

Due: Sunday 8/8, 10:00 PM  
Grace period until Sunday 8/8, 11:59 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.)

## 1 MGFs

Given a random variable  $X$ , the *moment generating function* of  $X$  is defined as the function  $M_X(t) = \mathbb{E}[e^{tX}]$ . Moment generating functions, or MGFs for short, are immensely useful because of the Taylor expansion

$$e^{tX} = 1 + tX + \frac{(tX)^2}{2!} + \frac{(tX)^3}{3!} + \dots = \sum_{n=0}^{\infty} \frac{(tX)^n}{n!}.$$

By taking the  $k$ th derivative of the MGF of  $X$  with respect to  $t$  and evaluating at  $t = 0$ , we can generate the  $k$ th moment of  $X$ , (i.e. the value of  $\mathbb{E}[X^k]$ ) without having to do any painful integration!

- Compute the moment generating function  $M_X(t)$  of  $X$ , where  $X \sim \text{Expo}(\lambda)$ , for  $t < \lambda$ .
- Suppose now that  $X \sim \text{Expo}(\lambda)$ , and further suppose that  $Y \sim \text{Poisson}(\mu)$ , where  $X$  and  $Y$  are independent and  $\mu < \lambda$ . Compute  $\mathbb{E}[X^Y]$ . (Hint: use conditional expectation and your answer in part (a))

## 2 Functions of Normals

Let  $Z \sim \text{Normal}(0, 1)$ .

- Let  $V = |Z|$ .
  - Find the cdf of  $V$  in terms of the standard normal cdf  $\Phi$ .
  - Find the pdf of  $V$  in terms of the standard normal pdf  $\phi$ .
- Let  $W = e^Z$ .
  - Find the cdf of  $W$  in terms of the standard normal cdf  $\Phi$ .
  - Find the pdf of  $W$  in terms of the standard normal pdf  $\phi$ .

### 3 Joint Practice

Suppose that  $X$  and  $Y$  are random variables with joint density

$$f_{X,Y}(x,y) = \begin{cases} Ax^2y^2 & \text{if } 0 \leq x \leq 1 \text{ and } 0 \leq y \leq 1, \\ 0 & \text{otherwise,} \end{cases}$$

where  $A$  is a positive constant.

- (a) What is the value of  $A$ ?
- (b) What is the marginal density of  $X$ ?
- (c) What is  $\text{cov}(X, Y)$ ?

### 4 Tightness of Inequalities

- (a) Show by example that Markov's inequality is tight; that is, show that given some fixed  $k > 0$ , there exists a discrete non-negative random variable  $X$  such that  $\mathbb{P}(X \geq k) = \mathbb{E}[X]/k$ .
- (b) Show by example that Chebyshev's inequality is tight; that is, show that given some fixed  $k \geq 1$ , there exists a random variable  $X$  such that  $\mathbb{P}(|X - \mathbb{E}[X]| \geq k\sigma) = 1/k^2$ , where  $\sigma^2 = \text{Var} X$ .

### 5 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 function  $\phi(x)$  which is monotonically increasing for  $x > 0$  and some constant  $\alpha > 0$ ,

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

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 positive 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).

We can plug in the minimizing value of  $c$  you found in part (b) to prove the following bound:

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

This bound is also known as Cantelli's inequality.

(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.

- (i) Why is this not always correct in general?
- (ii) 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,  $\mathbb{P}(X \geq \mathbb{E}[X] + \alpha) > \frac{\text{Var}(X)}{2\alpha^2}$  or  $\mathbb{P}(X \leq \mathbb{E}[X] - \alpha) > \frac{\text{Var}(X)}{2\alpha^2}$ .

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!

(d) 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  $\text{Var}(X) = 2$ .

- (i) What bound would Markov's inequality give for  $\mathbb{P}[X \geq 5]$ ?
- (ii) What bound would Chebyshev's inequality give for  $\mathbb{P}[X \geq 5]$ ?
- (iii) What bound would Cantelli's Inequality give for  $\mathbb{P}[X \geq 5]$ ? (*Note: Recall that Cantelli's Inequality only applies for zero-mean random variables.*)

## 6 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. Prove that the probability that Yiming wins is at most  $\frac{1}{8\alpha^2}$ .

## 7 Playing Pollster

As an expert in probability, the staff members at the Daily Californian have recruited you to help them conduct a poll to determine the percentage  $p$  of Berkeley undergraduates that plan to participate in the student sit-in. They've specified that they want your estimate  $\hat{p}$  to have an error of at most  $\varepsilon$  with confidence  $1 - \delta$ . That is,

$$\mathbb{P}(|\hat{p} - p| \leq \varepsilon) \geq 1 - \delta.$$

Recall from lecture and the notes that you have the bound

$$\mathbb{P}(|\hat{p} - p| \geq \varepsilon) \leq \frac{1}{4n\varepsilon^2},$$

where  $n$  is the number of students in your poll.

- Using the formula above, what is the smallest number of students  $n$  that you need to poll so that your poll has an error of at most  $\varepsilon$  with confidence  $1 - \delta$ ?
- At Berkeley, there are about 26,000 undergraduates and about 10,000 graduate students. Suppose you only want to understand the frequency of sitting-in for the undergraduates. If you want to obtain an estimate with error of at most 5% with 98% confidence, how many undergraduate students would you need to poll? Does your answer change if you instead only want to understand the frequency of sitting-in for the graduate students?
- It turns out you just don't have as much time for extracurricular activities as you thought you would this semester. The writers at the Daily Californian insist that your poll results are reported with at least 95% confidence, but you only have enough time to poll 500 students. Based on the bound above, what is the smallest error with which you can report your results and still ensure you have at least 95% confidence?

## 8 Uniform Estimation

Let  $U_1, \dots, U_n \stackrel{\text{iid}}{\sim} \text{Uniform}(-\theta, \theta)$  for some unknown  $\theta \in \mathbb{R}$ ,  $\theta > 0$ . We wish to estimate  $\theta$  from the data  $U_1, \dots, U_n$ .

- Why would using the sample mean  $\bar{U} = \frac{1}{n} \sum_{i=1}^n U_i$  fail in this situation?
- Find the PDF of  $U_i^2$  for  $i \in \{1, \dots, n\}$ .
- Consider the following variance estimate:

$$V = \frac{1}{n} \sum_{i=1}^n U_i^2.$$

Show that for large  $n$ , the distribution of  $V$  is close to one of the famous ones, and provide its name and parameters.

- (d) Use part (c) to construct an unbiased estimator for  $\theta^2$  that uses all the data.
- (e) Let  $\sigma^2 = \text{Var}[U_i^2]$ . We wish to construct a confidence interval for  $\theta^2$  with a significance level of  $\delta$ , where  $0 < \delta < 1$ .
- (i) Without any assumption on the magnitude of  $n$ , construct a confidence interval for  $\theta^2$  with a significance level of  $\delta$  using your estimator from part (d).
  - (ii) Suppose  $n$  is large. Construct an approximate confidence interval for  $\theta^2$  with a significance level of  $\delta$  using your estimator from part (d). You may leave your answer in terms of  $\Phi$  and  $\Phi^{-1}$ , the normal CDF and its inverse.