CS 70 Discrete Mathematics and Probability Theory
Summer 2012 Filiba HW 7

Due Tuesday, July 31, 4:59pm

1. (12 pts.) Misprints
A textbook has on average one misprint per page.

(a) What is the chance that you see exactly 4 misprints on page 1?

(b) What is the chance that you see exactly 4 misprints on some page in the textbook, if the textbook is 250 pages long?

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

2. (18 pts.) Packets Over the Internet
n packets are sent over the Internet (n even). Consider the following probability models for the packet loss process:

(i) Each packet is lost independently with probability p.

(ii) With probability p, all n packets are lost. Otherwise all packets are received.

(iii) The n packets are divided into $n/2$ groups of 2 packets, and each group is either entirely lost, with probability p, or entirely received. Losses of different groups are independent events.

In each of the three models, compute the distribution, mean and variance of the number of packets losses. Plot the distribution in each of the three cases. Does the answer depend on the probability model?

3. (16 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 $\text{Var}(W)$.

4. (12 pts.) Polling populations

Joe wishes to estimate the true fraction f of smokers in a large population without asking each and every person. He plans to select n people at random and then employ the estimator $F = S/n$, where S denotes the number of people in a size-n sample who are smokers. Joe would like to sample the minimum number of people, but also guarantee an upper bound p on the probability that the estimator F differs from the true value f by a value greater than or equal to d i.e., for a given accuracy d and given confidence p, Joe wishes to select the minimum n such that

$$\Pr(|F - f| \geq d) \leq p \quad .$$

For $p = 0.05$ and a particular value of d, Joe uses the Chebyshev inequality to conclude that n must be at least 50,000. Determine the new minimum value for n if:

(a) the value of d is reduced to half of its original value.

(b) the probability p is reduced to half of its original value, or $p = 0.025$.

5. (16 pts.) Those 3407 Votes

In the aftermath of the hotly contested 2000 US Presidential Election, many people claimed that the 3407 votes cast for independent candidate Pat Buchanan in Palm Beach County were 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 $\text{Var}(B)$.

(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 $\Pr[B \geq 3407]$ 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.

6. (15 pts.) Family Planning

Mr and Mrs Brown decide to continue having children until they either have their first girl or until they have five children. 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 C be the total number of children they have.

(a) Write down the sample space together with the probability of each sample point.

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

(c) Compute the expectations and variances of B, G and C using a direct calculation.

(d) Write down the joint distribution of G and C.

(e) Write down the conditional distributions of C given $G = i$ for all possible values i that G can take on.

7. (14 pts.) Bulb lifetimes

Earlier this year, after much searching and experimentation, I finally found a CFL light bulb I like: it's energy-efficient, bright, and has a pleasing color that complements my bedroom's color scheme nicely. Perfect.

Unfortunately, I just recently learned that this model of CFL light bulb has been discontinued and will no longer be manufactured. Oh no! Fortunately, when I found a model I liked early this year, I prudently bought a dozen light bulbs, so I currently have one bulb in my bedroom and a stockpile of 11 spares. As soon as one bulb burns out, I will immediately replace it with another from my stockpile, until I run out of stockpiled bulbs. How long will I have before the last of my light bulbs burns out?

Let $X_1, X_2, \ldots, X_{12}$ denote the number of hours that each bulb lasts. According to the manufacturer, each bulb has an average lifetime of 8000 hours, so let's model the lifetime of each bulb as a Geometric distribution: $X_i \sim \text{Geometric}(\frac{1}{8000})$. Assume that the lifetime of each bulb is independent of all other bulbs. Let X denote the total number of hours of pleasing lighting that I get from this set of 12 bulbs.

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

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

(c) I'd love it if I could get at least 16000 hours of light out of these 12 bulbs: given my usage patterns, that's over 10 years of life, and by then, there will probably be some better technology anyway (LEDs?). Find a lower bound on the probability that $X \geq 16000$.

8. (10 pts.) James Bond

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?

9. (6 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 her 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 dying 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?