EECS 70 Discrete Mathematics and Probability Theory
Spring 2013 Anant Sahai

HW 12

Due April 22

1. Using Chebyshev's inequality

Show that each of the following (claimed in Note 16) is true.

1. **Coin tosses.** Let X be the number of Heads in n tosses of a fair coin. The probability that X deviates from $\mu = \frac{n}{2}$ by more than $\sqrt{n}$ is at most $\frac{1}{4}$. The probability that it deviates by more than $5\sqrt{n}$ is at most $\frac{1}{100}$.
2. **Fixed points.** Let X be the number of fixed points in a random permutation of n items; recall that $\mathbf{E}[X] = \text{Var}(X) = 1$. Thus the probability that more than 10 students get their own homeworks after shuffling is at most $\frac{1}{100}$, however large n is.

2. Estimating pi

Use the following steps to estimate the value of pi:

- Generate random numbers within the unit square.
- Keep track of what percentage land within the unit circle.
- Do a little math to find an estimate of the value of pi (π).

How does your answer change as you use more and more random numbers?

3. True/false

For each of the following, determine whether the statement is true or false. If true, give a proof. If false, give a counterexample.

1. If $\mathbf{E}[XY] = \mathbf{E}[X] \times \mathbf{E}[Y]$, then X and Y are independent.
2. For independent random variables X, Y, $\mathbf{E}[XY] = \mathbf{E}[X] \times \mathbf{E}[Y]$.
3. For any random variables X, Y, $\mathbf{E}[XY] = \mathbf{E}[X] \times \mathbf{E}[Y]$.
4. Let the random variables X and Y be distributed independently and uniformly at random in the set $\{0, 1, ..., p-1\}$, where $p > 2$ is a prime. Let the random variable $S = (X+Y) \mod p$. Then $E[S] = (E[X] + E[Y]) \mod p$.

4. 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 $X_i = \text{Buchanan}$).

1. Compute the expectation $\mathbb{E}(B)$ and the variance $\text{Var}(B)$. [Hint: Notice that B has the binomial distribution!]
2. 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?
3. 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?
4. 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. Parameter Inference

We are given $x_1, x_2, \ldots, x_m$ be independent, identically distributed numbers drawn from a geometric distribution with an unknown success parameter p. By geometric distribution, we mean the distribution of the number of trials until a success (i.e. $x_i \geq 1$), as opposed to the number of failures until a success. The goal is estimate p using the data we are given.

1. In terms of p and x_i, what is the likelihood that one of the numbers x_i is generated?
2. In terms of p and the x_i, what is the likelihood that $x_1, x_2, \ldots, x_m$ are all generated?
3. For what value of p is the likelihood in part 2 maximized? (hint: prove as a lemma that x maximizes $f(x)$ if and only if x maximizes $\log f(x)$)

4. Why does this intuitively make sense?

Now instead of the geometric distribution, consider numbers $y_1, y_2, \ldots, y_m$ generated from the binomial distribution with known n and unknown p. Repeat the above four steps to provide an estimate of p.

6. When is BotRK worth it?

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
- +0.25 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 a 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 kill him. With a Bloodthirster, I would have 200 AD and would kill 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 $120 \times 2.5 = 425$ damage. With Blade of the Ruined King, my first attack would deal $25 + 0.05(1000) = 75$ damage, leaving him with 925 HP; my second attack would deal 71.25 damage.

In order to compare the items, we will estimate our damage per second (DPS) with each of the three items. Let H_0, 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.

1. In terms of the above variables, what is our expected damage per second with an Infinity Edge?
2. What is our expected damage per second with a Bloodthirster?

3. Assuming that $C_0 = 0$, 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.)
4. Using the average damage, estimate the expected DPS of Blade of the Ruined King for arbitrary C_0. (hint: assume that BotRK will deal the average amount of damage every hit).
5. 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, but as far as raw damage output goes, this is fairly accurate.

7. 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?