CS 70 Discrete Mathematics and Probability Theory
Summer 2021 **HW 6**

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

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

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

(a) Compute the moment generating function $M_X(t)$ of X, where $X \sim \text{Expo}(\lambda)$, for $t < \lambda$.

(b) 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 $\mathbb{E}\left[X^Y\right]$. (Hint: use conditional expectation and your answer in part (a))

Solution:

(a) We have that

$$\begin{aligned}
M_X(t) &= \mathbb{E}[e^{tX}] \\
&= \int_0^{\infty} e^{tx} \cdot \lambda e^{-\lambda x} \, dx \\
&= \int_0^{\infty} \lambda e^{(t-\lambda)x} \, dx \\
&= \boxed{\frac{\lambda}{\lambda - t}}.
\end{aligned}$$

Note that in the final step, we used the fact that $t - \lambda < 0$ to ensure that the integral converged.

(b) We proceed by using conditional expectation, conditioning on the value of Y. We have that

$$\begin{aligned}
\mathbb{E}\left[X^Y\right] &= \sum_{y=0}^{\infty} \mathbb{E}\left[X^Y \mid Y = y\right] \mathbb{P}[Y = y] \\
&= \sum_{y=0}^{\infty} \mathbb{E}[X^y] \cdot e^{-\mu} \frac{\mu^y}{y!}.
\end{aligned}$$

Now, we can use linearity of expectation in reverse to get that

$$\mathbb{E}\left[X^Y\right] = \mathbb{E}\left[\sum_{y=0}^{\infty} X^y \cdot e^{-\mu}\frac{\mu^y}{y!}\right] = e^{-\mu}\,\mathbb{E}\left[\sum_{y=0}^{\infty}\frac{(\mu X)^y}{y!}\right].$$

Finally, we observe that this is exactly the Taylor series for $e^{\mu X}$, so we may simplify

$$e^{-\mu}\,\mathbb{E}\left[\sum_{y=0}^{\infty}\frac{(\mu X)^y}{y!}\right] = e^{-\mu}\,\mathbb{E}\left[e^{\mu X}\right] = e^{-\mu}M_X(\mu),$$

where $M_X(t)$ is the moment generating function of X, which we derived in part (a). This at last becomes

$$e^{-\mu}M_X(\mu) = \boxed{\frac{\lambda e^{-\mu}}{\lambda-\mu}},$$

which is well-defined as in the problem we assume that $\lambda > \mu$.

Alternate Solution: We proceed as before up to the expression

$$\mathbb{E}\left[X^Y\right] = \sum_{y=0}^{\infty}\mathbb{E}[X^y]\cdot e^{-\mu}\frac{\mu^y}{y!},$$

but we instead compute each of these moments explicitly using the MGF $M_X(t)$. The yth moment of X is given by the yth derivative of M_X, evaluated at $t=0$. We claim that for any positive integer y,

$$\frac{d^y}{dt^y}\frac{\lambda}{\lambda-t} = \frac{y!\lambda}{(\lambda-t)^{y+1}}.$$

We will prove this by inducting on y. Firstly, for the case $y=1$, we have that

$$\frac{d}{dt}\frac{\lambda}{\lambda-t} = \frac{\lambda}{(\lambda-t)^2} = \frac{1!\lambda}{(\lambda-t)^{1+1}},$$

so the claim holds in this case. Now, suppose the claim holds for some $y=n$. Then, we have that

$$\frac{d^{n+1}}{dt^{n+1}}\frac{\lambda}{\lambda-t} = \frac{d}{dt}\left(\frac{d^n}{dt^n}\frac{\lambda}{\lambda-t}\right) = \frac{d}{dt}\frac{n!\lambda}{(\lambda-t)^{n+1}} = \frac{(n+1)!\lambda}{(\lambda-t)^{n+2}},$$

hence the claim holds for $y=n+1$, completing the induction. With this, we can now continue the problem by computing the yth moment using the MGF:

$$\begin{aligned}
\mathbb{E}[X^y] = M_X^{[y]}(0) &= \left.\frac{d^y}{dt^y}\right|_{t=0}\frac{\lambda}{\lambda-t} \\
&= \left.\frac{y!\lambda}{(\lambda-t)^{y+1}}\right|_{t=0} \\
&= \frac{y!}{\lambda^y}.
\end{aligned}$$

Plugging this in to our original expression, we get that

$$\sum_{y=0}^{\infty} \mathbb{E}[X^y] \cdot e^{-\mu} \frac{\mu^y}{y!} = \sum_{y=0}^{\infty} \frac{y!}{\lambda^y} \cdot e^{-\mu} \frac{\mu^y}{y!} = e^{-\mu} \sum_{y=0}^{\infty} \left(\frac{\mu}{\lambda}\right)^y.$$

Since $\lambda > \mu$, the ratio $\mu/\lambda < 1$, so this geometric series converges and our final answer is

$$\mathbb{E}\left[X^Y\right] = e^{-\mu} \sum_{y=0}^{\infty} \left(\frac{\mu}{\lambda}\right)^y = \frac{e^{-\mu}}{1 - \frac{\mu}{\lambda}} = \boxed{\frac{\lambda e^{-\mu}}{\lambda - \mu}}.$$

2 Functions of Normals

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

(a) Let $V = |Z|$.

(i) Find the cdf of V in terms of the standard normal cdf Φ.

(ii) Find the pdf of V in terms of the standard normal pdf ϕ.

(b) Let $W = e^Z$.

(i) Find the cdf of W in terms of the standard normal cdf Φ.

(ii) Find the pdf of W in terms of the standard normal pdf ϕ.

Solution:

(a) To find the CDF of V, we have to figure out what $\mathbb{P}[V \le t]$ is for all values of t. Firstly, if $t < 0$, then $\mathbb{P}[V \le t] = 0$, as V is by definition always nonnegative. Now, if $t \ge 0$, we have that

$$\mathbb{P}[V \le t] = \mathbb{P}[|Z| \le t] = \mathbb{P}[-t \le Z \le t] = \Phi(t) - \Phi(-t) = 2\Phi(t) - 1,$$

thus our CDF of V is

$$F_V(t) = \begin{cases} 2\Phi(t) - 1 & \text{if } t \ge 0 \text{, and} \\ 0 & \text{otherwise.} \end{cases}$$

To find the PDF from here, we just need to take the derivative, keeping in mind that $\Phi'(t) = \phi(t)$ and the chain rule. We have then that

$$f_V(t) = F_V'(t) = \begin{cases} 2\phi(t) & \text{if } t \ge 0 \text{, and} \\ 0 & \text{otherwise.} \end{cases}$$

Note that the value of the PDF at the discontinuity at $t = 0$ can also be 0.

(b) We proceed in a similar fashion to the previous part. Since the function e^x is strictly positive, we have that for any $t \le 0$, $\mathbb{P}[W \le t] = 0$. For $t > 0$, we can see that

$$\mathbb{P}[W \le t] = \mathbb{P}[e^Z \le t] = \mathbb{P}[Z \le \ln(t)] = \Phi(\ln(t)),$$

hence our CDF is

$$F_W(t) = \begin{cases} \Phi(\ln(t)) & \text{if } t > 0, \text{ and} \\ 0 & \text{otherwise.} \end{cases}$$

Likewise, we can find the PDF by taking the derivative:

$$f_W(t) = F_W'(t) = \begin{cases} \frac{1}{t}\phi(\ln(t)) & \text{if } t > 0, \text{ and} \\ 0 & \text{otherwise.} \end{cases}$$

3 Joint Practice

Suppose that X and Y are random variables with joint density

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

where A is a positive constant.

(a) What is the value of A?

(b) What is the marginal density of X?

(c) What is $\text{cov}(X,Y)$?

Solution:

(a) Since $f_{X,Y}$ is a joint density, we know that it must integrate to 1. Since $f_{X,Y}$ is only nonzero on the unit square, we can set up and solve the following integral:

$$1 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx\,dy = \int_0^1\int_0^1 Ax^2y^2\,dx\,dy = \frac{1}{9}\cdot A,$$

hence, we can see that $\boxed{A = 9}$.

(b) Since the joint density can only be nonzero when X is between 0 and 1, we know that outside of this interval, the marginal density of X must also be zero. Inside this interval, we can find the marginal density of X by integrating the joint density with respect to Y :

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy = \int_0^1 9x^2y^2\,dy = 3x^2.$$

Thus, we can see that the marginal density of X is given by

$$f_X(x) = \begin{cases} 3x^2 & \text{if } 0 \le x \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

As a sanity check, one can see that this function is both nonnegative and integrates to 1.

(c) There are two ways of approaching this. The first way is to use the fact that

$$\operatorname{cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y].$$

In order to apply this formula, we need to first find these values. We first find that

$$\mathbb{E}[XY] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy f_{X,Y}(x,y)\,dx\,dy = \int_0^1\int_0^1 9x^3y^3\,dx\,dy = \frac{9}{16}.$$

Moreover, we can compute that

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x)\,dx = \int_0^1 3x^3\,dx = \frac{3}{4}.$$

Finally, in order to compute $\mathbb{E}[Y]$, we first find the marginal density of Y. We can do this in a similar fashion to the previous part by integrating the joint density with respect to X. Since the joint density is zero when Y is not between 0 and 1, we know that the marginal density of Y must also be zero outside of this interval. When Y is inside this interval, we have that

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx = \int_0^1 9x^2y^2\,dx = 3y^2,$$

hence the full marginal density is

$$f_Y(y) = \begin{cases} 3y^2 & \text{if } 0 \le y \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

This then allows to compute

$$\mathbb{E}[Y] = \int_{-\infty}^{\infty} y f_Y(y)\,dy = \int_0^1 3y^3\,dy = \frac{3}{4},$$

hence

$$\operatorname{cov}(X,Y) = \frac{9}{16} - \frac{3}{4}\cdot\frac{3}{4} = \boxed{0}.$$

The second way of approaching this problem is to first compute the marginal density of Y. Upon doing so, one can check that for any pair of values x, y, we have that

$$f_{X,Y}(x,y) = f_X(x)\cdot f_Y(y),$$

meaning that X and Y are independent random variables. Thus, since the covariance between two independent random variables is always zero, we can conclude the desired result and finish.

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 $\sigma^2 = \operatorname{Var} X$.

Solution:

(a) In the proof of Markov's Inequality ($\mathbb{P}[X \geq \alpha] \leq \frac{\mathbb{E}[X]}{\alpha}$), the first time we lose equality is at this step:

$$\mathbb{E}[X] = \sum_{a}(a \cdot \mathbb{P}[X = a]) \geq \sum_{a \geq \alpha}(a \cdot \mathbb{P}[X = a])$$

We get an inequality because we drop all $a \cdot \mathbb{P}[X = a]$ terms where $a < \alpha$. Thus, we can only maintain equality if all of these dropped terms were actually 0. This would mean either $a = 0$ or $\mathbb{P}[X = a] = 0$ for an $a > 0$, which means X can put probability on 0, but should put no probability on any other value $< \alpha$.

The next time we lose equality in the proof is the step following the one above:

$$\sum_{a \geq \alpha}(a \cdot \mathbb{P}[X = a]) \geq \alpha \cdot \sum_{a \geq \alpha} \mathbb{P}[X = a]$$

We get an inequality because we treat all $a \geq \alpha$ in the summation as just α, so we can pull out the α term. The only way for us to maintain equality is if we never have to substitute α for some larger a. This tells us that X should not put probability on any value $> \alpha$.

Both of these facts drive the intuition behind our example: that X can only take values 0 and α.

Let X be the random variable which is 0 with probability $1 - p$ and k with probability p, where $k > 0$. Then, $\mathbb{E}[X] = kp$, and Markov's inequality says

$$\mathbb{P}(X \geq k) \leq \frac{\mathbb{E}[X]}{k} = \frac{kp}{k} = p,$$

which is tight.

(b) The proof of Chebyshev's Inequality ($\mathbb{P}[|X - \mathbb{E}[X]| \geq \alpha] \leq \frac{\operatorname{Var}(X)}{\alpha^2}$) comes from an application of Markov's Inequality to the variable $Y = (X - \mathbb{E}[X])^2$ being $\geq \alpha^2$. The only ways we can lose equality in the proof of Chebyshev's is if we lose equality in the application of Markov! Therefore, we need the variable Y to satisfy the conditions from Part (a) that ensure the application of Markov will be tight. To recap, we would need Y to only take values 0 and α^2. Thus, $(X - \mathbb{E}[X])$ can take on the values $\{-\alpha, 0, \alpha\}$.

Let

$$X = \begin{cases} -a & \text{with probability } k^{-2}/2 \\ a & \text{with probability } k^{-2}/2 \\ 0 & \text{with probability } 1 - k^{-2} \end{cases}$$

for $a > 0$. Note that $\operatorname{Var} X = a^2 k^{-2}$, so $k\sigma = a$, so Chebyshev's inequality gives

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq k\sigma) = \mathbb{P}(|X - \mathbb{E}[X]| \geq a) \leq \frac{1}{k^2},$$

which is tight.

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 $\mathbb{E}[X] = 0$, $\operatorname{Var}(X) = \sigma^2 < \infty$, 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 α such that using this method (dividing by two to bound one tail) is not correct, that is, $\mathbb{P}(X \geq \mathbb{E}[X] + \alpha) > \frac{\operatorname{Var}(X)}{2\alpha^2}$ or $\mathbb{P}(X \leq \mathbb{E}[X] - \alpha) > \frac{\operatorname{Var}(X)}{2\alpha^2}$.

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.)

Solution:

(a) Note that $\sigma^2 = \text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \mathbb{E}[X^2]$. Using the inequality presented in the problem, we have:

$$\mathbb{P}(X \geq \alpha) \leq \frac{\mathbb{E}[(X+c)^2]}{(\alpha+c)^2} = \frac{\mathbb{E}[X^2+2cX+c^2]}{(\alpha+c)^2} = \frac{\mathbb{E}[X^2]+2c\,\mathbb{E}[X]+c^2}{(\alpha+c)^2} = \frac{\sigma^2+c^2}{(\alpha+c)^2}$$

(b) We set the derivative with respect to c of the above expression equal to 0, and solve for c.

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}c}\frac{\sigma^2+c^2}{(\alpha+c)^2} &= 0 \\
\frac{2c(\alpha+c)^2 - 2(\alpha+c)(\sigma^2+c^2)}{(\alpha+c)^4} &= 0 \\
2c(\alpha+c)^2 - 2(\alpha+c)(\sigma^2+c^2) &= 0 \\
\alpha c^2 + (\alpha^2 - \sigma^2)c - \sigma^2\alpha &= 0 \\
c &= \frac{\sigma^2}{\alpha}
\end{aligned}$$

To get the last step we use the quadratic equation and take the positive solution.

(c) It is possible for one of the tails to contain more probability than the other. One example of a random variable which demonstrates this is X, where $\mathbb{P}(X = 0) = 0.75$ and $\mathbb{P}(X = 10) = 0.25$, with $\alpha = 7$. Here, $\mathbb{E}[X] = 2.5$ and $\text{Var}(X) = 100 \cdot 0.25 \cdot 0.75$, so we have:

$$\mathbb{P}(X \geq \mathbb{E}[X] + 7) = 0.25 > \frac{\text{Var}(X)}{2 \cdot 7^2} \approx 0.19$$

(d) Using Markov's: $\mathbb{P}(X \geq 5) \leq \frac{\mathbb{E}[X]}{5} = \frac{3}{5}$

Using Chebyshev's: $\mathbb{P}(X \geq 5) \leq \mathbb{P}(|X - \mathbb{E}[X]| \geq 2) \leq \frac{\text{Var}(X)}{2^2} = \frac{1}{2}$

Using bound shown above (Cantelli's):

Since we have the condition that this bound applies to zero-mean random variables, let us define $Y = X - \mathbb{E}[X] = X - 3$. Note that $\text{Var}(Y) = \text{Var}(X)$.

Then we get: $\mathbb{P}(X \geq 5) = \mathbb{P}(Y \geq 2) \leq \frac{\text{Var}(Y)}{2^2 + \text{Var}(Y)} = \frac{1}{3}$.

We see that Cantelli's inequality (the bound from part (b)) does better than Chebyshev's, which does better than Markov's (note that having a smaller upper bound is better)! This is a good demonstration on how we might derive better bounds using Markov's inequality, if we know further information about the random variable like its variance.

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 $c_1, c_2, \ldots, c_k$, 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 α.

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 α, otherwise it is a tie. Prove that the probability that Yiming wins is at most $\frac{1}{8\alpha^2}$.

Solution:

Let I_i be the indicator random variable indicating whether or not card i goes to Yiming. Define $S = \sum_{i=1}^{k} c_i I_i$ as the value of Yiming's hand. Then, we see that $\mathbb{E}[S] = \sum_{i=1}^{k} c_i \cdot \frac{1}{2} = 0$ and

$$\begin{aligned} \text{Var}(S) &= \sum_{i=1}^{k} \text{Var}(c_i I_i) \qquad \text{(due to independence) of } I_i \\ &= \sum_{i=1}^{k} c_i^2 \,\text{Var}(I_i) \end{aligned}$$

We know that I_i is a Bernoulli random variable, so its variance is $\frac{1}{4}$. Thus, we see that $\text{Var}(S) = \frac{1}{4}$.

By Chebyshev, we see that $\mathbb{P}(|S| \geq \alpha) \leq \frac{1}{4\alpha^2}$. Now we need to make a symmetry argument, specifically that for each value of x, $\mathbb{P}(S = x) = \mathbb{P}(S = -x)$. This is true because for each outcome where Yiming gets x, Jonathan gets $-x$, since the sum of the card values is 0. However, we also know that the reverse outcome, where Jonathan gets Yiming's cards and vice versa, has the same probability of happening.

Since the distribution of S is symmetric around 0, we see that $\mathbb{P}(|S| \geq \alpha) = 2\mathbb{P}(S \geq \alpha)$, and plugging into our bound yields $\mathbb{P}(S \geq \alpha) \leq \frac{1}{8\alpha^2}$.

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 ε 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.

(a) 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 ε with confidence $1-\delta$?

(b) 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?

(c) 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?

Solution:

(a) We know we need to have

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

Subtracting both sides from 1, it follows that we must have

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

Therefore if we choose n such that

$$\frac{1}{4n\varepsilon^2} \leq \delta,$$

we will have

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

and since $\mathbb{P}(|\hat{p}-p| > \varepsilon) \leq \mathbb{P}(|\hat{p}-p| \geq \varepsilon)$, this will meet the requirement that

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

Thus we must have that

$$\begin{aligned} \frac{1}{4n\varepsilon^2} &\leq \delta \\ \frac{1}{n} &\leq 4\varepsilon^2\delta \\ n &\geq \frac{1}{4\varepsilon^2\delta}. \end{aligned}$$

(b) Plugging in $\varepsilon = 0.05$ (our maximum error) and $\delta = 0.02$ (probability of being off by at least this error) to the bound you found above, you get that $n \geq 5000$. The answer is the same for graduate students; the size of the population does not affect the number of samples you need.

(c) If you only have time to poll 500 people and want to report your results with 95% confidence, you must report that the error in your estimate is at most 10%. You can find this by plugging in $1/(4 \cdot 500 \cdot \varepsilon^2) = .05$ and solving for ε.

8 Uniform Estimation

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

(a) Why would using the sample mean $\overline{U} = \frac{1}{n}\sum_{i=1}^{n} U_i$ fail in this situation?

(b) Find the PDF of U_i^2 for $i \in \{1, \ldots, n\}$.

(c) 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 θ^2 that uses all the data.

(e) Let $\sigma^2 = \text{Var}[U_i^2]$. We wish to construct a confidence interval for θ^2 with a significance level of δ, where $0 < \delta < 1$.

(i) Without any assumption on the magnitude of n, construct a confidence interval for θ^2 with a significance level of δ using your estimator from part (d).

(ii) Suppose n is large. Construct an approximate confidence interval for θ^2 with a significance level of δ using your estimator from part (d). You may leave your answer in terms of Φ and Φ^{-1}, the normal CDF and its inverse.

Solution:

(a) The sample mean would not work well as an estimator for θ because it has expected value 0, not θ.

(b) We will proceed by finding the CDF of U_i^2 first, and then taking the derivative after to get the PDF. Firstly, note that $0 \le U_i^2 \le \theta^2$, so we have that $\mathbb{P}[U_i^2 \le t] = 0$ when $t \le 0$ and $\mathbb{P}[U_i^2 \le t] = 1$ when $t \ge \theta^2$. When $0 < t < \theta^2$, we have that

$$\mathbb{P}[U_i^2 \le t] = \mathbb{P}[-\sqrt{t} \le U_i \le \sqrt{t}] = \frac{\sqrt{t}}{\theta},$$

hence the CDF of U_i^2 is

$$F(t) = \begin{cases} 0 & \text{if } t \le 0, \\ \frac{\sqrt{t}}{\theta} & \text{if } 0 < t < \theta^2, \text{ and} \\ 1 & \text{if } t \ge \theta^2. \end{cases}$$

Lastly, we take the derivative to get the PDF:

$$f(t) = F'(t) = \begin{cases} \frac{1}{2\theta\sqrt{t}} & \text{if } 0 < t < \theta^2, \text{ and} \\ 0 & \text{otherwise.} \end{cases}$$

(c) We can see that

$$nV = \sum_{i=1}^{n} U_i^2,$$

so by the Central Limit Theorem, we know that for large n,

$$\frac{nV - n\mathbb{E}[U_1^2]}{\sqrt{n\,\mathrm{Var}(U_1^2)}} \xrightarrow{\text{in distribution}} \mathcal{N}(0,1).$$

Hence, multiplying and adding, we can see that

$$V \xrightarrow{\text{in distribution}} \mathcal{N}\left(\mathbb{E}[U_1^2], \frac{1}{n}\mathrm{Var}\left(U_1^2\right)\right).$$

Now, it remains to calculate both the expectation and variance of U_1^2. We have that

$$\mathbb{E}[U_1^2] = \mathrm{Var}(U_1) + \mathbb{E}[U_1]^2 = \mathrm{Var}(U_1) = \frac{\theta^2}{3},$$

and we have that

$$\mathrm{Var}\left(U_1^2\right) = \mathbb{E}[U_1^4] - \mathbb{E}[U_1^2]^2 = \int_{-\theta}^{\theta} \frac{t^4}{2\theta}\,dt - \frac{\theta^4}{9} = \frac{\theta^4}{5} - \frac{\theta^4}{9} = \frac{4\theta^4}{45},$$

so $V \sim \mathcal{N}\left(\frac{\theta^2}{3}, \frac{4\theta^4}{45n}\right)$.

Alternatively, we can do these calculations using the distribution for U_1^2 derived in a previous part. We have that

$$\mathbb{E}[U_1^2] = \int_0^{\theta^2} t \cdot \frac{1}{2\theta\sqrt{t}}\, dt = \int_0^{\theta^2} \frac{\sqrt{t}}{2\theta}\, dt = \frac{\theta^2}{3},$$

and we have that

$$\mathrm{Var}(U_1^2) = \int_0^{\theta^2} t^2 \cdot \frac{1}{2\theta\sqrt{t}}\, dt - \frac{\theta^4}{9} = \int_0^{\theta^2} \frac{t^{\frac{3}{2}}}{2\theta}\, dt - \frac{\theta^4}{9} = \frac{4\theta^4}{45},$$

so again, $V \sim \mathcal{N}\left(\frac{\theta^2}{3}, \frac{4\theta^4}{45n}\right)$.

(d) We can use $3V$ as our unbiased estimator, as $\mathbb{E}[3V] = \theta^2$ and $\mathrm{Var}(3V) = \frac{4\theta^4}{5n} \to 0$ as $n \to \infty$.

(e) (i) We will use Chebyshev's inequality to bound the probability of deviation from the mean. Firstly, we can compute that

$$\mathrm{Var}(3V) = 9\,\mathrm{Var}(V) = \frac{9\sigma^2}{n}.$$

Moving forward, we have that

$$\mathbb{P}[|3V - \theta^2| \geq c] \leq \frac{\mathrm{Var}(3V)}{c^2} = \frac{9\sigma^2}{nc^2},$$

so in order to guarantee that this probability is less than δ, we need to set

$$\frac{9\sigma^2}{nc^2} \leq \delta \implies c \geq \frac{3\sigma}{\sqrt{\delta n}},$$

so our confidence interval is thus $[3V - \frac{3\sigma}{\sqrt{\delta n}}, 3V + \frac{3\sigma}{\sqrt{\delta n}}]$.

(ii) With the assumption that n is large, we can claim via the CLT that $3V \sim \mathcal{N}(\theta^2, \frac{9\sigma^2}{n})$, so in particular, $\frac{\sqrt{n}(3V-\theta^2)}{3\sigma}$ is a standard normal. Thus, we have that

$$\mathbb{P}[|3V - \theta^2| > c] = \mathbb{P}\left[\frac{\sqrt{n}|3V - \theta^2|}{3\sigma} > \frac{c\sqrt{n}}{3\sigma}\right] = 1 - \Phi\left(\frac{c\sqrt{n}}{3\sigma}\right) + \Phi\left(-\frac{c\sqrt{n}}{3\sigma}\right).$$

We can further simplify the right hand side of this to

$$\mathbb{P}[|3V - \theta^2| > c] = 2\Phi\left(-\frac{c\sqrt{n}}{3\sigma}\right),$$

hence to get a significance level of δ, we can set

$$2\Phi\left(-\frac{c\sqrt{n}}{3\sigma}\right) = \delta \implies c = -\frac{3\sigma}{\sqrt{n}}\Phi^{-1}\left(\frac{\delta}{2}\right).$$

Hence, our confidence interval is $\left[3V + \frac{3\sigma}{\sqrt{n}}\Phi^{-1}\left(\frac{\delta}{2}\right), 3V - \frac{3\sigma}{\sqrt{n}}\Phi^{-1}\left(\frac{\delta}{2}\right)\right]$.