Homework 11

Out: 29 Nov. Due: 6 Dec.

Instructions: *Start each problem on a new sheet. Write your name, SID number, section number and "CS70" on every sheet. Clip together (do not staple!) the sheets for different problems (but if you use more than one sheet for one problem, staple those). Put your solutions in the homework box on Soda level 2 by 5pm on Thursday. You are encouraged to form small groups (two to four people) to work through the homework, but you* **must** *write up all your solutions on your own.*

1. Total Annihilation

A superpower has 3450 missiles stored in well separated silos. An enemy is considering a sneak attack. However, for the attack to succeed every one of the missiles must be destroyed. Assume that each attacking warhead hits exactly one of the missile silos, with each silo being equally likely to be the one that is hit. How many warheads would you expect to be needed to ensure the complete destruction of every missile? Explain your answer; you may use any result from class without proof provided it is clearly stated.

2. More on Coupon Collecting

(a) Let X be the r.v. in the coupon collecting problem, i.e., X is the number of cereal boxes we need to buy before we have collected one copy of each of n baseball cards. Recall from Note 14 that $\mu = \mathrm{E}[X] = n\sum_{i=1}^{n}\frac{1}{i} \approx n(\ln n + \gamma)$. Assuming the fact that the variance of a geometric r.v. with parameter p is $\frac{1-p}{p^2}$, show that $\mathrm{Var}[X] = n^2\sum_{i=1}^{n}\frac{1}{i^2} - \mu$.

(b) It is well known that the series $\sum_{i=1}^{\infty}\frac{1}{i^2}$ converges to a constant value $C = \frac{\pi^2}{6} \approx 1.645$. Use part (a) together with Chebyshev's Inequality to deduce the smallest value of β for which you can say that the probability we need to buy more than $\mu + \beta n$ boxes is less than $\frac{1}{100}$.

(c) **[Extra credit: optional!]** Prove the fact assumed in part (a), i.e., that the variance of a geometric r.v. with parameter p is $\frac{1-p}{p^2}$. [HINT: You will need to sum a series like $S = \sum_{i=1}^{\infty} i^2 q^i$. One way to do this is to multiply S by q and subtract the result from S: this gives you a series for $(1-q)S$. Now if you look at this series carefully, you will see that you can split it into a series of the form $\sum_i iq^i$ and one of the form $\sum_i q^i$. But you know how to sum both of these: the first is like the *expectation* of X, and the second is just a geometric series.]

3. Normal Distribution

If a set of grades on a Discrete Math examination are approximately normally distributed with a mean of 68 and a standard deviation of 6.9, find:

(a) the lowest passing grade if the bottom 5% of the students fail the class;

(b) the highest B if the top 10% of the students are given A's.

NOTE: You may assume that if X is normal with mean 0 and variance 1, then $\Pr[X \leq 1.3] \approx 0.9$ and $\Pr[X \leq 1.65] \approx 0.95$.

[continued overleaf]

4. 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 $\mathrm{E}[B]$ 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.

5. Cardinalities

Answer the following two questions, giving a proof in each case. You may use without proof any result covered in class provided you state it clearly.

(a) Is the set of *pairs* of natural numbers countable or uncountable?

(b) Is the set of irrational numbers countable or uncountable?