

Due: Thursday, November 21, 2019 at 10 PM

## Sundry

Before you start writing your final homework submission, state briefly how you worked on it. Who else did you work with? List names and email addresses. (In case of homework party, you can just describe the group.)

*Note: This homework consists of two parts. The first part (questions 1-4) will be graded and will determine your score for this homework. The second part (questions 5-6) will be graded if you submit them, but will not affect your homework score in any way. You are strongly advised to attempt all the questions in the first part. You should attempt the problems in the second part only if you are interested and have time to spare.*

---

For each problem, justify all your answers unless otherwise specified.

## Part 1: Required Problems

### 1 Probabilistic Bounds

A random variable  $X$  has variance  $\text{Var}(X) = 9$  and expectation  $\mathbb{E}[X] = 2$ . Furthermore, the value of  $X$  is never greater than 10. Given this information, provide either a proof or a counterexample for the following statements.

- (a)  $\mathbb{E}[X^2] = 13$ .
- (b)  $\mathbb{P}[X = 2] > 0$ .
- (c)  $\mathbb{P}[X \geq 2] = \mathbb{P}[X \leq 2]$ .
- (d)  $\mathbb{P}[X \leq 1] \leq 8/9$ .
- (e)  $\mathbb{P}[X \geq 6] \leq 9/16$ .

## 2 Just One Tail, Please

Let  $X$  be some random variable with finite mean and variance which is not necessarily non-negative. The *extended* version of Markov's Inequality states that for a non-negative, monotonically increasing function  $\phi(X)$  and  $\alpha > 0$ ,

$$\mathbb{P}(X \geq \alpha) \leq \frac{\mathbb{E}[\phi(X)]}{\phi(\alpha)}$$

Suppose  $\mathbb{E}[X] = 0$ ,  $\text{Var}(X) = \sigma^2 < \infty$ , and  $\alpha > 0$ .

- (a) Use the extended version of Markov's Inequality stated above with  $\phi(x) = (x + c)^2$ , where  $c$  is some constant, to show that:

$$\mathbb{P}(X \geq \alpha) \leq \frac{\sigma^2 + c^2}{(\alpha + c)^2}$$

- (b) Note that the above bound applies for all positive  $c$ , so we can choose a value of  $c$  to minimize the expression, yielding the best possible bound. Find the value for  $c$  which will minimize the RHS expression (you may assume that the expression has a unique minimum). Plug in the minimizing value of  $c$  to prove the following bound:

$$\mathbb{P}(X \geq \alpha) \leq \frac{\sigma^2}{\alpha^2 + \sigma^2}.$$

- (c) Recall that Chebyshev's inequality provides a two-sided bound. That is, it provides a bound on  $\mathbb{P}(|X - \mathbb{E}[X]| \geq \alpha) = \mathbb{P}(X \geq \mathbb{E}[X] + \alpha) + \mathbb{P}(X \leq \mathbb{E}[X] - \alpha)$ . If we only wanted to bound the probability of one of the tails, e.g. if we wanted to bound  $\mathbb{P}(X \geq \mathbb{E}[X] + \alpha)$ , it is tempting to just divide the bound we get from Chebyshev's by two. Why is this not always correct in general? Provide an example of a random variable  $X$  (does not have to be zero-mean) and a constant  $\alpha$  such that using this method (dividing by two to bound one tail) is not correct, that is,  $\mathbb{P}(X \geq \mathbb{E}[X] + \alpha) \geq \frac{\text{Var}(X)}{2\alpha^2}$  or  $\mathbb{P}(X \leq \mathbb{E}[X] - \alpha) \geq \frac{\text{Var}(X)}{2\alpha^2}$ .
- (d) What is a sufficient condition to place on a random variable  $X$  such that dividing by two to bound one tail would be correct?

Now we see the use of the bound proven in part (b) - it allows us to bound just one tail while still taking variance into account, and does not require us to assume any property of the random variable. Note that the bound is also always guaranteed to be less than 1 (and therefore at least somewhat useful), unlike Markov's and Chebyshev's inequality!

- (e) Let's try out our new bound on a simple example. Suppose  $X$  is a positively-valued random variable with  $\mathbb{E}[X] = 3$  and  $\text{Var}(X) = 2$ . What bound would Markov's inequality give for  $\mathbb{P}[X \geq 5]$ ? What bound would Chebyshev's inequality give for  $\mathbb{P}[X \geq 5]$ ? What about for the bound we proved in part (b)? (*Note:* Recall that the bound from part (b) only applies for zero-mean random variables.)

### 3 Optimal Gambling

Jonathan has a coin that may be biased, but he doesn't think so. You disagree with him though, and he challenges you to a bet. You start off with  $X_0$  dollars. You and Jonathan then play multiple rounds, and each round, you bet an amount of money of your choosing, and then coin is tossed. Jonathan will match your bet, no matter what, and if the coin comes up heads, you win and you take both yours and Jonathan's bet, and if it comes up tails, then you lose your bet.

- (a) Now suppose you actually secretly know that the bias of the coin is  $\frac{1}{2} < p < 1$ ! You use the following strategy: on each round, you will bet a fraction  $q$  of the money you have at the start of the round. Let's say you play  $n$  rounds. What is the probability that you win exactly  $k$  of the rounds? What is the amount of money you would have if you win exactly  $k$  rounds? [Hint: Does the order in which you win the games affect your profit?]
- (b) Let  $X_n$  denote the amount of money you have on round  $n$ .  $X_0$  represents your initial assets and is a constant value. Show that  $\mathbb{E}[X_n] = ((1-p)(1-q) + p(1+q))^n X_0$ .

You may use the binomial theorem in your answer:

$$(a+b)^n = \sum_{k=0}^n \binom{n}{k} a^k b^{(n-k)}$$

[Hint: Try computing a sum over the number of rounds you win out of the  $n$  rounds you play - use your answers from the previous part.]

- (c) What value of  $q$  will maximize  $\mathbb{E}[X_n]$ ? (Use calculus!) For this value of  $q$ , what is the distribution of  $X_n$ ? Can you predict what will happen as  $n \rightarrow \infty$ ? [Hint: Under this betting strategy, what happens if you ever lose a round?]
- (d) The problem with the previous approach is that we were too concerned about expected value, so our gambling strategy was too extreme. Let's start over: again we will use a gambling strategy in which we bet a fraction  $q$  of our money at each round. Express  $X_n$  in terms of  $n$ ,  $q$ ,  $X_0$ , and  $W_n$ , where  $W_n$  is the number of rounds you have won up until round  $n$ . [Hint: Does the order in which you win the games affect your profit?]
- (e) By the law of large numbers, what does  $W_n/n$  converge to as  $n \rightarrow \infty$ ? Using this fact, what does  $(\ln X_n)/n$  converge to as  $n \rightarrow \infty$ ?
- (f) The rationale behind  $(\ln X_n)/n$  is that if  $(\ln X_n)/n \rightarrow c$ , where  $c$  is a constant, then that means for large  $n$ ,  $X_n$  is roughly  $e^{cn}$ . Therefore,  $c$  is the asymptotic growth rate of your fortune! Find the value of  $q$  that maximizes your asymptotic growth rate.
- (g) Using the value of  $q$  you found in the previous part, compute  $\mathbb{E}[X_n]$ .
- (h) Say Jonathan wishes to estimate the bias of the coin - he may want to use the value  $W_n/n$  as his estimate. What is the expectation of this value? What is a bound on the variance of this value? (Your bound should not include  $p$ .)

- (i) Let's say after playing 100 rounds, Jonathan observes that 74 heads have appeared. Help Jonathan construct a 75% confidence interval using Chebyshev's inequality on the estimator from the previous part.

## 4 Random Cuckoo Hashing

Cuckoo birds are parasitic beasts. They are known for hijacking the nests of other bird species and evicting the eggs already inside. Cuckoo hashing is inspired by this behavior. In cuckoo hashing, when we get a collision, the element that was already there gets evicted and rehashed.

We study a simple (but ineffective, as we'll see) version of cuckoo hashing, where all hashes are random. Let's say we want to hash  $n$  pieces of data  $d_1, d_2, \dots, d_n$  into  $n$  possible hash buckets labeled  $1, \dots, n$ . We hash the  $d_1, \dots, d_n$  in that order. When hashing  $d_i$ , we assign it a random bucket chosen uniformly from  $1, \dots, n$ . If there is no collision, then we place  $d_i$  into that bucket. If there is a collision with some other  $d_j$ , we evict  $d_j$  and assign it another random bucket uniformly from  $1, \dots, n$ . (It is possible that  $d_j$  gets assigned back to the bucket it was just evicted from!) We again perform the eviction step if we get another collision. We keep doing this until there is no more collision, and we then introduce the next piece of data,  $d_{i+1}$  to the hash table.

- (a) What is the probability that there are no collisions over the entire process of hashing  $d_1, \dots, d_n$  to buckets  $1, \dots, n$ ? What value does the probability tend towards as  $n$  grows very large?
- (b) Assume we have already hashed  $d_1, \dots, d_{n-1}$ , and they each occupy their own bucket. We now introduce  $d_n$  into our hash table. What is the expected number of collisions that we'll see while hashing  $d_n$ ? (*Hint*: What happens when we hash  $d_n$  and get a collision, so we evict some other  $d_i$  and have to hash  $d_i$ ? Are we at a situation that we've seen before?)
- (c) Generalize the previous part: Assume we have already hashed  $d_1, \dots, d_{k-1}$  successfully, where  $1 \leq k \leq n$ . Let  $C_k$  be the number of collisions that we'll see while hashing  $d_k$ . What is  $\mathbb{E}[C_k]$ ?
- (d) Let  $C$  be the total number of collisions over the entire process of hashing  $d_1, \dots, d_n$ . What is  $\mathbb{E}[C]$ ? You may leave your answer as a summation.

**Note:** This concludes the first part of the homework. The problems below are optional, will not affect your score, and should be attempted only if you have time to spare.

---

## Part 2: Optional Problems

### 5 Subset Card Game

Jonathan and Yiming are playing a card game. Jonathan has  $k > 2$  cards, and each card has a real number written on it. Jonathan tells Yiming (truthfully), that the sum of the card values is

0, and that the sum of squares of the values on the cards is 1. Specifically, if the card values are  $c_1, c_2, \dots, c_k$ , then we have  $\sum_{i=1}^k c_i = 0$  and  $\sum_{i=1}^k c_i^2 = 1$ . Jonathan and Yiming also agree on a positive target value of  $\alpha$ .

The cards are then going to be dealt randomly in the following fashion: for each card in the deck, a fair coin is flipped. If the coin lands heads, then the card goes to Yiming, and if the coin lands tails, the card goes to Jonathan. Note that it is possible for either player to end up with no cards/all the cards.

A player wins the game if the sum of the card values in their hand is at least  $\alpha$ , otherwise it is a tie.

- (a) Prove that the probability that Yiming wins is at most  $\frac{1}{8\alpha^2}$ .
- (b) Find a deck of  $k$  cards and target value  $\alpha$  where the probability that Yiming wins is exactly  $\frac{1}{8\alpha^2}$ .

## 6 Easy A's

A friend tells you about a course called “Laziness in Modern Society” that requires almost no work. You hope to take this course next semester to give yourself a well-deserved break after working hard in CS 70. At the first lecture, the professor announces that grades will depend only on two homework assignments. Homework 1 will consist of three questions, each worth 10 points, and Homework 2 will consist of four questions, also each worth 10 points. He will give an A to any student who gets at least 60 of the possible 70 points.

However, speaking with the professor in office hours you hear some very disturbing news. He tells you that, in the spirit of the class, the GSIs are very lazy, and to save time the grading will be done as follows. For each student's Homework 1, the GSIs will choose an integer randomly from a distribution with mean  $\mu = 5$  and variance  $\sigma^2 = 1$ . They'll mark each of the three questions with that score. To grade Homework 2, they'll again choose a random number from the same distribution, independently of the first number, and will mark all four questions with that score.

If you take the class, what will the mean and variance of your total class score be? Use Chebyshev's inequality to conclude that you have less than a 5% chance of getting an A when the grades are randomly chosen this way.