Stat 110 Homework 2, Fall 2013

Due: Friday 9/20 at 1:10 pm, in the Stat 110 dropbox on the 7th floor of the Science Center. No late homework will be accepted. Please write your name, and staple your homework. Show your work and make sure to give clear, careful, convincing explanations.

Reading: Sections 2.1-2.6.

1. A researcher wants to estimate the percentage of people in some population who have used illegal drugs, by conducting a survey. Concerned that a lot of people would lie when asked a sensitive question like "Have you ever used illegal drugs?", the researcher uses a method known as *randomized response*. A hat is filled with slips of paper, each of which says either "I have used illegal drugs" or "I have not used illegal drugs". Let p be the proportion of slips of paper that say "I have used illegal drugs" (p is chosen by the researcher in advance).

   Each participant chooses a random slip of paper from the hat and answers (truthfully) "yes" or "no" to whether the statement on that slip is true. The slip is then returned to the hat. The researcher does not know which type of slip the participant had. Let y be the probability that a participant will say "yes", and d be the probability that a participant has used illegal drugs.

(a) Find y, in terms of d and p.

(b) What would be the worst possible choice of p that the researcher could make in designing the survey? Explain.

(c) Now consider the following alternative system. Suppose that proportion p of the slips of paper say "I have used illegal drugs", but that now the remaining $1 - p$ say "I was born in winter" rather than "I have not used illegal drugs". Assume that $1/4$ of people are born in winter, and that a person's season of birth is independent of whether they have used illegal drugs. Find d, in terms of y and p.

2. Alice, Bob, and 100 other people live in a small town. Let C be the set consisting of the 100 other people, let A be the set of people in C who are friends with Alice, and let B be the set of people in C who are friends with Bob. Suppose that for each person in C, Alice is friends with that person with probability $1/2$, and likewise for Bob, with all of these friendship statuses independent.

(a) Let $D \subseteq C$. Find $P(A = D)$.

(b) Find $P(A \subseteq B)$.

(c) Find $P(A \cup B = C)$.

3. A fair coin is flipped 3 times (with all possible outcomes equiprobable). Each result of a toss is recorded on a slip of paper (writing "H" if Heads and "T" if Tails), and the 3 slips of paper are thrown into a hat.

(a) Explain what is wrong with the following argument: "there is a 50% chance that the 3 tosses all landed the same way, since obviously there are two out of the three that match, and then the other toss has a 50% chance of matching these two".

(b) Find the probability that all 3 tosses landed Heads, given that at least 2 were Heads.

(c) Two of the slips of paper are drawn from the hat, and both show the letter "H". Given this information, what is the probability that all 3 tosses landed Heads?

(d) Give a clear explanation in words for why the inequality or equality you found between the answers to (b) and to (c) makes sense (e.g., if your answer to (b) is greater than your answer to (c), explain why).

4. Massachusetts voters are surveyed one at a time, independently. Suppose that the party affiliation of a voter is Democrat with probability p_1, Republican with probability p_2, Green-Rainbow with probability p_3, and other with probability p_4. What is the probability that the sample will obtain a Green-Rainbow party member before obtaining a Republican?

5. A standard deck of cards will be shuffled and then the cards will be turned over one at a time until the first Ace is revealed. Let B be the event that the *next* card in the deck will also be an Ace.

(a) Intuitively, how do you think $P(B)$ compares in size with $1/13$ (the overall proportion of Aces in a deck of cards)? Explain your intuition. (Note: the only way to lose points on this part is by trying to do mathematical calculations rather than explaining your intuition!)

(b) Let C_j be the event that the first Ace is at position j in the deck. Find $P(B|C_j)$ in terms of j, fully simplified.

(c) Using the law of total probability, find an expression for $P(B)$ as a sum. (The sum does not need to be simplified, but it should be something that could easily be calculated with R or Wolfram Alpha.)

(d) Find a fully simplified expression for $P(B)$ using a symmetry argument.
Hint: if you were deciding whether to bet on the next card after the first Ace being an Ace or to bet on the last card in the deck being an Ace, would you have a preference?

6. (Exercise 2.14 in book) Fred has just tested positive for a certain disease.

(a) Given this information, find the posterior odds that he has the disease, in terms of the prior odds, the sensitivity of the test, and the specificity of the test.

(b) Not surprisingly, Fred is much more interested in P(have disease|test positive), known as the *positive predictive value*, than in the sensitivity P(test positive|have disease). A handy rule of thumb in biostatistics and epidemiology is as follows:

For a rare disease and a reasonably good test, specificity matters much more than sensitivity in determining the positive predictive value.

Explain intuitively why this rule of thumb works. For this part you can make up some specific numbers and interpret probabilities in a "frequentist" way as proportions in a large population, e.g., assume the disease afflicts 1% of a population of 10000 people and then consider various possibilities for the sensitivity and specificity.

7. (Exercise 2.15 in book) A family has two children. Let C be a characteristic that a child can have, and assume that each child has characteristic C with probability p, independently of each other and of gender. For example, C could be the characteristic "born in winter" as in Example 2.2.6. Show that the probability that both children are girls given that at least one is a girl with characteristic C is $\frac{2-p}{4-p}$, which is $1/3$ if $p = 1$ (agreeing with the first part of Example 2.2.4) and approaches $1/2$ from below as $p \to 0$ (agreeing with Example 2.2.6).