CS 70 Discrete Mathematics and Probability Theory
Spring 2011 Demmel HW 11

Due April 15, 8:00am

You may work in groups of up to 3 people (no larger!). Please read the group collaboration policies on bSpace or `www.cs.berkeley.edu/~demmel/cs70_Spr11` before beginning group work. 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 begin your answer to each question on a new sheet of paper, and make sure that each sheet is labeled with your name, section number, GSI name, the assignment number, the question number, and "CS70–Spring 2011".

Turn in each question in a different box in 283 Soda Hall: Question 1 in the box labeled "CS70 - 1", Question 2 in the box labeled "CS70 - 2", etc. Reason: Different problems will be graded in parallel by different readers.

Warning: You risk receiving no credit, or losing points, for any homework that does not conform to the above regulations! Please take the time to write clear and concise solutions; we will not grade messy or unreadable submissions. No late homeworks will be accepted. We will drop the lowest two homework scores.

1. **(20 pts.) Family Planning**

   Mr and Mrs Brown are really desperate for girls and decide to continue having children until they have their first girl, no matter how long they have to wait. Assume that each child is equally likely to be a boy or a girl, independent of all other children, and that there are no multiple births. Let B and G denote the numbers of boys and girls respectively that the Browns have. Let T be the total number of children they have.

   (a) List the outcomes of the sample space and assign probabilities to the outcomes.

   (b) Write down the distributions of the random variables B, G and T.

   (c) Write down the expectations of B, G and T.

   Suppose instead the Browns decide to have children until they have *two* girls.

   (d) List the outcomes in the new sample space and assign probabilities to the outcomes.

   (e) Compute the distribution and the expectation of the total number of boys that the Browns have under this new strategy.

2. **(20 pts.) Misprints**

   A textbook has on average one misprint per page.

   (a) What is the chance that you see exactly 5 misprints on page 1?

   (b) What is the chance that you see exactly 5 misprints on some page in the textbook, if the textbook is 500 pages long?

[HINT: You may assume that misprints are "rare events" that obey the Poisson distribution.]

3. (20 pts.) Random Homeworks

Consider the example where the homeworks of n students are returned randomly to the students, with all $n!$ permutations equally likely. Let X be the number of students who got their own homeworks back.

(a) Compute $\mathbb{E}(X)$.

(b) Compute $\text{Var}(X)$.

(c) Find an upper bound on the probability that at least 5 students get back their homeworks in a class of 190 students.

4. (20 pts.) Practice with variance

Here are some calculations that we are hoping you will have down solid. Please make sure you are comfortable with these calculations.

(a) Let X be an indicator random variable for the event that the top card of a well-shuffled 52-card deck is the Ace of Spades. Calculate $\mathbb{E}(X)$ and $\text{Var}(X)$.

(b) Let Y be a random variable with the following distribution: $Y = 2$ with probability $\frac{1}{3}$, $Y = 0$ with probability $\frac{1}{3}$, and $Y = -2$ with probability $\frac{1}{3}$. Calculate $\text{Var}(Y)$.

(c) Let Z be a random variable with the following distribution: $Z = 5$ with probability $\frac{1}{6}$, $Z = 2$ with probability $\frac{1}{3}$, $Z = 0$ with probability $\frac{1}{3}$, and $Z = -4$ with probability $\frac{1}{6}$. Calculate $\text{Var}(Z)$.

(d) With Z as defined in part (c), calculate $\text{Var}(Z+10)$.

Let R, S, T be independent r.v.'s with $R \sim \text{Binomial}(100, \frac{1}{2})$, $S \sim \text{Binomial}(20, \frac{1}{2})$, and $T \sim \text{Binomial}(90, \frac{1}{3})$.

(e) Find $\text{Var}(R)$, $\text{Var}(S)$, and $\text{Var}(T)$.

(f) Let $U = R+S$. Calculate $\text{Var}(U)$.

(g) Let $V = R+T$. Calculate $\text{Var}(V)$.

(h) Let $W = 2R+T$. Calculate $\text{Var}(W)$.

5. (20 pts.) Say What?

In a binary communication channel the transmitter sends *zero* or *one*, but at the receiver there are three possibilities: a *zero* is received, a *one* is received, and an *undecided* bit is received. Define the event $T_1 = \{$ 1 is sent $\}$ and $T_0 = \{$ 0 is sent $\}$ and assume that they are equally probable. At the receiver we have three events: $R_1 = \{$ 1 is received $\}$, $R_0 = \{$ 0 is received $\}$, $R_u = \{$ cannot decide the bit $\}$. We assume that we have the following conditional probabilities: $\Pr[R_0|T_0] = \Pr[R_1|T_1] = 0.9$, $\Pr[R_u|T_0] = \Pr[R_u|T_1] = 0.09$.

(a) Find the probability that a transmitted bit is received as *undecided*.

(b) Find the probability that a bit is received in error (error means sending *one* while receiving *zero* OR sending *zero* while receiving *one*).

(c) Given that we received a *zero*, what is the conditional probability that a *zero* was sent? What is the conditional probability that a *one* was sent?

(d) Now suppose to increase the reliability, the transmitter sends a *zero* three times if it wants to send a *zero* and it sends a *one* three times if it wants to send a *one*. Suppose now the receiver sees three *zero*'s. Is the model completely specified so that you can compute the conditional probability that a *zero* is sent? If so, compute it. If not, complete the specification of the model as you wish and then compute the conditional probability.

6. (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?