(\sqrt{n \log^n}\right)$

Colecionador de Cupons

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

↝ Seq. de var. ind $X_1, \ldots$ cada uma indica o índice que saiu.
↝ Cada índice tem prob. $1/m$ de sair.

Defina $W_{m,k}$ = # de sorteios realizados até que se tenha obtido k itens distintos, dos m existentes

↝ Estamos interessados em $E[W_{m,m}]$.

↝ Parece com bolas x latas
- sorteios
- itens existentes

m latas, infinitas bolas

DEFINA $W_{n,k}$ = # DE SORTEIOS REALIZADOS ATÉ QUE SE TENHA OBTIDO k ITENS DISTINTOS, DOS n 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 $p_i = \frac{n-(i-1)}{n}$

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

$\Rightarrow E(Z_i) = 1/p_i = \frac{n}{n-(i+1)}$

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

$E[W_{n,k}] = \sum_{i=1}^{n} \frac{n}{n-i+1} \Rightarrow E[W_{n,n}] = \sum_{i=1}^{n} \frac{n}{n-i+1} = n \sum_{i=1}^{n} \frac{1}{i}$

$= n \ln n + \Theta(n)$.

EX: MENSAGEM EM PACOTES CLIENTE → SERVIDOR

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

⇒ O NÚMERO DE PACOTES QUE PRECISAM SER RECEBIDOS PARA TERMOS INFO DE TODOS OS ROTEADORES É

$$n \log n + \Theta(n)$$

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

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

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 2 (PROB. $1/k$) E NÃO TROCAR DEPOIS (PROB. $\prod_{i=k+1}^{n} \frac{i-1}{i}$)

$$\Rightarrow \Pr(\text{GUARDAR A K-INFO}) = \frac{1}{k} \cdot \frac{k}{k+1} \cdot \frac{k+1}{k+2} \cdot \ldots \cdot \frac{n-1}{n} = \frac{1}{n}$$

Análise Probabilística

- Algoritmos determinísticos executam da mesma forma para a mesma instância. Por isso os analizamos 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.

→ Analizar 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. reordene S de forma aleatória
2. Quicksort (S)

- Diferentes execuções de algs. randomizados possuem tempos diferentes p/ a mesma instância. Por isso analizamos 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 $n = 2^m$ 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 $O(n)$

- SEJA X_j O # DE ELTS NO BUCKET j.

→ BOLAS (↓ ENTRADA) E LATAS (↓ BUCKETS) } X_j É V.A. BINOM. $B(n, 1/n)$

- 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

$$E[X] \overset{\text{lin esp.}}{=} E\left[\sum_j c X_j^2\right] = c \cdot n \cdot E[X_1^2].$$

→ LEMBREMOS QUE COMO X_1 É V.A. BIN. $B(n,p)$, TEMOS

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

$$X = Y_1 + \cdots + Y_n$$

$$X^2 = \sum_{i=1}^{n} Y_i^2 + \sum_{i \neq j} Y_i Y_j$$

$$E(X_i^2) = n\,E(Y_i) + n(n-1)p^2$$

→ 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

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

- 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} COLORAÇÕES

↝ 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"

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

QUEREMOS CALCULAR $P_R(\cap \bar{M}_i) = 1 - P_R(\cup M_i)$

PELO LIMITE DA UNIÃO, TEMOS

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

LOGO, $P_R(\cap \bar{M}_i) > 0.427 > 0$ ∎