EECS 70 Discrete Mathematics and Probability Theory
Fall 2014 Anant Sahai

Homework 13

This homework is due December 3, 2014, at 12:00 noon.

1. **Section Rollcall!**

   In your self-grading for this question, give yourself a 10, and write down what you wrote for parts (a) and (b) below as a comment. Put the answers in your written homework as well.

   (a) What discussion did you attend on Monday last week? If you did not attend section on that day, please tell us why.

   (b) What discussion did you attend on Wednesday last week? If you did not attend section on that day, please tell us why.

2. **Practice Makes Perfect**

   For this question, do 5 of the online practice problems. For your answer, write down which problems you did (the problem set title and the number of the question).

3. **Hashing & Drunk Man Lab**

   Please complete the Virtual Lab that was released in Homework 12. For consistency, rename the lab file to `lab13.ipynb`, then zip and submit it on the instructional server as `hw13.zip`.

4. **Law of Large Numbers**

   Recall that the *Law of Large Numbers* holds if, for every $\varepsilon > 0$,

   $$\lim_{n\to\infty} \Pr\left[\left|\tfrac{1}{n}S_n - \mathbb{E}(\tfrac{1}{n}S_n)\right| > \varepsilon\right] = 0.$$

   In class, we saw that the Law of Large Numbers holds for $S_n = X_1 + \cdots + X_n$, where the X_i's are i.i.d. random variables. This problem explores if the Law of Large Numbers holds under other circumstances.

   Packets are sent from a source to a destination node over the Internet. Each packet is sent on a certain route, and the routes are disjoint. Each route has a failure probability of p and different routes fail independently. If a route fails, all packets sent along that route are lost. You can assume that the routing protocol has no knowledge of which route fails.

   For each of the following routing protocols, determine whether the Law of Large Numbers holds when S_n is defined as the total number of received packets out of n packets sent. Answer YES if the Law of Large Number holds, or NO if not, and give a brief justification of your answer. (Whenever convenient, you can assume that n is even.)

   (a) YES or NO: Each packet is sent on a completely different route.

   (b) YES or NO: The packets are split into $n/2$ pairs of packets. Each pair is sent together on its own route (i.e., different pairs are sent on different routes).

(c) YES or NO: The packets are split into 2 groups of $n/2$ packets. All the packets in each group are sent on the same route, and the two groups are sent on different routes.

(d) YES or NO: All the packets are sent on one route.

5. Simplified Self-Grading

There are about $n = 500$ self-graded question parts in this iteration of EECS 70. For this simplified version of self-grading, we use a scale from 0 to 4 instead of the $0, 2, 5, 8, 10$ scale currently being used. On each of them, a student assigns a grade S_i. For each homework, readers randomly grade a subset of the problems. Assume that $n/5$ of the question parts are graded by the readers (chosen uniformly over all the problem parts) and the readers assign grades R_i. Assume that R_i may deviate from an honest self-grade S_i according to the conditional probabilities given in table 1.

R_i \ S_i	0	1	2	3	4
0	3/4	1/4	0	0	0
1	1/4	1/2	1/4	0	0
2	0	1/4	1/2	1/4	0
3	0	0	1/4	1/2	1/4
4	0	0	0	1/4	3/4

Table 1: $\mathbf{P}(R_i|S_i)$.

We do the following check: we add up all of the $S_i - R_i$ for a particular student (for the subset of problems graded by readers only). If the result is too high, we suspect that a student might be inflating their grades.

a) Suppose that a student is honest. Let $p_0 = \mathbf{P}(S_i = 0)$ and $p_4 = \mathbf{P}(S_i = 4)$. Let $X_i = S_i - R_i$. Express the distribution of X_i as a function of p_0 and p_4.

b) Give the best upper-bounds you can on both $\mathbf{E}[X_i]$ and $\text{Var}(X_i)$. Your bounds shall not depend on p_0 or p_4.

c) Using Chebyshev's inequality and the above parts, compute the smallest threshold T that we should choose so that $\sum_i X_i \leq T$ for 95% of honest students?

d) Repeat the above using the Central Limit Theorem to get an approximate answer for T.

e) For simplicity, we are going to focus our attention on a hypothetical student who never truly deserves full points and never truly deserves a zero on any question part. Recompute better upper bounds on both $\mathbf{E}[X_i]$ and $\text{Var}(X_i)$ that are valid for this student. Recalculate the relevant theshold T using the Central Limit Theorem.

f) Assume this student is inflating their true self-grades S_i by adding 1 point to a question part with probability $1/2$. What is their risk of being caught (i.e., above the threshold T)? (Here, explain how you are modeling things to be true to the spirit of this problem.)

g) If this student is willing to accept a 50% chance of being caught cheating, by how much can they systematically inflate their grade(i.e. inflates his/her grade to every question by some constant number of points)? Assume that they can inflate by no more than 3 points per question part. (Because inflating a 0 to a 4 would get them slammed the first time they did it.) We will assume that $5, 6$ and 7 are allowed as self-reported grades to keep things simple.

h) Is it worth trying to cheat on self-grading, even for a grade-maximizing sociopath[1] student with no internal sense of morality or "decent respect to the opinions of mankind." ?

6. Those 3407 Votes

In the aftermath of the 2000 US Presidential Election, many people have claimed that unusually large number of 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 $\mathbf{E}[B]$ 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) Suppose that your bound b in part (b) is exactly accurate, i.e., assume that $\Pr[X \geq 3407]$ is exactly equal to b. [*In fact the true value of this probability is much smaller*] 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?

7. Median

Given a list of numbers with an odd length, the median is obtained by sorting the list and seeing which number occupies the middle position. (e.g. the median of $(1,0,1,0,2,1,2)$ is 1 because the list sorts to $(0,0,1,1,1,2,2)$ and there is a 1 in the middle. Meanwhile, the median of $(1,0,2,0,0,1,0)$ is 0 because the list sorts to $(0,0,0,0,1,1,2)$ and there is a 0 in the middle.)

[1] This sociopathic model of a selfish maximizer is referred to as a "rational agent" in the formal language of economics. Showing that cheating is not substantially attractive in the context of a mechanism even for a sociopath is one way to show that the mechanism is probably safe against normal humans too — since actual human beings are caring, loving, altruistic, and have senses of integrity and honor.

a) Consider an iid sequence of random variables X_i with probability mass function $P_X(0) = \frac{1}{3}$, $P_X(1) = \frac{1}{4}$, $P_X(2) = \frac{5}{12}$. Let M_i be the random variable that is the median of the random list $(X_1, X_2, \ldots, X_{2i+1})$. **Show that $P(M_i \neq 1)$ goes to zero exponentially fast in i.**

(HINT: What has to happen with the $\{X_j\}$ for $M_i = 0$ or $M_i = 2$?)

b) Generalize the argument you have made above to state a law of large numbers for the median of a sequence of iid discrete-valued random variables. Sketch a proof for this law.

(HINT: Let Mid(X) be that value t for which $P(X < t) < \frac{1}{2}$ and $P(X > t) < \frac{1}{2}$. If such a t doesn't exist, then don't worry about that case at all.)

8. (Optional) Best Question NA

In the game League of Legends, two champions fight each other and the one with greater prowess wins. Each champion can carry items to enhance their fighting ability. In this problem, we will do a (slightly simplified) analysis of three items in the game.

Each champion has

- H_0: Health (HP) - how much damage they can take before dying. e.g. 3000 HP
- A_0: Base attack damage (AD) - how much damage they do without any items every time they attack. e.g. 50 AD
- C_0: Base critical strike percentage (crit chance) - Every attack has a certain chance of dealing a critical strike, which doubles the amount of damage dealt. C_0 is the probability of this happening without items. e.g. 0.05 crit chance
- S_0: Base attack speed (AS) - how many times the champion attacks per second. e.g. 1.1 AS

We will analyze three common items: Infinity Edge, Bloodthirster, and Blade of the Ruined King. Their abilities are outlined below:

Infinity Edge

- +70 attack damage
- +25% critical strike percentage
- Critical strikes will deal 250% damage, instead of 200%

Bloodthirster

- +100 attack damage

Blade of the Ruined King

- +25 attack damage
- Will grant additional attack damage equal to 5% of the opponents current health.
- +0.4 attack speed

Here is an example if you are still confused as to how this all works: Suppose I am fighting against an enemy champion with 1000 HP. I have 100 base AD, 0 base crit chance, and 1 base attack speed. Without any items, it would take 10 seconds for me to defeat him. With a Bloodthirster, I would have 200 AD and would defeat him 5 seconds. With an Infinity Edge, I would have 170 AD and a critical strike (which would happen 25% of the time) would deal $170 \times 2.5 = 425$ damage. With Blade of the Ruined King, my first attack would deal $125 + 0.05(1000) = 175$ damage, leaving him with 825 HP; my second attack would deal 166.25 damage.

In order to compare the items, we will estimate our damage per second (DPS) with each of the three items. Let $H_0 = \infty, A_0, C_0, S_0$ denote our champion's statistics and H_0', A_0', C_0', S_0' denote our opponent's statistics.

(a) In terms of the above variables, what is our expected damage per second with an Infinity Edge?

(b) What is our expected damage per second with a Bloodthirster?

(c) What is our average damage per second with a Blade of the Ruined King? (Your damage will be lower after every hit against the enemy since his HP will go down. Compute the average amount of damage you deal until the enemy is dead.)

(Hint(s): First, take the average damage per second to mean the average damage dealt per second till the enemy is defeated. So, it would be H_0' divided by number of seconds to defeat the enemy.

Secondly, in order to actually calculate this, set up a recurrence relation of the enemy's health after each attack. This would be of the form $h_n = A \cdot C^n + B$ where h_n is the health after the n^{th} attack. To solve this, write the enemy's health after the $n+1^{th}$ attack as a function of their health after the n^{th} attack. Given that you know the form, plug in a couple of numbers for n and solve for A, B and C.

Finally, incorporate your attack speed back into this as you calculated it per attack. Now you have your DPS!)

(d) An item is better if it has higher expected DPS. Come up with 3 different scenarios (values for H_0, A_0, etc.) where each of the three items is the optimal choice.

For LoL players: Now you know which item is best for your AD carry in different situations! Granted, the above analysis leaves out some of the finer details (e.g. life steal, builds, actives, etc.), but as far as raw damage output goes, this is fairly accurate.

9. Write your own problem

Write your own problem related to this week's material and solve it. You may still work in groups to brainstorm problems, but each student should submit a unique problem. What is the problem? How to formulate it? How to solve it? What is the solution?

You might have noticed that this homework is shorter than usual. If you want more practice questions, please come talk to us. Enjoy your Thanksgiving!