UNIVERSIDAD DE CHILE
ESCUELA DE INGENIERIA
DEPARTAMENTO DE CIENCIAS DE LA COMPUTACION

Diseño y Análisis de Algoritmos – CC4102
Examen - Semestre Otoño 2015

1. (1,5pts) Recuerde que una fórmula en 3-CNF es una expresión de la lógica proposicional de la forma:

$$(\ell_1^1 \vee \ell_1^2 \vee \ell_1^3) \wedge (\ell_2^1 \vee \ell_2^2 \vee \ell_2^3) \wedge \ldots \wedge (\ell_n^1 \vee \ell_n^2 \vee \ell_n^3),$$

donde cada ℓ_i^j ($1 \leq i \leq n$ y $1 \leq j \leq 3$) es un *literal*, es decir, una variable proposicional o su negación. Cada expresión en esta fórmula de la forma $(\ell_i^1 \vee \ell_i^2 \vee \ell_i^3)$ se denomina una *cláusula*. Para propósitos de este problema asumimos que cada fórmula en 3-CNF cumple lo siguiente: (a) cada cláusula contiene exactamente tres literales distintos, y (b) ninguna cláusula contiene una variable proposicional y su negación.

Definimos MAX-3-CNF-SAT como el problema de encontrar una valuación para las variables proposicionales de una fórmula en 3-CNF que maximice el número de cláusulas que se satisfacen. Considere un algoritmo probabilista para este problema que simplemente asigna a cada variable proposicional un 1 con probabilidad 1/2 y un 0 con probabilidad 1/2. Demuestre que el número esperado de cláusulas satisfechas después de correr este algoritmo en una fórmula con n cláusulas es $7n/8$.

Respuesta: La probabilidad de que este proceso no satisfazga una cláusula cualquiera en la fórmula corresponde a la probabilidad de que todos sus literales reciban el valor 0. Debido a nuestra suposición en la forma de las fórmulas en 3-CNF podemos asumir por independencia que este es el producto de la probabilidad de que cada uno de sus tres literales reciba el valor 0, es decir, 1/8. Por tanto, la probabilidad de que una cláusula sí sea satisfecha por este proceso es 7/8.

Sea Y una variable aleatoria que cuenta el número de cláusulas que son satisfechas por una valuación. Por tanto, $Y = Y_1 + \cdots + Y_n$, donde Y_j es una variable indicatoria que vale 1 si y solo si la j-esima cláusula es satisfecha por la valuación. Por tanto, $E(Y_j) = 7/8$. Concluímos que $E(Y) = 7n/8$, que es lo que queríamos demostrar.

2. Dados un universo U con n elementos y una colección $\mathcal{S} = \{S_1, \ldots, S_m\}$ de subconjuntos de U que contienen a cada elemento de U, queremos encontrar un subconjunto $\mathcal{S}'$ de $\mathcal{S}$ que *cubra* a todos los elementos de U, es decir, $\bigcup_{S \in \mathcal{S}'} S = U$, y que contenga la menor cantidad de conjuntos posibles.

A continuación presentamos un algoritmo aproximado para resolver este problema: Elija el conjunto S que contenga la mayor cantidad de elementos de U. Elimine los elementos en S de $\mathcal{S}$. Repita sobre la colección obtenida hasta obtener una colección de elementos en $\mathcal{S}$ que cubra a U.

Demuestre que si la solución óptima contiene k conjuntos en $\mathcal{S}$, entonces la solución dada por el algoritmo aproximado contiene a lo más $k \log n$ tales conjuntos.

Hint: Demuestre que en cada iteración se borran al menos una fracción de $1/k$ de los elementos que quedan en $\mathcal{S}$.

Solución: Dado que U se cubre con k conjuntos en $\mathcal{S}$, debe existir uno de esos conjuntos que contenga al menos n/k elementos en U. Por tanto, luego de la primera iteración quedan a lo más $n(1 - 1/k)$

elementos en $\mathcal{S}$. Para la segunda ietración podemos asumir también que uno de los conjuntos en la solución óptima cubre al menos una fracción de $1/k$ de los elementos restantes. Por tanto, el máximo número de elementos que pueden quedar en $\mathcal{S}$ después de esa iteración es $n(1-1/k)^2$. En general, después de la i-ésima iteración tenemos que en $\mathcal{S}$ quedan a lo más $n(1-1/k)^i$ elementos. Por tanto, después de $i = k \ln n$ iteraciones hay a lo más $n(1-1/k)^{k \ln n}$ elementos en $\mathcal{S}$. Este valor es estrictamente menor que $n(1/e)^{k \ln n} \leq 1$.

3. (2,5pts) 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) (0,3pts) ¿Cuál es la probabilidad Q_k de que exactamente k llaves sean insertadas en un registro dado?

   b) (0,4pts) Demuestre que $Q_k < (e/k)^k$. (Recuerde que $k! \geq (k/e)^k$ debido a la aproximación de Stirling).

   c) (0,3pts) 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) (0,7pts) 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) (0,8pts) 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:

a) 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 $Q_k = \binom{n}{k}(n-1)^{n-k}/n^n$.

b) 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 $Q_k < (e/k)^k$.

c) 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, $P_k = Pr(\bigcup_{i=1}^{n} Pr(E_i)) \leq \sum_{i=1}^{n} Pr(E_i) \leq nQ_k$.

d) 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. 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 $k \geq k_0$.

e) Note que

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

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