CS 70 Discrete Mathematics for CS
Spring 2008 David Wagner
HW 12

Due Thursday, March 1st

Important: Show your work on all problems on this homework.

1. (5 pts.) More on the birthday paradox

We invite n people to a party. Assume that each person's birthday is uniformly and independently chosen at random from among the 365 days of the year. Let the r.v. X denote the number of pairs of people who have the same birthday.

Calculate $\text{Var}(X)$. Your answer should be a simple function of n.

2. (10 pts.) A poor shuffle

Consider the following bad method for "shuffling" (i.e., randomly permuting the elements of) a 52-element array A.

1. Initialize an array B to contain 52 "empty" indicators.
2. For j = 0 to 51, do the following:
   (a) Repeatedly generate a random integer k between 0 and 51 until we choose one where $B[k]$ is empty.
   (b) Copy $A[j]$ to $B[k]$.
3. Copy B, which now contains the shuffled elements, back to A.

Answer the following questions.

(a) Let the r.v. X_j denote the number of random integers k that are generated in step 2(a), for a fixed value of j. Calculate $\mathbf{E}[X_j]$. Your answer should be a simple function of j.

(b) Let the r.v. $X =$ the total number of random integers generated throughout this algorithm. Calculate $\mathbf{E}[X]$. Compute this value to within 2 digits of precision.

3. (5 pts.) The binomial distribution

Suppose $X \sim \text{Binomial}(n,p)$ and $Y \sim \text{Binomial}(m,p)$, i.e., X and Y are independent binomially distributed random variables with parameters n,p and m,p, respectively. Let $Z = X+Y$. Show that Z has a binomial distribution. What are its parameters?

Hint: It may be useful to know that $\binom{a}{0}\binom{b}{c} + \binom{a}{1}\binom{b}{c-1} + \cdots + \binom{a}{a}\binom{b}{c-a} = \binom{a+b}{c}$. (The intuition: to choose a subset of c items from a set of $a+b$ items, we can choose some subset of some number of items from the first a items, say i of them, and then we can choose a subset of $c-i$ items from the second b items.) You may use this equality without proof. There are several ways to solve this problem, so if you don't need to use this equality, that's fine, too.

4. (11 pts.) Illegal immigration

The population of the US consists of 280 million legal residents and 5 million illegal aliens. Legislators

implement a get-tough policy. All legal residents will be issued noncounterfeitable ID cards. Then, the police will stop people at random to check their ID, immediately deporting any illegal alien who is stopped and letting legal residents go. Assume that the random stops are implemented in a scrupulously even-handed manner, so we can model them as the police picking a person uniformly at random from the entire US population and stopping them to check their ID. Assume that no new illegal aliens enter the US and that no illegal alien obtains a valid ID.

(a) On average, how many random stops will the police have to conduct (in total) before every illegal alien is deported? Compute this value to within 2 digits of precision.
Hint: If you end up with a sum of the form $\sum_i \frac{n-i}{m-i}$ and you have a hard time evaluating it, try reversing the order of summation (e.g., replace i by $m-i$) or using the equality $\frac{n-i}{m-i} = \frac{n-m}{m-i} + 1$. There are many ways to do this problem, so if you don't end up with something of this form, that's okay, too.

(b) On average, how many times will each legal resident be stopped before every illegal alien is deported? Compute this value to within 2 digits of precision.

(c) Re-do part (a) and (b), but this time assuming that the police cease stopping people once half of the illegal aliens are deported.

5. (15 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. One match is found, and the police promptly prosecute the corresponding individual. You are appointed a member of the jury. 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. In the following, compute all probabilities to at least two digits of precision.

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

(b) 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. Let G denote the event that this defendant did indeed commit the crime. Let I_0 denote event that this defendant is innocent, and the crime was instead committed by someone whose DNA fingerprint is in the police database, and I_1 the event that the crime was committed by someone whose DNA fingerprint is not in the database. Calculate $\Pr[X = 0|I_1]$.

(c) Calculate $\Pr[X = 1|I_1]$.

(d) Calculate $\Pr[X = 1|I_0]$. 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 other false matches.

(e) Calculate $\Pr[X = 1|G]$. As before, assume that, if this defendant is guilty, then it's certain that his DNA will match the DNA left at the crime scene.

(f) ~~Suppose that there is no other evidence against the defendant. Assume that this defendant is one of a population of a million people who, if not for the DNA fingerprint evidence, could all be considered equally likely suspects—i.e., $\Pr[G] = 1/10^6$. Assume that half of those other people have their DNA fingerprint on file in the DNA database, and half do not. Calculate $\Pr[G|X = 1]$.~~

Changed 4/30: Deleted 5(f). Also, please see the newsgroup for a clarification on 5(b)–5(e).

6. (4 pts.) Sudden infant death syndrome

In 1999, UK prosecutors charged British mother S.C. with murdering her two infant children. (This is a

true story; I have abbreviated the accused's name, for privacy.) It seems that both of her two children unexplainedly died in their sleep. Her defense was that both infants had died of sudden infant death syndrome (SIDS), also known as crib death or cot death. The prosecution observed that the rate of crib death was approximately 1 in 8,550 for a well-off family, like hers. The prosecution reasoned that this means that the probability of a double crib-death occuring by chance is $(1/8550)^2 \approx 1/73000000$, and thus concluded that the chances of both her children dieing naturally in this way is 1 in 73,000,000. The prosecution argued that therefore S.C. was almost surely guilty. What do you think of the prosecution's argument?