ICS 261—Fall 2002
Data Structures
Suggested Homework 1—answer to #7

7. Part (b) is easier, so let's do that first.

   b) In the notation of slide 2-20, assume we have 1-byte keys, and let $k_1 =< 0 >$ and $k_2 =< 1 >$. Then note that regardless of the choice of m and a, $h_a(k_1)$ is always 0. Thus the hash functions do not meet the stronger condition.

   a) This part was considerably harder. Here are two ways it could be done.

   Note that $\sum_{i=0}^{m-1} n_i$ is always n; thus $\mathsf{E}\left[\sum_{i=0}^{m-1} n_i\right] = n$.

   To compute $\mathsf{E}\left[\sum_{i=0}^{m-1} n_i^2\right]$ we can relate it to the total number of collisions that take place with chaining. Since n_i is the number of keys going to bucket i, the number of collisions in bucket i is $0 + 1 + 2 + \cdots + (n_i - 1) = \binom{n_i}{2}$, so, letting c be the expectation of the total number of collisions, we have

$$c = \mathsf{E}\left[\sum_{i=0}^{m-1} \binom{n_i}{2}\right]. \tag{1}$$

   From slide 2-19 we know that the expected number of collisions involving any key is $(n-1)/m$. Adding this together for all of the n keys gives a sum of $n(n-1)/m$. However, for any x and y, we have counted any collision between x and y twice in this sum. Thus the expected total number of collisions is

$$c = \frac{n(n-1)}{2m}. \tag{2}$$

   Now note that since $k^2 = 2\binom{k}{2} + k$, we have

$$\mathsf{E}\left[\sum_{i=0}^{m-1} n_i^2\right] = \mathsf{E}\left[\sum_{i=0}^{m-1} 2\binom{n_i}{2} + n_i\right] = 2\,\mathsf{E}\left[\sum_{i=0}^{m-1} \binom{n_i}{2}\right] + \mathsf{E}\left[\sum_{i=0}^{m-1} n_i\right] = 2c + n. \tag{3}$$

   where we used (1) in the last step. Using (2), this implies

$$\mathsf{E}\left[\sum_{i=0}^{m-1} n_i^2\right] = 2\frac{n(n-1)}{2m} + n = n\left(1 + \frac{n-1}{m}\right).$$

   Note that this analysis is very similar to the analysis on page 248 of our text (second edition).

   Here is another way to compute $\mathsf{E}\left[\sum_{i=0}^{m-1} n_i^2\right]$. Define

$$X_{si} = \begin{cases} 1 & \text{if key } s \text{ goes to slot } i \\ 0 & \text{otherwise.} \end{cases}$$

Note that $\mathsf{E}[X_{si}]$ is equal to the probability that key s goes to slot i, that $n_i = \sum_{s=1}^{n} X_{si}$, and that

$$n_i^2 = \left(\sum_{s=1}^{n} X_{si}\right)\left(\sum_{r=1}^{n} X_{ri}\right) = \sum_{s=1}^{n}\sum_{r=1}^{n} X_{si}X_{ri}.$$

Hence

$$\mathsf{E}\left[\sum_{i=0}^{m-1} n_i^2\right] = \mathsf{E}\left[\sum_{i=0}^{m-1}\sum_{s=1}^{n}\sum_{r=1}^{n} X_{si}X_{ri}\right] = \sum_{i=0}^{m-1}\sum_{s=1}^{n}\sum_{r=1}^{n} \mathsf{E}[X_{si}X_{ri}]. \tag{4}$$

We now compute the value of $\mathsf{E}[X_{si}X_{ri}]$. For any given s and i, let A be the event that key s goes to slot i, and $\overline{A}$ be the event that it does not. Note that A holds if and only if $X_{si} = 1$, so $\Pr\{A\} = \mathsf{E}[X_{si}]$. Then

$$\mathsf{E}[X_{si}X_{ri}] = \Pr\{A\}\,\mathsf{E}[X_{si}X_{ri} \mid A] + \Pr\{\overline{A}\}\,\mathsf{E}\left[X_{si}X_{ri} \mid \overline{A}\right].$$

When A does not hold, $X_{si} = 0$ so $X_{ri}X_{si} = 0$, so this simplifies to

$$\mathsf{E}[X_{si}X_{ri}] = \Pr\{A\}\,\mathsf{E}[X_{si}X_{ri} \mid A].$$

When A does hold, $X_{si} = 1$ and by the properties of universal hashing $\Pr\{X_{ri} = 1\} = 1/m$ if $r \neq s$, and of course $X_{ri}X_{si} = X_{ri}$ when $r = s$. Thus this simplifies to

$$\mathsf{E}[X_{si}X_{ri}] = \begin{cases} \mathsf{E}[X_{si}] & \text{if } s = r \\ \mathsf{E}[X_{si}]/m & \text{if } s \neq r. \end{cases} \tag{5}$$

Substituting (5) into (4) gives

$$\begin{aligned}
\mathsf{E}\left[\sum_{i=0}^{m-1} n_i^2\right] &= \sum_{i=0}^{m-1}\sum_{s=1}^{n}\sum_{r=1}^{n} \mathsf{E}[X_{si}] \times \begin{cases} 1 & \text{if } s = r \\ 1/m & \text{if } s \neq r \end{cases} \\
&= \sum_{i=0}^{m-1}\sum_{s=1}^{n} \left(1 + \frac{n-1}{m}\right) \mathsf{E}[X_{si}] \\
&= \sum_{i=0}^{m-1} \left(1 + \frac{n-1}{m}\right) \sum_{s=1}^{n} \mathsf{E}[X_{si}] \\
&= \sum_{i=0}^{m-1} \left(1 + \frac{n-1}{m}\right) \mathsf{E}[n_i] \\
&= \left(1 + \frac{n-1}{m}\right) \sum_{i=0}^{m-1} \mathsf{E}[n_i] \\
&= \left(1 + \frac{n-1}{m}\right) n.
\end{aligned}$$