CS174 **Midterm 1 Solutions** **Spring 99**
J. Canny **Mar 4**

1. Suppose a positive integer k is chosen uniformly at random from $\{1, \ldots, 1200\}$. Then the following are random variables:

$$\begin{array}{ll} X = 1 \text{ if } k \text{ is a multiple of 4,} & 0 \text{ otherwise} \\ Y = 1 \text{ if } k \text{ is a multiple of 5,} & 0 \text{ otherwise} \\ Z = 1 \text{ if } k \text{ is a multiple of 6,} & 0 \text{ otherwise} \end{array}$$

   (a) What is $E[X]$?
   Since X is an indicator variable $E[X] = Pr[X = 1] = Pr[k$ **divisible by** $4] = 1/4$

   (b) What is $E[Y + Z]$?
   By linearity of expectation $E[Y + Z] = E[Y] + E[Z]$ **which is** $1/5 + 1/6 = 11/30$

   (c) What is $E[XZ]$?
   XZ **is a** $\{0, 1\}$**-valued variable so** $E[XZ] = Pr[XZ = 1]$**. Now** X **and** Z **are not independent, but** $Pr[XZ = 1] = Pr[X = 1 \wedge Z = 1] = Pr[k$ **divisible by** $12] = 1/12$

2. Let a biased coin have $Pr[Heads] = p$, where p is not necessarily 0.5. The coin is tossed repeatedly, until a total of 3 heads (not necessarily consecutive) have appeared. Let X be a random variable which is the number of tosses up to and including the 3rd head.

   (a) What is $E[X]$?
   First, write $X = X_1 + X_2 + X_3$ **where the** X_i **are random variables representing the number of tosses after the** $(i-1)^{st}$ **head up to and including the** i^{th} **head. Then each** X_i **is a geometric random variable with parameter** $1/p$**. For each** X_i**,** $E[X_i] = 1/p$**, so** $E[X] = 3E[X_i] = 3/p$

   (b) What is $\sigma^2[X]$?
   Write $X = X_1 + X_2 + X_3$ **where the** X_i **are geometric random variables as before. Note that the** X_i **are independent. So the variance of** X **is the sum of the variances of the** X_i**. So** $\sigma^2[X] = 3\sigma^2[X_i] = 3(1-p)/p^2$

   (c) What is the distribution of X? i.e. give $Pr[X = k]$ as a function of k.
   We know that the k **tosses must consist of exactly 3 heads and the rest tails. For every such sequence, the probability is** $p^3(1-p)^{k-3}$**. To count the number of sequences, notice that the order is arbitrary, except that the last toss is a head. The other two heads can occur anywhere in the other** $k-1$ **positions. So the overall probability is**
$$\binom{k-1}{2} p^3 (1-p)^{k-3}$$

3. Suppose we run the proposal algorithm (for stable marriages) on m males and n females. The algorithm is the same as before, namely:

   Males Each unmarried male proposes to the highest-ranked female on his preference list who has not turned him down before.

   Females Each female accepts a proposal if she is not married, or if the proposer ranks higher on her preference list than her current spouse.

   Suppose further that we implement proposals in rounds. In each round, all the unmarried males propose to their current favorite. Each female accepts the highest ranked proposal in that round, unless she has a spouse who ranks higher. This is equivalent to the usual algorithm. Assume males and females have random preference lists.

   (a) How large should m be (in terms of n) to be confident every female receives a proposal in the first round?

   This is an instance of coupon collecting. We need m to be at least $n \ln n + \Omega(n)$, e.g. $n \ln n + 5n$

   (b) If $m = n$, what is the expected number of unmarried males (or females) after the first round?

   Think of this as placing n balls randomly in n bins, and asking for the expected number of empty bins. The probability that a given bin is empty is $1/e$, so the expected number of empty bins is n/e

   (c) If $m \neq n$, what is the expected number of females who receive more than one proposal in the first round? Simplify your result as much as possible, assuming large n and m.

   Think of this as placing m balls into n bins, and asking how many bins have > 1 ball. Let X be a random variable which is the number of bins with $> 1 ball$. Then $E[X] = nPr[X_i > 1]$ where X_i is the number of balls in bin i. And $Pr[X_i > 1] = 1 - Pr[X_i = 0] - Pr[X_i = 1]$.

   The probability that bin i contains k balls is

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

   which simplifies (as we did in class) to a Poisson probability $\lambda^k e^{-\lambda}/k!$ with $\lambda = m/n$. So $Pr[X_i = 0] \approx e^{-m/n}/k!$ and $Pr[X_i = 1] \approx (m/n)e^{-m/n}/k!$. Then $Pr[X_i > 1] \approx 1 - (1 + m/n)e^{-m/n}/k!$ and

$$E[X] \approx n - (m + n)e^{-m/n}/k!$$

4. The following examples define random variables and probabilities. In each case, suggest a tail bound (Markov, Chebyshev or Chernoff) that could be used to compute a bound on the probability. Choose the method that is applicable and gives the best bound. Then compute the bound. In each case, assume $\mu = E[X]$:

   (a) Let X have the uniform distribution on $\{1, \ldots, n\}$. What is $Pr[X > 1.5\mu]$?

   You can't use Chernoff here, so the choice is between Markov and Chebyshev. Markov is usually weaker, but Chebyshev is two sided and may not be as strong for this one-sided bound. So its best to try both. For Markov $Pr[X > k] \leq E[X]/k = \mu/(1.5\mu) = 2/3$.

   For Chebyshev, the mean is $(n+1)/2$, and the variance is $E[X^2] - E[X]^2$. Now

   $$E[X^2] = \frac{1}{n}\sum_{i=1}^{n} i^2 \approx n^2/3$$

   and $\sigma^2[X] \approx n^2/3 - n^2/4 \approx n^2/12$ and $\sigma \approx n/2\sqrt{3}$. For Chebyshev, we have $Pr[|X - \mu| > 0.5\mu]$ where $t\sigma = 0.5\mu = n/4$. Solving for t we get $t = \sqrt{3}/2$ but this is < 1 so

   $$Pr[X > 1.5\mu] < 1/t^2 = 4/3$$

   which is greater than 1, and not a probability bound. So Markov is better in this case.

   (b) Let X have the binomial distribution with parameters $p = 0.1$ and $n = 1000$. What is $Pr[X > 4\mu]$?

   Here Chernoff applies, and n is large, so Chernoff will be a much better bound than Markov or Chebyshev. $\mu = np = 100$ and $\delta = 3$. The upper tail bound is

   $$Pr[X > 4\mu] = \left(\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right)^{\mu} = \left(\frac{e^3}{4^4}\right)^{100} = 0.0785^{100} = 2.9 \times 10^{-111}$$

   (c) Let X have the geometric distribution with parameter $p = 0.2$. What is $Pr[X > 3\mu]$?

   We cant use Chernoff. Clearly Markov applies, and it trivially gives a bound of $1/3$, but its unlikely to be tight compared to Chebyshev, because we are several standard deviations away from the mean. First of all $\sigma^2 = (1-p)/p^2 = 20$ and $\sigma = 2\sqrt{5}$. We have $Pr[X > 3\mu] \leq Pr[|X - \mu| > 2\mu]$ so $t\sigma = 2\mu$. Solving, we get $t = \sqrt{5}$. So

   $$Pr[X > 3\mu] < 1/t^2 = 1/5$$