12: Tail Bounds

PSTAT 120A: Summer 2022

Ethan P. Marzban

July 20, 2022

University of California, Santa Barbara

Where We've Been

- Axioms of Probability, Probability Spaces, Counting
- Conditional Probabilities, independence, etc.
- Basics of Random Variables (classification, p.m.f., c.m.f., moments)
- Discrete Distributions
- Continuous Distributions
- Transformations of Random Variables
- Double Integrals
- Random Vectors and the basics of multivariate probability
- Independence of random variables, and covariance/correlation
- Sums of Random Variables; Indicators
- Moment Generating Functions

Leadup

- Up until now, we've been relying heavily on our knowledge about distributions; specifically, "famous" or "known" ones (e.g. Binomial, Poisson, Normal, Gamma, etc.)
- We now shift our attention to a slightly different situation: what happens when we have *very little* information about a random variable? As in, what if we don't even know what distribution it follows?
- In general, there is not a whole lot we can do in such situations.
- So, let's suppose I have a random variable X and that I do not know its distribution, but I do know its mean. Can I say anything about probabilities involving X?
- Surprisingly... yes!

Tail Bounds

Markov's Inequality

Theorem: Markov's Inequality

For a nonnegative function u of a random variable X and a constant $c > 0$,

$$\mathbb{P}(u(X) \geq c) \leq \frac{\mathbb{E}[u(X)]}{c}$$

- When X is nonnegative, "Markov's Inequality" is often reported as the above theorem with u taken to be the identity function.

Proof.

On the Chalkboard (time permitting) $\square$

Example

- As an example, suppose X is a nonnegative random variable with mean 2. Then

$$\mathbb{P}(X \geq 3) \leq \frac{2}{3}$$

- Pretty neat, huh?
- Well, suppose instead of considering $\mathbb{P}(X \geq 3)$, I had considered $\mathbb{P}(X \geq 1)$. Then, by Markov's Inequality,

$$\mathbb{P}(X \geq 1) \leq \frac{2}{1} = 2$$

  Wait... couldn't we have done better without even applying Markov's?
- So, this last example illustrates an important point: though Markov's inequality is very useful, it is *very conservative*. Sometimes, in fact, it is so conservative that the bound it provides, though correct, is entirely useless.

Example

- By the way, can we use Markov's Inequality to bound quantities like $\mathbb{P}(X < c)$? Certainly!
- Note that

$$\mathbb{P}(X < c) = 1 - \mathbb{P}(X \geq c) \implies \mathbb{P}(X \geq c) = 1 - \mathbb{P}(X < c)$$

- By Markov's Inequality,

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

meaning

$$1 - \mathbb{P}(X < c) \leq \frac{\mathbb{E}[X]}{c}$$

or, isolating $\mathbb{P}(X < c)$, we find

$$\left[1 - \frac{\mathbb{E}[X]}{c}\right] \geq \mathbb{P}(X < c)$$

- So, in this case, note that Markov's Inequality provides a *lower* bound.

Chebyshev's Inequality

- Alright, suppose someone has decided to take pity on us and in addition to telling us the mean of X, they have also told us the variance.
- We can now modify our bounds on probabilities, by way of:

Theorem: Chebyshev's Inequality

Suppose X is a random variable with finite mean μ and finite variance σ^2; then, for any $c > 0$,

$$\mathbb{P}(|X - \mu| \geq c) \leq \frac{\sigma^2}{c^2}$$

Chebyshev's Inequality

Theorem: Chebyshev's Inequality

Suppose X is a random variable with finite mean μ and finite variance σ^2; then, for any $c > 0$,

$$\mathbb{P}(|X - \mu| \geq c) \leq \frac{\sigma^2}{c^2}$$

Proof.

- Note $|X - \mu|$ is a nonnegative random variable; additionally, since $c > 0$ we see that

$$\{|X - \mu| \geq c\} \iff \{(X - \mu)^2 \geq c^2\}$$

- We can apply Markov's Inequality:

$$\mathbb{P}(|X - \mu| \geq c) = \mathbb{P}\left[(X - \mu)^2 \geq c^2\right] \leq \frac{\mathbb{E}\left[(X - \mu)^2\right]}{c^2}$$

- Finally, recall that $\mathbb{E}[(X - \mu)^2] =: \mathrm{Var}(X) = \sigma^2$.

$\square$

Chebyshev's Inequality

- By the way, note the following:

$$\{X \geq \mu + c\} \subseteq \{|X - \mu| \geq c\}$$
$$\{X \leq \mu - c\} \subseteq \{|X - \mu| \geq c\}$$

  (sketch a number line if you're not convinced!)

- Therefore, Chebyshev's Inequality also gives:

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

- Also, we can manipulate Chebyshev's Inequality to give lower bounds (I leave this as an exercise to the reader.)

Example

Suppose X is a random variable with mean 2 and variance 1. Then

$$\mathbb{P}(|X - 2| \geq 4) \leq \frac{1}{4^2} = \frac{1}{16}$$

Comparing the Bounds

- Okay, so it seems we have two ways of bounding probabilities: Markov's Inequality and Chebyshev's Inequality. Which is better?
- Well, firstly, what do we mean by "better?" (Sometimes we also refer to a "good" bound as a "tight" bound.)
- More concretely, suppose we are trying to bound $\mathbb{P}(X \geq c)$; suppose Markov's inequality tells us $\mathbb{P}(X \geq c) \leq a$ for some a, and Chebyshev's tells us $\mathbb{P}(X \geq c) \leq b$ for some b.
- Clearly, the upper bound that is *smaller* provides more information, as it eliminates a greater number of values! Remember, at the end of the day, we'd like to compute $\mathbb{P}(X \geq c)$ exactly. So, we'd like to eliminate as much uncertainty as possible.
- So, in practice, we often use **both** Markov's and Chebyshev's inequalities, and then report the bound that is tightest.

Comparing the Bounds

- I would also like to say- the bounds reported by Markov's and Chebyshev's inequalities are just that- bounds. Both tend to be quite conservative!
- As an illustration, suppose $X \sim \text{Geom}(1/2)$. Further suppose we wished to bound $\mathbb{P}(X \geq 6)$.
- Markov's inequality tells us

$$\mathbb{P}(X \geq 6) \leq \frac{\mathbb{E}[X]}{6} = \frac{2}{6} = \frac{1}{3} = 0.\overline{3}$$

- Chebyshev's tells us

$$\mathbb{P}(X \geq 6) = \mathbb{P}(X \geq 2 + 4) \leq \frac{\text{Var}(X)}{4^2} = \frac{4}{16} = \frac{1}{4} = 0.25$$

- So, in this case, Chebyshev's gives us a tigher bound.
- **BUT**, for this particular problem, we can actually compute $\mathbb{P}(X \geq 6)$ *exactly*:

$$\mathbb{P}(X \geq 6) = \sum_{k=6}^{\infty} \left(\frac{1}{2}\right)^k = \left(\frac{1}{2}\right)^5 \approx 0.03$$

- Notice how far off *both* Markov's and Chebyshev's were!

Conclusion

- So, in conclusion, Markov's and Chebyhev's Inequalities are very useful in that the require so few assumptions.
- However, as the saying goes, "there is no free lunch-" because they operate under so few assumptions, they tend to be wildly conservative to compensate, sometimes being *so* conservative as to report bounds that are totally useless (i.e. upper bounds greater than 1, or lower bounds less than 0). Additionally, even the best bound between the two Inequalities will often be far from the true value.