STAT 421

Fall 2010 Homework 10, due Fri. Dec. 3

Note: there are four parts to this homework: text problems, in-class problems, additional problems, and a graduate problem for each of the graduate students. Problems marked by * are required for graduate students, optional for undergraduates. I encourage you to use R on problems where you are asked to compute probabilities for named distributions (see the R Intro on the web page for details on the commands).

From DeGroot and Schervish: Sec. 4.8 (pp. 235-6): #2,6,9,15*,16*; Sec. 4.10 (pp. 243-5): #22; Sec. 5.2 (pp. 250-1): #4,6,10; Sec. 5.3 (p. 255): #2,5,8*; Sec. 5.4 (pp. 262-3): #6,8,9,10*,15*

In-class, Wed., Dec. 1:

Josh B.: #4, p. 235

Josh G.: (Continuation of example from class): let X be number of trials until first success in independent Bernoulli trials with probability of success p. We showed in class that $\mathrm{E}(X) = 1/p$ by conditioning on the result of the first trial Y and using $\mathrm{E}(X) = \mathrm{E}[\mathrm{E}(X|Y)]$. Find $\mathrm{Var}(X)$ in a similar way by using $\mathrm{Var}(X) = \mathrm{E}[\mathrm{Var}(X|Y)] + \mathrm{Var}[\mathrm{E}(X|Y)]$.

Additional problems:

1. Cell phone problem: we have seen that, for any n, the expected value and variance of the number of people who get their own cell phone back are both 1. Verify this with a simulation for several different values of n.

2. In the derivation of the variance of the hypergeometric distribution on p. 253 of the text, the authors use the expression

$$\mathrm{Var}(X_1 + \ldots + X_n) = \sum_{i=1}^{n} \mathrm{Var}(X_i) + 2 \sum_{i<j} \sum \mathrm{Cov}(X_i, X_j).$$

   They first substitute the expression for $\mathrm{Var}(X_i)$ from the bottom of p. 252. They are then able to derive $\mathrm{Cov}(X_i, X_j)$ in an indirect way. For this problem:

   (a) Justify their expression for $\mathrm{Var}(X_i)$.

   (b) Derive the expression for $\mathrm{Cov}(X_i, X_j)$ directly from the definition of covariance rather than in the roundabout way they derive it.

3. *A spelunker is lost in a cave from which three tunnels exit. The first tunnel takes her to safety after 2 hours travel. The second tunnel returns her to this same room after 3 hours travel. The third door returns her to this room after 7 hours.

   (a) Assume the spelunker selects one of the tunnels randomly and continues to select one randomly each time she returns to the room (the tunnels look identical). What are the expected value and variance of the number of hours until she reaches safety?

(b) Suppose the spelunker marks a tunnel with chalk before she tries it and so chooses randomly only among those tunnels which she has not already tried. What are the expected value and variance of the number of hours until she reaches safety?

Graduate problems. Each graduate student is responsible for one problem.

1. (Rachel) You roll a die repeatedly and if you get 2,3,4,5, or 6 you add the amount to your total. Your turn ends when either a) you decide to stop in which case your score is your accumulated points, or b) you roll a 1 in which case you get a score of 0.

   (a) Find the optimal strategy if your goal is to maximize your expected score. That is, give a rule for deciding, based on your accumulated points, whether to keep rolling or to stop.

   (b) Let's change the game a little. You play against an opponent in the following way. You alternate turns. On each turn, you follow the scenario in part (a). You accumulate points gained from turn to turn. If you roll a 1 in any turn, you get a 0 for that turn but do not lose your accumulated points from previous turns. Your goal is to be the first to get to at least 100 points. Assume that your opponent has 99 points and you have $x < 100$ points and it is your turn. What is your optimal strategy (the one that maximizes your probability of winning) as a function of x? What is it if your opponent has 98 points?

2. (Micky) Here are several problems involving having children. For all these problems, we'll assume, to be completely general, that the probability of having a boy is p (in fact, p is slightly greater than .5). We'll also assume that the sexes of children in the same family are independent.

   (a) Suppose a family intends to keep having children until they have a boy. What is the expected number of children as a function of p? What is the expected number if $p = .5$?

   (b) Suppose a family intends to keep having children until they have a boy or until they have n children, whichever comes first. What are the expected number of boys, the expected number of girls, and the expected number of children in terms of n and p? What are the values if $p = .5$?

   (c) A couple decides to keep having children until they have children of both sexes. What is the expected number of children they will have as a function of p? What is the expected number if $p = 0.5$?

   (d) Suppose the couple decides that they will have children until they have a child that's the same sex as the first one. How many children can they expect to have as a function of p?

3. (Grant) The standard binomial sampling scheme assumes n independent trials with a constant probability of success. Suppose, for example, that we are given a single fair die, and that the success event consists in throwing, say, a 5. Thus, each single throw results in a success with

probability $p = 1/6$. With $n = 144$ independent trials, we would thus expect about $np = 24$ successes, and the variance of the number of successes would be equal to $np(1-p) = 20$.

(a) Suppose we are given two biased dice. Die A will show a 5 with probability $p = 1/4$, and die B shows a 5 with probability $p = 1/12$. With these two dice 144 trials are conducted, as follows. For the first 72 trials we use die A, and during the last 72 trials we use die B. Given that the average success probability across both dice is equal to $(1/4+1/12)/2 = 1/6$, we would again expect 24 successes. Is the variance of the number of successes larger than, equal to, or smaller than with the standard binomial sampling scheme?

(b) Again we are given the two dice A and B described in part (a). This time, however, we select at random one die and then throw this selected die 144 times, observing again the number of throws yielding a 5. Is the variance of the number of successes larger than, equal to, or smaller than with the standard binomial sampling scheme?

(c) Once again we are given the two dice A and B. This time, however, in each of 144 trials we start by choosing at random one of the two dice, then throw it and observe whether or not the trial yields a 5. Is the variance of the number of successes larger than, equal to, or smaller than with the standard binomial sampling scheme?

4. (Holly) One foggy morning, a Coast Guard station receives a radio distress call from a ship at sea: "This is the *Thomas Bayes.* Am sinking fast near ≪ *static burst* ≫ Island. Send help fast!" And then the ship's transmitter goes dead. Unfortunately, there are two islands, call them A and B, that could be the approximate location of the sinking ship and they are in opposite directions. The station commander knows that the *Thomas Bayes* is a sight-seeing ship and could be at either island with a load of tourists. Time is of the essence and he must split his fleet of N search boats and search the waters off of both islands at the same time. Because of the time of day, the commander thinks that the *Thomas Bayes* is more likely to be at one island than the other; in fact, he estimates that there is probability p_A that the ship is at Island A and $1 - p_A$ that it is at Island B. Assume that each ship has probability p_s of finding the ship if it is sent to the correct island and that all the search boats search independently. How should the commander divide his fleet of boats to maximize the probability of finding the sinking ship? That is, what is n, the number of search boats that should be assigned to Island A?

   Find a general solution (remembering that n must be an integer) and then plot the optimal n as p_A varies from 0 to 1 for the case of $N = 13$ search boats and $p_s = 0.2$. Repeat for $N = 30$ search boats and $p_s = 0.1$. Write an R function with inputs N and p_s which will plot the optimal n as a function of p_A.