Colóquio Brasileiro de Matemática - Exercícios de Algoritmos Randomizados

Capítulo 1

Exercício 2.

Considere os seguintes eventos associados a uma execução do algoritmo que consiste na aplicação do exame de sangue em um paciente:

$\triangleright$ A_S: o algoritmo responde `SIM`.

$\triangleright$ A_N: o algoritmo responde `NÃO`.

$\triangleright$ C_S: a resposta correta para aquela entrada é `SIM`.

$\triangleright$ C_N: a resposta correta para aquela entrada é `NÃO`.

Pelos dados do enunciado, temos que:

- $Pr[A_S|C_S] = 0{,}95$: Se a pessoa realmente está doente, o exame responde "sim" com 95% de certeza.
- $Pr[A_N|C_S] = 0{,}05$: Existe 5% de chance de erro quando a pessoa está doente (ou seja, o exame não indica doença).
- $Pr[A_S|C_N] = 0$: Como o algoritmo (exame) é baseado no sim, não há erro quando ele indica doença.
- $Pr[A_N|C_N] = 100$: Como o algoritmo (exame) é baseado no sim, não há possibilidade de erro quando o paciente não está doente (ou seja, não há falso positivo).

Observe ainda, que nos são informados os valores $Pr[C_S] = 1/3$ e $Pr[C_N] = 2/3$.

Suponha que nosso algoritmo consiste não apenas em aplicar o exame uma vez, mas em repeti-lo k vezes. Neste caso, $Pr[A_N|C_S] = (0{,}05)^k$ e $Pr[A_S|C_S] = 1 - (0{,}05)^k$.

Nosso objetivo é descobrir, dado que o exame não indicou doença, qual a real probabilidade de o paciente estar doente (queremos que este valor seja menor que 1%). Ou seja, queremos saber $Pr[C_S|A_N]$. Temos que

$$Pr[C_S|A_N] = \frac{Pr[C_S, A_N]}{Pr[A_N]}.$$

Ou seja,

$$Pr[C_S|A_N] = \frac{Pr[A_N|C_S]Pr[C_S]}{Pr[A_N|C_S]Pr[C_S] + Pr[A_N|C_N]Pr[C_N]}.$$

Substituindo pelos valores conhecidos,

$$Pr[C_S|A_N] = \frac{(0{,}05)^k(1/3)}{(0{,}05)^k(1/3) + 2/3}.$$

Basta, agora, descobrir para que valores de k a expressão acima é menor do que 1%. Efetuando as contas, vemos que, para uma única repetição do exame, a chance de "falso negativo" é de 2,45%. Se efetuarmos dois exames, essa chance já será de 0,12%, atingindo nosso objetivo com apenas duas repetições.

Exercício 3

Podemos utilizar o resultado obtido no exercício anterior, de que

$$Pr[C_S|A_N] = \frac{Pr[A_N|C_S]Pr[C_S]}{Pr[A_N|C_S]Pr[C_S] + Pr[A_N|C_N]Pr[C_N]}.$$

No entanto, agora, assumimos que a probabilidade de o paciente estar doente é de $Pr[R_r = S] = 80\%$. Se o exame indica que não existe doença, então a probabilidade de que o paciente esteja doente cai para 16,67%.

Capítulo 2

Exercício 1.

(a) São necessárias 8 pessoas para que, com certeza, haja duas aniversariando no mesmo dia da semana. (Se houvesse apenas 7, poderíamos ter uma em cada dia da semana.)

(b) A probabilidade p_k de que **não haja** duas pessoas com aniversário no mesmo dia, em um grupo com k pessoas, é

$$p_k = \frac{7}{7} \times \frac{6}{7} \times \frac{5}{7} \times \ldots \times \frac{7-k+1}{7}.$$

Para que seja mais provável que **haja** duas pessoas com aniversário no mesmo dia, devemos ter $p < 1/2$. Calculando este valor para $k = 3$, temos

$$p_3 = \frac{7}{7} \times \frac{6}{7} \times \frac{5}{7} = \frac{30}{49} > 1/2;$$

e para $k = 4$,

$$p_4 = \frac{7}{7} \times \frac{6}{7} \times \frac{5}{7} \times \frac{4}{7} \approx 0{,}35 < 1/2.$$

Ou seja, em um grupo com 4 pessoas, a probabilidade $1-p$ de dois aniversários no mesmo dia da semana é cerca de 65%.

Exercício 2.

Usaremos o paradigma do colecionador de cupons, onde queremos que todos os cupons (possíveis resultados) tenham sido obtidos. Temos que

$$E[W_{6,6}] = 6 \times \sum_{i=1}^{6} \frac{1}{i} = \frac{36}{21} \approx 14{,}7.$$

Se usarmos a desigualdade de Markov para aproximar $P(W_{6,6} \geq 10)$, teremos:

$$P(W_{6,6} \geq 10) \leq \frac{E[W_{6,6}]}{10} \approx 1{,}47.$$

Mas isso não nos fornece nenhuma informação, já que, certamente, a probabilidade será no máximo 1. Tentemos, então, usar a desigualdade de Chebyshev para melhorar nossa estimativa. Para isso, precisaremos calcular a variância de $W_{6,6}$. Mas, como essa variável aleatória é igual à soma de variáveis aleatórias geométricas independentes, então, a variância de $W_{6,6}$ será soma das variâncias destas variáveis aleatórias geométricas. Ou seja,

$$Var[W_{6,6}] = \sum_{i=1}^{6} \frac{1 - p_i}{p_i^2},$$

onde $p_i = 1-(i-1)/6$ é a probabilidade associada à i-ésima variável aleatória geométrica. Calculando, obtemos

$$Var[W_{6,6}] = 6 \times \sum_{i=1}^{6} \frac{(i-1)}{(6-i+1)^2} \approx 6{,}5.$$

Podemos, agora, usar a desigualdade de Chebyshev e obter uma approximação para a probabilidade de que $W_{6,6}$ **não esteja** entre 10 e 19,4:

$$P(|W_{6,6} - E[W_{6,6}]| \geq 4{,}7) \leq \frac{Var[W_{6,6}]}{4{,}7^2} \approx 0,3.$$

Ou seja, a probabilidade de que $W_{6,6}$ **esteja** entre 10 e 19,4 (inclusive) é maior que 70%. Mas essa probabilidade é menor que a probabilidade de que $W_{6,6}$ seja maior que ou igual a 10. Logo, podemos concluir que a probabilidade de que o número de lançamentos maior que ou igual a 10 é de pelo menos 70%. (Observação: poderíamos melhorar mais um pouco nossa estimativa substituindo 4,7 por 5,69, em nossa última conta. Por que isso é possível?)

Exercício 5.

Denotemos por $C_{x_i x_j}$ uma variável aleatória de Bernoulli cujo sucesso indica que houve inversão entre os elementos x_i e x_j. É fácil ver que os elementos x_i e x_j terão suas posições trocadas se o maior deles está à esquerda do menor deles. Ou seja,

$$C_{x_i x_j} = \begin{cases} 1 & \text{se o maior dos elementos está à esquerda do menor;} \\ 0 & \text{caso contrário.} \end{cases}$$

Como a entrada é aleatória, ambos os casos podem acontecer com igual probabilidade, ou seja, $C_{x_i x_j}$ é uma variável aleatória de Bernoulli com probabilidade de sucesso 1/2.

O número total de inversões efetuadas pelo algoritmo é uma variável aleatória T que é a soma de todas as variáveis aleatórias $C_{x_i x_j}$:

$$T = \sum_{x_i, x_j} C_{x_i x_j}.$$

Temos, então, para o valor esperado de T,

$$E[T] = E\left[\sum_{x_i, x_j} C_{x_i x_j}\right] = \binom{n}{2} \times \frac{1}{2} = \frac{n(n-1)}{4} = \Theta(n^2).$$

Capítulo 5

Exercício 1.

Escolha, uniforme e aleatoriamente, valores verdade para as variáveis. A probabilidade de que cada cláusula seja satisfeita é $1-(1/2)^3 = 7/8$. Logo, o número esperado de cláusulas satisfeitas é $7m/8$ e, portanto, existe alguma atribuição que satisfaça a pelo menos essa quantidade de cláusulas.

Exercício 2.

(a) Isso equivale a mostrar que toda 2-coloração das arestas do K_6 contém algum triângulo monocromático, e que o mesmo não vale para K_5.

Primeiramente, contemos o número de triplas ordenadas de vértices (x, y, z) tais que xy é amarela e yz é azul. Um vértice qualquer poderá ser o vértice central de $0 \times 5 = 0$, $1 \times 4 = 4$ ou $2 \times 3 = 6$ dessas triplas, pois existem 5 arestas incidentes a ele, e as combinações de cores para estas arestas são 0 e 5, 1 e 4 ou 2 e 3. Portanto, o número máximo para tais triplas é $6 \times 6 = 36$. Observe, agora, que dado um triângulo não-monocromático xyz, devem existir exatamente duas de tais triplas. Portanto, podem existir, no máximo, $36/2 = 18$ triângulos não-monocromáticos. Como existem 20 triângulos no K_6, pelo menos dois deles vão ser monocromáticos.

Além disso, é possível construir uma 2-coloração das arestas do K_5 sem que haja triângulo monocromático. Tal coloração consiste em um subgrafo C_5 colorido de amarelo e seu complementar (que também é um C_5) colorido de azul. Concluímos, assim, a demonstração.

(b) Da seção 5.1.1, sabemos que é possível 2-colorir as arestas de um K_{40} de forma que não haja clique monocromática de tamanho maior ou igual a 8. Ou seja, $R(8, 8) > 40$.

(c) Considere o grafo completo G com n vértices e suponha que atribuímos as cores amarelo ou azul a cada aresta de G de forma aleatória e uniforme. Seja M_i o evento em que a i-ésima das C_n^5 cliques de tamanho 5 de G **é monocromática**. Sabemos que $Pr[M_i] = 2^{-(C_5^2-1)} = 2^{-9}$. A probabilidade de que todas as cliques de tamanho 5 de G sejam **não-monocromáticas** é

$$Pr\left[\bigcap_{i=1}^{C_n^5} \overline{M_i}\right] = 1 - Pr\left[\bigcup_{i=1}^{C_n^5} M_i\right].$$

Pelo limite da união,

$$Pr\left[\bigcup_{i=1}^{C_n^5} M_i\right] \leq \sum_{i=1}^{C_n^5} Pr[M_i] = C_n^5 2^{-9}.$$

Se este valor for menor que 1, então a probabilidade de que todas as cliques de tamanho 5 sejam não-monocromáticas será positiva. Então, pelo método probabilístico, saberemos que existe alguma coloração do K_n em que todas as cliques K_5 são não-monocromáticas, e, portanto, que $R(5, 5) > n$. Observe que

$$C_n^5 2^{-9} \leq 2^{-9} \frac{n^5}{5!}$$

e, portanto, se $n^5 < 5!/2^{-9} \therefore n < 9{,}07$, existe coloração de K_n em que todas as cliques K_5 são não-monocromáticas. Assim, $R(5,5) \geq 10$.

Exercício 3.

(a) Considere uma coloração aleatória de K_n (escolhida uniformemente dentre todas as colorações possíveis) e seja X a variável aleatória que descreve o número de cliques K_4 monocromáticas. Iremos calcular $E[X]$. Seja A_i o evento que indica que a i-ésima clique K_4 (dentre as C_n^4) é monocromática. Como cada clique possui 6 arestas, $Pr[A_i] = 2^{-5}$. Temos, então,

$$E[X] = E\left[\sum_{i=1}^{C_n^4} A_i\right] \leq \sum_{i=1}^{C_n^4} E[A_i] = \binom{n}{4} \times 2^{-5}.$$

Logo, pelo método da esperança, existe pelo menos uma coloração com $C_n^4 2^{-5}$ ou menos cliques K_4 monocromáticas.

(b) O algoritmo randomizado consistirá em simplesmente escolher, aleatória e uniformemente, uma das colorações de K_n e contar, dentre as $C_n^4 = \Theta(n^4)$ cliques, quantas são monocromáticas.