Due: Sunday, August 9, 2020 at 10:00 PM
Grace period until Sunday, August 9, 2020 at 11:59 PM

1 Tightness of Inequalities

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

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 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). 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 α such that using this method (dividing by two to bound one tail) is not correct, that is, $\mathbb{P}(X \geq \mathbb{E}[X]+\alpha) > \frac{\text{Var}(X)}{2\alpha^2}$ 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 $\text{Var}(X) = 2$. 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 Erasures, Bounds, and Probabilities

Alice is sending 1000 bits to Bob. The probability that a bit gets erased is p, and the erasure of each bit is independent of the others.

Alice is using a scheme that can tolerate up to one-fifth of the bits being erased. That is, as long as Bob receives at least 801 of the 1000 bits correctly, he can decode Alice's message.

In other words, Bob becomes unable to decode Alice's message only if 200 or more bits are erased. We call this a "communication breakdown", and we want the probability of a communication breakdown to be at most 10^{-6}.

1. Use Markov's inequality to upper bound p such that the probability of a communications breakdown is at most 10^{-6}.
2. Use Chebyshev's inequality to upper bound p such that the probability of a communications breakdown is at most 10^{-6}.
3. As the CLT would suggest, approximate the fraction of erasures by a Gaussian random variable (with suitable mean and variance). Use this to find an approximate bound for p such that the probability of a communications breakdown is at most 10^{-6}.

4 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,

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

Assume that you've been given 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 worst-case error with which you can report your results?

5 Oski's Markov Chain

When Oski Bear is studying for CS70, he splits up his time between reading notes and working on practice problems. To do this, every so often he will make a decision about what kind of work to do next.

When Oski is already reading the notes, with probability a he will decide to switch gears and work on a practice problem, and otherwise, he will decide to keep reading more notes. Conversely, when Oski is already working on a practice problem, with probability b he will think of a topic he needs to review, and will decide to switch back over to the notes; otherwise, he will keep working on practice problems.

Assume that (unlike real life, we hope!) Oski never runs out of work to do.

(a) Draw a 2-state Markov chain to model this situation.

(b) In the remainder of this problem, we will learn to work with the definitions of some important terms relating to Markov Chains. These definitions are as follows:

  (a) (Irreducibility) A Markov chain is irreducible if, starting from any state i, the chain can transition to any other state j, possibly in multiple steps.

  (b) (Periodicity) $d(i) := \gcd\{n > 0 \mid P^n(i,i) = \mathbb{P}[X_n = i \mid X_0 = i] > 0\}$, $i \in \mathscr{X}$. If $d(i) = 1\ \forall i \in \mathscr{X}$, then the Markov chain is aperiodic; otherwise it is periodic.

  (c) (Matrix Representation) Define the transition probability matrix P by filling entry (i,j) with probability $P(i,j)$.

(d) (Invariance) A distribution π is invariant for the transition probability matrix P if it satisfies the following balance equations: $\pi = \pi P$.

For what values of a and b is the Markov chain irreducible?

(c) For $a = 1$, $b = 1$, prove that the Markov chain is periodic.

(d) For $0 < a < 1$, $0 < b < 1$, prove that the Markov chain is aperiodic.

(e) Construct a transition probability matrix using the Markov chain.

(f) Write down the balance equations for the Markov chain and solve them. Assume that the Markov chain is irreducible.

6 Markov Chains: Prove/Disprove

Prove or disprove the following statements, using the definitions from the previous question.

(a) There exists an irreducible, finite Markov chain for which there exist initial distributions that converge to different distributions.

(b) There exists an irreducible, aperiodic, finite Markov chain for which $\mathbb{P}(X_{n+1} = j \mid X_n = i) = 1$ or 0 for all i, j.

(c) There exists an irreducible, non-aperiodic Markov chain for which $\mathbb{P}(X_{n+1} = j \mid X_n = i) \neq 1$ for all i, j.

(d) For an irreducible, non-aperiodic Markov chain, any initial distribution not equal to the invariant distribution does not converge to any distribution.

7 Balls, meet Bins

Alice and Bob are tasked with throwing balls into bins (to set up a probability problem for later). They decide to make a game out of it: Alice and Bob will each take a ball, and once per minute, they will both simultaneously (and independently) attempt to throw their balls into a bin.

Once Alice or Bob successfully lands a ball in a bin, that person stops while the other person continues to try until they also land a throw. When this happens, the game is over.

Suppose that on every try, the probability of successfully landing the throw in a bin is p. What is the expected number of minutes until the game is over? Solve this using a Markov chain with three states. Then, state how your solution can be interpreted in terms of two geometric random variables.

8 Homework Process and Study Group

Citing sources and collaborators are an important part of life, including being a student! We also want to understand what resources you find helpful and how much time homework is taking, so we can change things in the future if possible.

1. **What sources (if any) did you use as you worked through the homework?**
2. **If you worked with someone on this homework, who did you work with?** List names and student ID's. (In case of homework party, you can also just describe the group.)
3. **How did you work on this homework?** (For example, *I first worked by myself for 2 hours, but got stuck on problem 3, so I went to office hours. Then I went to homework party for a few hours, where I finished the homework.*)
4. **Roughly how many total hours did you work on this homework?**