Stat S-110 Homework 2, Summer 2015

Due: Wednesday 7/8 at noon (Eastern Time), submitted as a PDF file via the course website or as a hard copy in class. Show your work and explain your steps.

Collaboration Policy: You are welcome to discuss the problems with others, but *you must write up your solutions yourself and in your own words.* Copying someone else's solution, or just making trivial changes for the sake of not copying verbatim, is not acceptable. For example, in problems where you have to make up a "story" or example, two students should not have the exact same answer, or almost the same answer except one has an example with dogs chasing cats and the other has an example with cats chasing mice, with the same structure and the same numbers.

The following problems are from Chapters 2-3 of the book.

1. (BH 2.41) You are the contestant on the Monty Hall show. Monty is trying out a new version of his game, with rules as follows. You get to choose one of three doors. One door has a car behind it, another has a computer, and the other door has a goat (with all permutations equally likely). Monty, who knows which prize is behind each door, will open a door (but not the one you chose) and then let you choose whether to switch from your current choice to the other unopened door.

Assume that you prefer the car to the computer, the computer to the goat, and (by transitivity) the car to the goat.

(a) Suppose for this part only that Monty always opens the door that reveals your less preferred prize out of the two alternatives, e.g., if he is faced with the choice between revealing the goat or the computer, he will reveal the goat. Monty opens a door, revealing a goat (this is again for this part only). Given this information, should you switch? If you do switch, what is your probability of success in getting the car?

(b) Now suppose that Monty reveals your less preferred prize with probability p, and your more preferred prize with probability $q = 1 - p$. Monty opens a door, revealing a computer. Given this information, should you switch (your answer can depend on p)? If you do switch, what is your probability of success in getting the car (in terms of p)?

2. (BH 2.45) A gambler repeatedly plays a game where in each round, he wins a dollar with probability $1/3$ and loses a dollar with probability $2/3$. His strategy is "quit when he is ahead by \$2", though some suspect he is a gambling addict anyway.

Suppose that he starts with a million dollars. Show that the probability that he'll ever be ahead by \$2 is less than $1/4$.

3. (BH 2.51)

(a) There are two crimson jars (labeled C_1 and C_2) and two mauve jars (labeled M_1 and M_2). Each jar contains a mixture of green gummi bears and red gummi bears. Show by example that it is possible that C_1 has a much higher percentage of green gummi bears than M_1, and C_2 has a much higher percentage of green gummi bears than M_2, yet if the contents of C_1 and C_2 are merged into a new jar and likewise for M_1 and M_2, then the combination of C_1 and C_2 has a lower percentage of green gummi bears than the combination of M_1 and M_2.

(b) Explain how (a) relates to Simpson's paradox, both intuitively and by explicitly defining events A, B, C as in the statement of Simpson's paradox.

4. (BH 2.54) Fred decides to take a series of n tests, to diagnose whether he has a certain disease (any individual test is not perfectly reliable, so he hopes to reduce his uncertainty by taking multiple tests). Let D be the event that he has the disease, $p = P(D)$ be the prior probability that he has the disease, and $q = 1 - p$. Let T_j be the event that he tests positive on the jth test.

(a) Assume for this part that the test results are conditionally independent given Fred's disease status. Let $a = P(T_j|D)$ and $b = P(T_j|D^c)$, where a and b don't depend on j. Find the posterior probability that Fred has the disease, given that he tests positive on all n of the n tests.

(b) Suppose that Fred tests positive on all n tests. However, some people have a certain gene that makes them *always* test positive. Let G be the event that Fred has the gene. Assume that $P(G) = 1/2$ and that D and G are independent. If Fred does *not* have the gene, then the test results are conditionally independent given his disease status. Let $a_0 = P(T_j|D, G^c)$ and $b_0 = P(T_j|D^c, G^c)$, where a_0 and b_0 don't depend on j. Find the posterior probability that Fred has the disease, given that he tests positive on all n of the tests.

5. (BH 2.55) A certain hereditary disease can be passed from a mother to her children. Given that the mother has the disease, her children independently will have it with probability $1/2$. Given that she doesn't have the disease, her children won't have it either. A certain mother, who has probability $1/3$ of having the disease, has two children.

(a) Find the probability that neither child has the disease.

(b) Is whether the elder child has the disease independent of whether the younger child has the disease? Explain.

(c) The elder child is found not to have the disease. A week later, the younger child is also found not to have the disease. Given this information, find the probability that the mother has the disease.

6. (BH 3.2)

(a) Independent Bernoulli trials are performed, with probability 1/2 of success, until there has been at least one success. Find the PMF of the number of trials performed.

(b) Independent Bernoulli trials are performed, with probability 1/2 of success, until there has been at least one success and at least one failure. Find the PMF of the number of trials performed.

7. (BH 3.16) Let $X \sim \text{DUnif}(C)$, and B be a nonempty subset of C. Find the conditional distribution of X, given that X is in B.

8. (BH 3.20) Suppose that a lottery ticket has probability p of being a winning ticket, independently of other tickets. A gambler buys 3 tickets, hoping this will triple the chance of having at least one winning ticket.

(a) What is the distribution of how many of the 3 tickets are winning tickets?

(b) Show that the probability that at least 1 of the 3 tickets is winning is $3p-3p^2+p^3$, in two different ways: by using inclusion-exclusion, and by taking the complement of the desired event and then using the PMF of a certain named distribution.

(c) Show that the gambler's chances of having at least one winning ticket do not quite triple (compared with buying only one ticket), but that they do *approximately* triple if p is small.

9. (BH 3.22) There are two coins, one with probability p_1 of Heads and the other with probability p_2 of Heads. One of the coins is randomly chosen (with equal probabilities for the two coins). It is then flipped $n \geq 2$ times. Let X be the number of times it lands Heads.

(a) Find the PMF of X.

(b) What is the distribution of X if $p_1 = p_2$?

(c) Give an intuitive explanation of why X is *not* Binomial for $p_1 \neq p_2$ (its distribution is called a *mixture* of two Binomials). You can assume that n is large for your explanation, so that the frequentist interpretation of probability can be applied.

10. (BH 3.23) There are n people eligible to vote in a certain election. Voting requires registration. Decisions are made independently. Each of the n people will register with probability p_1. Given that a person registers, he or she will vote with probability p_2. Given that a person votes, he or she will vote for Kodos (who is one of the candidates) with probability p_3. What is the distribution of the number of votes for Kodos (give the PMF, fully simplified, or the name of the distribution, including its parameters)?

11. (BH 3.24) Let X be the number of Heads in 10 fair coin tosses.

(a) Find the conditional PMF of X, given that the first two tosses both land Heads.

(b) Find the conditional PMF of X, given that at least two tosses land Heads.

12. (BH 3.34) There are n students at a certain school, of whom $X \sim \text{Bin}(n, p)$ are Statistics majors. A simple random sample of size m is drawn ("simple random sample" means sampling without replacement, with all subsets of the given size equally likely).

(a) Find the PMF of the number of Statistics majors in the sample, using the law of total probability (don't forget to say what the support is). You can leave your answer as a sum (though with some algebra it can be simplified, by writing the binomial coefficients in terms of factorials and using the binomial theorem).

(b) Give a story proof derivation of the distribution of the number of Statistics majors in the sample; simplify fully.

Hint: Does it matter whether the students declare their majors before or after the random sample is drawn?