Auxiliar 10

Profesor: Pablo Barceló, Gonzalo Navarro
Auxiliar: Matilde Rivas, Bernardo Subercaseaux

P1. Conflictos en tablas de hash

Suponga que tenemos una tabla de hash con n registros (*slots*), y que la utilizamos para insertar n llaves resolviendo los conflictos por encadenamiento. Cada llave tiene la misma probabilidad de ser insertada en cada registro. Sea M el número máximo de llaves en algún registro después de que todas las llaves han sido insertadas.

(a) ¿Cuál es la probabilidad Q_k de que exactamente k llaves sean insertadas en un registro dado?

(b) Demuestre que $Q_k < (e/k)^k$. (Recuerde que $k! \geq (k/e)^k$ debido a la aproximación de Stirling).

(c) Sea P_k la probabilidad de que $M = k$, es decir, de que el registro que contiene más llaves contenga exactamente k llaves. Demuestre que $P_k \leq nQ_k$.

(d) Demuestre que existe una constante $c > 1$ tal que $Q_{k_0} < 1/n^3$, donde $k_0 = c \ln n / \ln \ln n$. Concluya que $P_k < 1/n^2$ para todo $k \geq k_0$.

(e) Explique por qué se cumple que el valor esperado $E(M)$ de M está acotado por la siguiente expresión

$$Pr(M > k_0) \cdot n + Pr(M \leq k_0) \cdot k_0.$$

Concluya que $E(M) = O(\ln n / \ln \ln n)$.

Hint: Exprese $E(M)$ como $\sum_{i=1}^{k_0} i \cdot Pr(M = i) + \sum_{i=k_0+1}^{n} i \cdot Pr(M = i)$.

Solución:

i. Primero debemos elegir las k llaves, para las cuales tenemos $\binom{n}{k}$ opciones. Esas k llaves deben caer en el registro i, lo que ocurre con probabilidad $(1/n)^k$. Las demás deben caer en cualquier otro registro, lo que ocurre con probabilidad $(n-1/n)^{n-k}$. Concluímos que $Q_k = \binom{n}{k}(n-1)^{n-k}/n^n$.

ii. Podemos escribir Q_k como $\binom{n}{k}\frac{1}{n^k}(1-\frac{1}{n})^{n-k}$. Dado que $(1-\frac{1}{n})^{n-k} < 1$ podemos concluir que $Q_k < \binom{n}{k}\frac{1}{n^k}$. Por tanto, $Q_k < \frac{n^k}{k!}\frac{1}{n^k}$. Pero por aproximación de Stirling tenemos entonces que $Q_k < (e/k)^k$.

iii. Sea E_i el evento de que el slot i contenga exactamente k llaves y sea el slot con más llaves. Por tanto, $Pr(E_i)$ está acotada por la probabilidad de que el slot i tenga exactamente k llaves, lo que corresponde a Q_k. Ahora, $P_k = Pr(\bigcup_{i=1}^{n} E_i) \leq \sum_{i=1}^{n} Pr(E_i) \leq nQ_k$.

iv. Necesitamos un $c > 1$ tal que $(e/k_0)^{k_0} < 1/n^3$. Es decir, se debe cumplir que:

$$\begin{aligned}
k_0(\ln e - \ln k_0) &< -3\ln n \\
\frac{c\ln n}{\ln\ln n}(\ln e - \ln\frac{c\ln n}{\ln\ln n}) &< -3\ln n \\
\frac{c\ln n}{\ln\ln n}(\ln e - \ln c - \ln\ln n + \ln\ln\ln n) &< -3\ln n \\
\frac{c}{\ln\ln n}(\ln c - \ln e + \ln\ln n - \ln\ln\ln n) &> 3.
\end{aligned}$$

Note que para lograr esto basta que $c > e$ y que $c > 3\ln\ln n/(\ln\ln n - \ln\ln\ln n)$. Es decir, necesitamos que

$$c > \frac{3}{1 - \frac{\ln\ln\ln n}{\ln\ln n}}.$$

La segunda expresión en el denominador tiende a 0 y nunca supera el valor 1/2 (esto se puede ver estudiando el máximo de la función $f(x) = \frac{\log(x)}{x}$. Por tanto, basta con que $c > 6$ para que esto se cumpla.

Esto nos dice que $Q_{k_0} < 1/n^3$. Dado que Q_k es una función decreciente para $k > e$, concluímos que $Q_k \leq 1/n^3$ para todo $k \geq k_0$ (ya que $k_0 \geq e$). Por tanto, $P_k \leq nQ_k < 1/n^2$ para todo $k \geq k_0$.

v. Note que

$$\begin{aligned}
E(M) &= \sum_{i=1}^{k_0} i \cdot Pr(M = i) + \sum_{i=k_0+1}^{n} i \cdot Pr(M = i) \\
&\leq k_0 \cdot \sum_{i=1}^{k_0} Pr(M = i) + n \cdot \sum_{i=k_0+1}^{n} Pr(M = i) \\
&= k_0 \cdot Pr(M \leq k_0) + n \cdot Pr(M > k_0) \\
&\leq c\ln n/\ln\ln n + n(n - k_0)1/n^2 \quad \text{(por parte (d))} \\
&\leq 1 + c\ln n/\ln\ln n.
\end{aligned}$$

Esto último es claramente $O(\ln n/\ln\ln n)$.

P2. Ski rental

Recuerde el problema del arriendo de skies visto en clases. En tal problema el costo de comprar skies es de \$b y el de arrendar es de \$a, de tal forma que $a << b$ y a divide a b. Considere un algoritmo en línea (no determinístico) que realiza lo siguiente: Se lanza una moneda. Si esta sale cara se decide comprar skies el día $b/a - 1$, si sale sello se decide comprar el día

$b/a+1$. Analice el radio entre el costo esperado que tiene este procedimiento y el algoritmo *offline* óptimo. ¿Cómo se compara esto con el del algoritmo en línea determinístico que vimos en clases?

Solución: Se puede realizar la siguiente tabla:

número de días	costo óptimo	costo esperado algoritmo en línea
$d < b/a - 1$	da	da
$d = b/a - 1$	$b - a$	$1/2(b - 2a + b) + 1/2(b - a)$
$d = b/a$	b	$1/2(b - 2a + b) + b/2$
$d \geq b/a + 1$	b	$1/2(b - 2a + b) + 1/2(b + b)$

Puede observarse que la fila que maximiza el radio entre ambos costos es la última. En tal caso el radio vale $2 - a/b$. Esto es igual al radio del algoritmo que obtuvimos en clases.

P3. Verificando el producto de matrices

Dadas 3 matrices cuadradas A, B y Cde tamaño $n \times n$. Muestre un algoritmo probabilístico que tome tiempo $O(n^2)$ en determinar si $AB = C$.

Solución: Elegimos un vector aleatorio r uniformemente en $\{0,1\}^n$. Respondemos `true` si $A(Br) = Cr$ y `false` en otro caso. Se cumple que cuando es cierto que $AB = C$ el algoritmo responde correctamente `true`, sin embargo, en caso que $AB \neq C$ puede que el algoritmo de todas formas responda `true` erróneamente lo que sucede en caso que $A(Br) = Cr$ o equivalentemente $(AB - C)r := Dr = P = 0$.

Como $D \neq 0$, existe un $d_{i,j} \neq 0$, la probabilidad de que la coordenada i de P sea 0 es:

$$\begin{aligned}
\mathbb{P}\left(p_i = 0\right) &= \mathbb{P}\left(\sum_{k=1}^{n} d_{i,k} r_k = 0\right) \\
&= \mathbb{P}\left(d_{i,j} r_j + y = 0\right) \\
&= \mathbb{P}\left(d_{i,j} r_j + y = 0 | y = 0\right) \mathbb{P}(y = 0) + \mathbb{P}\left(d_{i,j} r_j + y = 0 | y \neq 0\right) \mathbb{P}(y \neq 0) \\
&= \mathbb{P}\left(r_j = 0\right) \mathbb{P}(y = 0) + \mathbb{P}\left(r_j = 1 \wedge y = -d_{i,j}\right) (1 - \mathbb{P}(y = 0)) \\
&\leq \mathbb{P}\left(r_j = 0\right) \mathbb{P}(y = 0) + \mathbb{P}\left(r_j = 1\right) (1 - \mathbb{P}(y = 0)) \\
&= \frac{1}{2} \mathbb{P}(y = 0) + \frac{1}{2}(1 - \mathbb{P}(y = 0)) \\
&= \frac{1}{2}
\end{aligned}$$

, luego la probabilidad de error del algoritmo en este caso es:

$$\mathbb{P}(P = 0) \leq \mathbb{P}(p_i = 0) \leq \frac{1}{2}$$

Note que al probar con k vectores obtenemos una probabilidad de $1/2^k$, y que si los elementos del vector aleatorio, en lugar de elegidos sobre $0, 1$, fuesen elegidos sobre $0, 1, 2, ..., M$, la probabilidad sería de $1/M^k$.