CS 70 Spring 2005 — Discrete Mathematics for CS — Clancy/Wagner — HW 12

Due Thursday, April 28

Coverage: This assignment involves topics from the lectures of April 19 and 21, and from Rosen section 5.3.

Administrative reminders: We will accept only unformatted text files or PDF files for homework submission. Include your name, login name, section number, and partner list in your submission. Give the command `submit hw12` to submit your solution to this assignment.

Homework exercises:

1. **(8 pts.) Independence**

   (a) Show that, for independent random variables X, Y, we have $\mathbb{E}[XY] = \mathbb{E}[X] \times \mathbb{E}[Y]$.
   HINT: Show first—carefully!—that, even if the random variables are *not* independent, $\mathbb{E}[XY] = \sum_a \sum_b ab \times \Pr[X = a \wedge Y = b]$.

   (b) Give a simple example to show that the conclusion of the previous part is not necessarily true when X and Y are not independent.

2. **(16 pts.) Those 3407 Votes**
   In the aftermath of the 2000 US Presidential Election, many people have claimed that the 3407 votes cast for Pat Buchanan in Palm Beach County are statistically highly significant, and thus of dubious validity. In this problem, we will examine this claim from a statistical viewpoint.

   The total percentage votes cast for each presidential candidate in the entire state of Florida were as follows:

Gore	Bush	Buchanan	Nader	Browne	Others
48.8%	48.9%	0.3%	1.6%	0.3%	0.1%

   In Palm Beach County, the actual votes cast (before the recounts began) were as follows:

Gore	Bush	Buchanan	Nader	Browne	Others	Total
268945	152846	3407	5564	743	781	432286

   To model this situation probabilistically, we need to make some assumptions. Let's model the vote cast by each voter in Palm Beach County as a random variable X_i, where X_i takes on each of the six possible values (five candidates or "Others") with probabilities corresponding to the Florida percentages. (Thus, e.g., $\Pr[X_i = \text{Gore}] = 0.488$.) There are a total of $n = 432286$ voters, and their votes are assumed to be mutually independent. Let the r.v. B denote the total votes cast for Buchanan in Palm Beach County (i.e., the number of voters i for which $X_i = \text{Buchanan}$).

(a) Compute the expectation $\mathbb{E}[B]$ and the variance $\mathrm{Var}[B]$.

(b) Use Chebyshev's inequality to compute an *upper bound* b on the probability that Buchanan receives at least 3407 votes, i.e., find a number b such that

$$\Pr[B \geq 3407] \leq b.$$

Based on this result, do you think Buchanan's vote is significant?

(c) Now suppose that your bound b in part (b) is in fact sharp, i.e., assume that $\Pr[X \geq 3407]$ is *equal to b*. [*In fact the true value of this probability is quite a bit smaller than b.*] Suppose also that all 67 counties in Florida have the same number of voters as Palm Beach County, and that all behave independently according to the same statistical model as Palm Beach County. What is the probability that in *at least one* of the counties, Buchanan receives at least 3407 votes? How would this affect your judgement as to whether the Palm Beach tally is significant?

(d) Our model assumes that all voters behave like the fabled "swing voters," in the sense that they are undecided when they go to the polls and end up making a random decision. A more realistic model would assume that only a fraction (say, about 20%) of voters are in this category, the others having already decided. Suppose then that 80% of the voters in Palm Beach County vote deterministically according to the state-wide proportions for Florida, while the remaining 20% behave randomly as described earlier. Does your bound b in part (b) increase, decrease or remain the same under this model? Justify your answer.

3. (10 pts.) Find the joker

(a) Take an ordinary deck of 52 playing cards, and add a single joker. Shuffle it and turn up the cards one at a time until the joker appears. On average, how many cards are required until you see the joker?

(b) Now take a deck of 52 playing cards, add two jokers, shuffle, and turn up cards one at a time until the first time that a joker appears. On average, how many cards are required until you see the first joker?

(c) Check your work by writing a program. Run it for some large number of trials (say, a million trials) and compute the average number of cards needed to see the first joker. What do you get?

4. (20 pts.) Random variables mod p

Let the random variables X and Y be distributed independently and uniformly at random in the set $\{0, 1, \ldots, p-1\}$, where $p > 2$ is a prime.

(a) What is the expectation $\mathbb{E}[X]$?

(b) Let $S = (X+Y) \bmod p$ and $T = XY \bmod p$. What are the distributions of S and T?

(c) What are the expectations $\mathbb{E}[S]$ and $\mathbb{E}[T]$?

(d) By linearity of expectation, we might expect that $\mathbb{E}[S] = (\mathbb{E}[X] + \mathbb{E}[Y]) \bmod p$. Explain why this does not hold in the present context; i.e., why does the value for $\mathbb{E}[S]$ obtained in part (c) not contradict linearity of expectation?

(e) Since X and Y are independent, we might expect that $\mathbb{E}[T] = \mathbb{E}[X]\mathbb{E}[Y] \bmod p$. Does this hold in this case? Explain why/why not.

5. (6 pts.) Geometric distribution

James Bond is imprisoned in a cell from which there are three possible ways to escape: an air-conditioning duct, a sewer pipe and the door (which is unlocked). The air-conditioning duct leads him on a two-hour trip whereupon he falls through a trap door onto his head, much to the amusement of his captors. The sewer pipe is similar but takes five hours to traverse. Each fall produces temporary amnesia and he is returned to the cell immediately after each fall. Assume that he always immediately chooses one of the three exits from the cell with probability $\frac{1}{3}$. On the average, how long does it take before he realizes that the door is unlocked and escapes?

Hint: if you're doing a lot of algebra, you're taking the wrong approach.