Solutions to Math 416 Homework due November 5, Chapter 11

November 5, 2004

1 Problem 11-2

a. Fix an arbitrary slot, s. The probability that exactly k keys hash to s is the probability that a particular set of k keys is exactly the set that hashes to s times the number of sets of size k. The probability that the keys in the set hash to s is $(1/n)^k$; the probability that the keys outside the set do not hash to s is $(1-1/n)^{n-k}$, and there are $\binom{n}{k}$ sets. Thus

$$Q_k = \left(\frac{1}{n}\right)\left(1-\frac{1}{n}\right)^{n-k}\binom{n}{k}.$$

b. If the slot containing the most keys contains k keys, then, in particular, *some* slot contains exactly k keys. That is, slot 1 contains k keys or slot 2 contains k keys or ..., and each event in this n-wise disjunction has probability Q_k. Thus the probability P_k that the fullest slot has k keys is at most nQ_k.

c. We will use a slightly weaker approximation to $n!$, namely, $n! = \Theta\left(\sqrt{n}(n/e)^n\right)$. Then

$$\begin{aligned}
\binom{n}{k} &= \frac{n!}{k!(n-k)!} \\
&= \Theta\left(\sqrt{\frac{n}{k(n-k)}}\frac{(n/e)^n}{(k/e)^k((n-k)/e)^{n-k}}\right) \\
&= \Theta\left(\sqrt{\frac{n}{k(n-k)}}\frac{n^n}{k^k(n-k)^{n-k}}\right) \\
&\leq O\left(\frac{n^n}{k^k(n-k)^{n-k}}\right),
\end{aligned}$$

since $k \geq n/2$ or $n-k \geq n/2$. Thus

$$\begin{aligned}
Q_k &\leq O\left(\frac{1}{n^k}\left(\frac{n-1}{n}\right)^{n-k}\frac{n^n}{k^k(n-k)^{n-k}}\right) \\
&\leq O\left(\frac{1}{k^k}\left(\frac{n-1}{n-k}\right)^{n-k}\right) \\
&\leq O\left(\frac{1}{k^k}\left(1+\frac{k-1}{n-k}\right)^{n-k}\right) \\
&\leq O\left(\frac{1}{k^k}\left(e^{\frac{k-1}{n-k}}\right)^{n-k}\right) \\
&\leq O\left(\frac{1}{k^k}e^{k-1}\right),
\end{aligned}$$

using the fact that $1+x \leq e^x$. Thus $Q_k \leq O(e^k/k^k)$.

d. We want a constant $c > 1$ such that $Q_{k_0} < 1/n^3$ for $k_0 = c \lg n / \lg\lg n$.

Taking natural logs, we have $\ln Q_k < k(1 - \ln k) + O(1) = -k(\ln k - 1) + O(1)$. For large enough k, we have $\ln Q_k < -k\ln(k)/2$, and we need this to be at most $\ln(1/n^3) = -3\ln n$, or $k\ln(k) \geq 6\ln n$.

Put $k = \frac{12\ln(n)}{\ln\ln(n)} = \frac{12\lg(n)}{\lg\ln(n)} = \frac{12\lg(n)}{\lg\lg(n) - \lg\lg(e)}$; thus $k \leq O\left(\frac{\lg(n)}{\lg\lg(n)}\right)$. Then $\ln(k) = \ln(12) + \ln\ln(n) - \ln\ln\ln(n)$, and $k\ln(k) \geq (\ln(12) + \ln\ln(n) - \ln\ln\ln(n))\frac{12\ln(n)}{\ln\ln(n)} \geq 6\ln(n)$ for sufficiently large n, as desired. (To handle smaller n, just increase c.)

We also assume that $k_0 \geq e$, so that e^k/k^k is strictly decreasing as $k \geq k_0$ increases. It follows that, for all $k > \frac{12\ln(n)}{\ln\ln(n)}$, we have $Q_k < 1/n^3$. Thus $P_k \leq nQ_k < 1/n^2$ for such k.

e (and conclusion). We have

$$\begin{aligned} E[M] &= \sum_k kP_k \\ &= \sum_{k_0 < k \leq n} kP_k + \sum_{k \leq k_0} kP_k \\ &\leq \sum_{k_0 < k \leq n} n(1/n^2) + \sum_{k \leq k_0} k_0 P_k \\ &\leq 1 + k_0 \sum_{k \leq k_0} P_k \\ &\leq 1 + k_0. \end{aligned}$$

2 Problem 11-4

a. Suppose $\mathcal{H}$ is 2-univeral and suppose $i_1 \neq i_2$. Let h be a random element of $\mathcal{H}$. Then

$$\Pr(h(i_1) = h(i_2)) = \sum_t \Pr(h(i_1) = h(i_2) = t) = \sum_t (1/m^2) = 1/m.$$

b (second printing, 2001), and c (third printing, 2002). Suppose $h_{a,b}(x) = \left(\sum_{j=0}^{n-1} a_j x_j + b\right) \mod p$, suppose $x \neq y$, and suppose $s, t \in B$. Then $h(x) = s$ and $h(y) = t$ iff

$$\begin{pmatrix} x_0 & x_1 & x_2 & \cdots & x_{n-1} & 1 \\ y_0 & y_1 & y_2 & \cdots & y_{n-1} & 1 \end{pmatrix} \cdot \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_{n-1} \\ b \end{pmatrix} = \begin{pmatrix} s \\ t \end{pmatrix}.$$

Since $x \neq y$, we have $x_i \neq y_i$ for some i, or

$$\begin{pmatrix} x_i & 1 \\ y_i & 1 \end{pmatrix} \cdot \begin{pmatrix} a_i \\ b \end{pmatrix} = \begin{pmatrix} s - \sum_{j \neq i} a_j x_j \\ s - \sum_{j \neq i} a_j y_j \end{pmatrix}.$$

We claim that the statement holds even conditioning on all the values of a_j's *except* a_i. Since the matrix

$$\begin{pmatrix} x_i & 1 \\ y_i & 1 \end{pmatrix}$$

is invertible, there is exactly one setting (out of m^2) of a_i and b that works. Thus $\mathcal{H}$ is 2-universal.

b (third printing, 2002). We now define $h_a(x) = \left(\sum_{j=0}^{n-1} a_j x_j\right) \mod p$. We need to show that the family of h_a's is universal, but not 2-universal. First, suppose $x \neq y$. Then $x_i \neq y_i$ for some i. Then $\Pr(h_a(x) =$

$h_a(y)) = \Pr(h_a(x-y)=0) = \Pr(a_i(x_i - y_i) = \sum_{j\neq i} a_j(x_j - y_j)) = \Pr(a_i = (x_i - y_i)^{-1} \sum_{j\neq i} a_j(x_j - y_j))$. Conditioned on any setting of a_j's for $j \neq i$, this probability is $1/m$. Thus the family is universal.

To show that it is not 2-universal, suppose $x = 0$ and $t \neq 0$. Then $\Pr(h_a(x) = t \;\&\; h_a(y) = s) \leq \Pr(h_a(x) = t) = 0 \neq 1/m^2$.

c (second printing) and d (third printing).

This piece is not for credit. But please think about it!

What we really want is $\Pr_{a,b}(h(m') = t' | h(m) = t) = \Pr(h(m) = t \;\&\; h(m') = t') / \Pr(h(m) = t)$. (It may take some thought to see this.) Because the family is 2-universal, the numerator is $1/p^2$ and the denominator is $1/p \neq 0$. So the probability is $1/p$.