CS 70 Discrete Mathematics and Probability Theory
Fall 2011 Rao HW 11

Due Monday, November 14, 5:00pm

You *must* write up the solution set entirely on your own. You must never look at any other students' solutions (not even a draft), nor share your own solutions (not even a draft).

Please put your answer to each problem on its own sheet of paper, and paper-clip (don't staple!) the sheets of paper together. Label each sheet of paper with your name, your discussion section number (101–108), your login, and "CS70–Fall 2011". Turn in your homework and problem x into the box labeled "CS70 – Fall 2011, Problem x" whereon the 2nd floor of Soda Hall. Failure to follow these instructions will likely cause you to receive no credit at all.

1. (14 pts.) Basics: joint distributions
Consider Figure 1 in Lecture 17 of the Reader.

(a) What are the distributions of X and Y?

(b) What is $\Pr[X = 1 \wedge Y = 1]$?

(c) What is $\Pr[X = 1|Y = 1]$?

(d) What is the distribution of the random variable Y conditioned on $Y = 1$?

(e) What is the distribution of the random variable X conditioned on $Y = 1$?

(f) Say X represents one of three variants of a gene, and $Y = 0$ represents the event that an individual is healthy, $Y = 1$ has a type 1 variant of a disease, and $Y = 2$ has a type 2 variant of the disease.
What is the probability that an individual is sick? Given that X is not 1, what is the probability that the individual is sick?

2. (16 pts.) Who's better
I am playing in a tennis tournament, and I am up against a player I have watched but never played before. Based on what I have seen, I consider three possible models for our relative strengths:

- Model A: We are evenly matched, so that each of us is equally likely to win each game.
- Model B: I am slightly better, so that I win each game independently with probability 0.6.
- Model C: My opponent is slightly better and wins each game independently with probability 0.6.

Before we play, I consider each of these possibilities to be equally likely.

In our match, we play until one player wins three games. I win the second game, but my opponent wins the first, third and fourth games. After the match, with what probability should I believe in model C (i.e., that my opponent is slightly better than me)?

[Hint: The unknown r.v. here is X, which takes values A, B, C. My prior knowledge about X is a uniform distribution over these three values. The observed r.v.'s are the outcomes Y_1, Y_2, Y_3, Y_4 of the four games, which take values 1 (denoting a win for me) and 0 (denoting a win for my opponent). As usual we know the conditional probabilities, e.g., $\Pr[Y_1 = 1 \mid X = A] = 0.5, \Pr[Y_1 = 1|X = B] = 0.6$ etc.]

3. (20 pts.) Communication over noisy channels

Recall the communication problem we considered in the lecture (see also Lecture 17 of the Reader).

(a) Suppose the noise probability of the channel is $p = 0.25$. Using the summation formula at the beginning of the last section of Lecture 17 of the Reader, compute the exact error probability numerically for $n = 5, 10, 15$ repetitions. Compare this to the bound obtained from Chebyshev's Inequality near the end of the same section. How good is the bound? Also, using the same exact formula, find the smallest number of repetitions n^* such that the error probability is less than 0.01. How does this compare to the bound of 300 given by Chebyshev?

(b) In class, we considered the case when X is equally likely to be 1 and 0. Now suppose $\Pr[X = 1] = q$. Re-derive the maximum a posterior (MAP) decision rule in this more general case. Make your rule as explicit as possible. Is it still the simple majority rule we derived in class? For $q = 0.1$, $p = 0.25$ and $n = 5, 10, 15$, work out explicitly what the decision rule is. How does it compare to the simple majority rule? Does it make intuitive sense?

4. (20 pts.) The myth of fingerprints

A crime has been committed. The police discover that the criminal has left DNA behind, and they compare the DNA fingerprint against a police database containing DNA fingerprints for 20 million people. Assume that the probability that two DNA fingerprints (falsely) match by chance is 1 in 10 million. Assume that, if the crime was committed by someone whose DNA fingerprint is on file in the police database, then it's certain that this will turn up as a match when the police compare the crime-scene evidence to their database; the only question is whether there will be any false matches.

Let D denote the event that the criminal's DNA is in the database; $\neg D$ denotes the event that the criminal's DNA is not in the database. Assume that it is well-documented that half of all such crimes are committed by criminals in the database, i.e., assume that $\Pr[D] = \Pr[\neg D] = 1/2$. Let the random variable X denote the number of matches that are found when the police run the crime-scene sample against the DNA database.

(a) Calculate $\Pr[X = 1|D]$.

(b) Calculate $\Pr[X = 1|\neg D]$.

(c) Calculate $\Pr[\neg D|X = 1]$. Evaluate the expression you get and compute this probability to at least two digits of precision.

As it happens, the police find exactly one match, and promptly prosecute the corresponding individual. You are appointed a member of the jury, and the DNA match is the only evidence that the police present. During the trial, an expert witness testifies that the probability that two DNA fingerprints (falsely) match by chance is 1 in 10 million. In his summary statement, the prosecutor tells the jury that this means that the probability that the defendant is innocent is 1 in 10 million.

(d) What is wrong with the prosecutor's reasoning in the summary statement?

(e) Do you think the defendant should be convicted? Why or why not?