Solutions for CS174 Homework 3

1. (a) Note that here we assume all balls and bins are distinguishable. For each ball, there are n possibilities. So for m balls, there are n^m distributions.

   (b) The probability that there are k balls in bin 1 is

$$\binom{m}{k} \frac{1}{n}^k \left(1 - \frac{1}{n}\right)^{m-k}$$

   and we simply multiply this probability by the number of distributions n^m to get

$$\binom{m}{k} (n-1)^{m-k}$$

   distributions.

   (c) We need inclusion-exclusion here. Let E_i be the property that bin i is empty. Then $\Pr[E_i] = (1-1/n)^m$, $\Pr[E_i \wedge E_j] = (1-2/n)^m$ etc. and each S_k in the inclusion/exclusion formula is

$$S_k = \binom{n}{k} (1 - k/n)^m$$

   We want exactly two of these events to be true, so the inclusion/exclusion formula is

$$\Pr[\text{two empty bins}] = S_2 - \binom{3}{2} S_3 + \cdots + (-1)^k \binom{k}{2} S_k + \cdots + (-1)^n \binom{n}{2} S_n$$

   and substituting we have

$$\Pr[\text{two empty bins}] = \sum_{k=2}^{n} (-1)^k \binom{k}{2} \binom{n}{k} (1 - k/n)^m$$

   which rewrites to

$$\Pr[\text{two empty bins}] = \sum_{k=2}^{n} (-1)^k \binom{n}{2, k-2, n-k} (1 - k/n)^m$$

   using a familar approximation, we can write $(1 - k/n)^m \approx e^{-mk/n}$ or $e^{-\lambda k}$ where $\lambda = m/n$, and the above becomes:

$$\Pr[\text{two empty bins}] = \sum_{k=2}^{n} (-1)^k \binom{n}{2, k-2, n-k} e^{-\lambda k}$$

which is now part of the expansion of a trinomial (you can prove this by applying binomial expansion twice). The terms correspond to the coefficient of x^2 in the expansion of

$$(-xe^{-\lambda} + 1 - e^{-\lambda})^n$$

which we can write directly as:

$$\binom{n}{2} e^{-2\lambda}(1 - e^{-\lambda})^{n-2}$$

finally, we multiply by n^m to count the number of distributions with this property:

$$n^m \binom{n}{2} e^{-2\lambda}(1 - e^{-\lambda})^{n-2}$$

2. Assume the hash table has size k. From the birthday paradox computation,

$$\Pr[\text{all k bins have less than 1 element}] \le \exp(-n(n-1)/2k).$$

So when $\exp(-n(n-1)/2k) < 1/2$, the hash table has probability at least a half for a collision. So $k < \frac{n(n-1)}{\ln 2}$.

3. Let X be the number of heads got after 1000 tosses of a fair coin. So $\mathrm{E}[X] = 500, \mathrm{Var}[X] = 1000 * 1/2 * 1/2 = 250$.

   (a) Using Markov bound, $\Pr[X > 800] < \frac{E[X]}{800} = \frac{5}{8} = 0.625$.

   (b) Because the probability that there are more than 800 heads equals to the probability that there are more than 800 tails, so $\Pr[X > 800] = \Pr[X < 200]$. Using Chebyshev bound,

$$\Pr[X > 800] = \frac{1}{2}\Pr[|X - 500| > 300] < \frac{1}{2} \cdot \frac{250}{300^2} = \frac{1}{72} \doteq 0.014.$$

4. (a) From lecture 5, $\Pr[X_i \ge 20] \le (\frac{me}{nk})^k \cdot \frac{1}{1-\frac{me}{nk}} = 1.8 * 10^{-5}$.

   $\mathrm{E}[X_i] = \mathrm{Var}[X_i] = \lambda = 4$.

   Using Markov bound, $\Pr[X_i \ge 20] \le \frac{\mathrm{E}[X_i]}{20} = 0.2$.

   Using Chebyshev bound, $\Pr[X_i \ge 20] \le \Pr[|X_i - 4| \ge 16] \le \frac{\mathrm{Var}[X_i]}{16^2} = 0.0156$.

   (b) $\Pr[X_i \ge k] \le (\frac{me}{nk})^k \cdot \frac{1}{1-\frac{me}{nk}} = O(k^{-k})$.

   Using Markov bound, $\Pr[X_i \ge k] \le \frac{\mathrm{E}[X_i]}{k} = O(k^{-1})$.

   Using Chebyshev bound, $\Pr[X_i \ge k] \le \Pr[|X_i - \mathrm{E}[X_i]| \ge k - \mathrm{E}[X_i]] \le \frac{\mathrm{Var}[X_i]}{(k-\mathrm{E}[X_i])^2} = O(k^{-2})$.