# Homework 6

Due: Tuesday, August 6, 4:00pm

CS 70: Discrete Mathematics and Probability Theory, Summer 2013

1. [8 points] Two faulty machines, $M_1$ and $M_2$, are repeatedly run synchronously in parallel (i.e., both machines execute one run, then both execute a second run, and so on). On each run, $M_1$ fails with probability $p_1$ and $M_2$ fails with probability $p_2$, all failure events being independent. Let the random variables $X_1$, $X_2$ denote the number of runs until the first failure of $M_1$, $M_2$ respectively; thus $X_1$, $X_2$ have geometric distributions with parameters $p_1$, $p_2$ respectively. Let $X$ denote the number of runs until the first failure of *either* machine. Show that $X$ also has a geometric distribution, with parameter $p_1 + p_2 - p_1 p_2$.

2. [10 points] Suppose you take a deck of $n$ cards and repeatedly perform the following step: take the current top card and put it back in the deck at a uniformly random position. (I.e., the probability that the card is placed in any of the $n$ possible positions in the deck — including back on top — is $1/n$.) Consider the card that starts off on the bottom of the deck. What is the expected number of steps until this card rises to the top of the deck? (Hint: Let $T$ be the number of steps until the card rises to the top. We have $T = T_n + T_{n-1} + \cdots + T_2$, where the random variable $T_i$ is the number of steps until the bottom card rises from position $i$ to position $i - 1$. Thus, for example, $T_n$ is the number of steps until the bottom card rises off the bottom of the deck, and $T_2$ is the number of steps until the bottom card rises from second position to top position. What is the distribution of $T_i$? What is the connection to coupon collecting?)

3. [20 points] Use the Poisson distribution to answer these questions.

   3a. [4 points] You type 2000 lines of code, and on each line you make a syntax error (independently) with probability $1/500$. Then you compile your code. (If you are a good programmer, you would have compiled much sooner than this!) What is the (approximate) probability that you will have exactly 3 lines with a syntax error?

   3b. [4 points] A textbook has 250 pages, and has on average one misprint per page. What is the probability that there are exactly 4 misprints on page 37?

   3c. [4 points] Suppose that on average, 20 people enter your small bookstore per day. What is the probability that exactly 7 people enter your store tomorrow?

   3d. [4 points] Suppose that on average, you catch a cold twice per year. What is the probability that you will catch *at most* one cold in 2014?

   3e. [4 points] Suppose that on average, there are 5.7 accidents per day on California state highways. What is the probability there will be *at least* 3 accidents throughout the *next two days* on California state highways?

4. [8 points] This problem will give you practice computing standard deviations.

   4a. [4 points] Let $X$ be a random variable with the following distribution: $X = 1$ with probability $1/3$, $X = 2$ with probability $1/2$, and $X = 3$ with probability $1/6$. Calculate $\sigma(X)$.

   4b. [4 points] Let $X$ have a $\text{Bin}(n, p)$ distribution and $Y$ have a $\text{Pois}(\lambda)$ distribution, and suppose $X$ and $Y$ are independent. What is the standard deviation of $5X + 6Y$?

5. [14 points] This problem will give you practice using the "standard method" to compute the variance of a sum of random variables that are not pairwise independent (so you cannot use "linearity" of variance).

   5a. [7 points] A building has $n$ floors numbered $1, 2, \ldots, n$, plus a ground floor G. At the ground floor, $m$ people get on the elevator together, and each gets off at a uniformly random one of the $n$ floors (independently of everybody else). What is the *variance* of the number of floors the elevator *does not* stop at? (In fact, the variance of the number of floors the elevator *does* stop at must be the same (do you see why?) but the former is a little easier to compute.)

   5b. [7 points] A group of three friends has $n$ books they would all like to read. Each friend (independently of the other two) picks a random permutation of the books and reads them in that order, one book per week (for $n$ consecutive weeks). Let $X$ be the number of weeks in which all three friends are reading the same book. Compute $\text{Var}(X)$.

6. [8 points] Consider any random variable $X$ with a finite sample space. Prove that $\text{E}(X^2) \geq \text{E}(X)^2$, and that $\text{E}(X^2) = \text{E}(X)^2$ iff $X$ is a constant random variable. (Hint: Think in terms of variance.)

7. [16 points] A friend tells you about a course called "Laziness in Modern Society" that requires almost no work. You hope to take this course so that you can devote all of your time to CS70. At the first lecture, the professor announces that grades will depend on only a midterm and a final. The midterm will consist of three questions, each worth 10 points, and the final will consist of four questions, also each worth 10 points. He will give an A to any student who gets at least 60 of the possible 70 points. However, speaking with the professor in office hours you hear some very disturbing news. He tells you that to save time he will be grading as follows. For each student's midterm, he'll choose a number randomly from some distribution with mean $\mu = 5$ and variance $\sigma^2 = 1$. He'll mark each of the three questions with that score. To grade the final, he'll again choose a random number from the same distribution, independent of the first number, and will mark all four questions with that score.

   7a. [4 points] What will the mean (i.e., expectation) of your total score be?

   7b. [4 points] Use Markov's Inequality to prove an upper bound on the probability of getting an A.

7c. [4 points] What will the variance of your total score be?

7d. [4 points] Use Chebyshev's Inequality to prove an upper bound on the probability of getting an A.

8. [16 points] In the aftermath of the hotly contested 2000 US Presidential Election, many people claimed that the 3407 votes cast for independent candidate Pat Buchanan in Palm Beach County were statistically highly significant, and thus of dubious validity. In this problem, we will examine this claim from a statistical viewpoint. The total percentage votes cast for each presidential candidate in the entire state of Florida were as follows:

| Gore | Bush | Buchanan | Nader | Browne | Others |
|---|---|---|---|---|---|
| 48.8% | 48.9% | 0.3% | 1.6% | 0.3% | 0.1% |

In Palm Beach County, the actual votes cast (before the recounts began) were as follows:

| Gore | Bush | Buchanan | Nader | Browne | Others | Total |
|---|---|---|---|---|---|---|
| 268945 | 152846 | 3407 | 5564 | 743 | 781 | 432286 |

To model this situation probabilistically, we need to make some assumptions. Let's model the vote cast by each voter in Palm Beach County as being random, with probabilities corresponding to the Florida percentages. Thus, e.g., a voter votes for "Gore" with probability 0.488, votes for "Others" with probability 0.001, etc. There are a total of $n = 432286$ voters, and their votes are assumed to be mutually independent. Let the random variable $B$ denote the total votes cast for Buchanan in Palm Beach County.

8a. [4 points] Compute the expectation $\mathrm{E}(B)$ and the variance $\mathrm{Var}(B)$.

8b. [4 points] Use Chebyshev's Inequality to compute an *upper bound* $b$ on the probability that Buchanan receives at least 3407 votes; i.e., find a number $b$ such that $\Pr[B \geq 3407] \leq b$. Based on this result, do you think Buchanan's vote is significant?

8c. [4 points] Now suppose that your bound $b$ in part (b) is in fact sharp, i.e., assume that $\Pr[B \geq 3407]$ is *equal to* $b$. [*In fact the true value of this probability is quite a bit smaller than* $b$.] Suppose also that all 67 counties in Florida have the same number of voters as Palm Beach County, and that all behave independently according to the same statistical model as Palm Beach County. What is the probability that in *at least one* of the counties, Buchanan receives at least 3407 votes? How would this affect your judgement as to whether the Palm Beach tally is significant?

8d. [4 points] Our model assumes that all voters behave like the fabled "swing voters," in the sense that they are undecided when they go to the polls and end up making a random decision. A more realistic model would assume that only a fraction (say, about 20%) of voters are in this category, the others having already decided. Suppose then that 80% of the voters in Palm Beach County vote deterministically according to the state-wide proportions for Florida, while the remaining 20% behave randomly as described earlier. Does your bound $b$ in part (b) increase, decrease, or remain the same under this model? Justify your answer.