Math 20 – Inequalities of Markov and Chebyshev

Often, given a random variable X whose distribution is unknown but whose expected value μ *is* known, we may want to ask how likely it is for X to be 'far' from μ, or how likely it is for this random variable to be 'very large.' This would give us some idea of the spread of the distribution, though perhaps not a complete picture.

Proposition 1 (Markov's Inequality)**.** *Let X be a random variable that takes only nonnegative values. Then for any positive real number a,*

$$P(X \geq a) \leq \frac{E(X)}{a}$$

provided $E(X)$ exists.

For example, Markov's inequality tells us that as long as X doesn't take negative values, the probability that X is twice as large as its expected value is at most $\frac{1}{2}$, which we can see by setting $a = 2E(X)$. More generally, the probability that a random variable is at least k times its expected value is at most $\frac{1}{k}$. Notice that the only things we assumed about this random variable are that it can't be negative and has finite mean; we don't need to know anything about its variance or it's probability distribution, in general.

Proof. We'll prove this for discrete RVs, but the proof for continuous RVs is essentially the same, replacing sums with integrals.

By definition, $E(X) = \sum_x xP(X = x)$. We'll split this sum into two pieces, depending on whether or not $x \geq a$.

$$\begin{aligned} E(X) &= \sum_{x \geq a} xP(X = x) + \sum_{x < a} xP(X = x) \\ &\geq \sum_{x \geq a} aP(X = x) + 0 \qquad \text{(since in the first sum we assume } x \geq a) \\ &= a \sum_{x \geq a} P(X = x) \\ &= aP(X \geq a) \end{aligned}$$

□

Example 2. Suppose that the average grade on the upcoming Math 20 exam is 70%. Give an upper bound on the proportion of students who score at least 90%.

$$P(X \geq 90) \leq \frac{E(X)}{90} = \frac{7}{9}$$

so at most 77.8% of students can possible score this high. But in order to achieve this average, we would need $\frac{7}{9}$ of the class to score a 90 and the remaining $\frac{2}{9}$ to score a 0...

Example 3. A coin is weighted so that its probability of landing on heads is 20%. Suppose the coin is flipped 20 times. Find a bound for the probability it lands on heads at least 16 times.

We actually do *know* this distribution; it's the the binomial distribution with $n = 20$ and $p = \frac{1}{5}$. It's expected value is 4. Markov's inequality tells us that

$$P(X \geq 16) \leq \frac{E(X)}{16} = \frac{1}{4}.$$

Let's compare this to the *actual* probability that this happens:

$$P(X \geq 16) = \sum_{k=16}^{20} \binom{20}{k} 0.2^k \cdot 0.8^{20-k} \approx 1.38 \cdot 10^{-8}.$$

So it seems like this is not a very good estimate. We'll see later that this distribution (at least, for large n) is close to normal, and Markov's inequality doesn't get close to the true value for such "compact" distributions.

Example 4 (Markov's Inequality is Tight)**.** Consider a random variable X that takes the value 0 with probability $\frac{24}{25}$ and the value 1 with probability $\frac{1}{25}$. Then

$$E(X) = \frac{1}{25} \cdot 5 = \frac{1}{5}.$$

Let's use Markov's inequality to find a bound on the probability that X is at least 5:

$$P(X \geq 5) \leq \frac{E(X)}{5} = \frac{1/5}{5} = \frac{1}{25}.$$

But this is exactly the probability that $X = 5$! We've found a probability distribution for X and a positive real number k such that the bound given by Markov's inequality is exact; we say that Markov's inequality is *tight* in the sense that in general, no better bound (using only $E(X)$) is possible.

If $P(X = 0) = 1 - \frac{1}{k^2}$ and $P(X = k) = \frac{1}{k^2}$ (as the example above in which $k = 5$), Markov's inequality gives the best possible bound, but as we saw earlier, there are cases for which Markov's inequality provides a terrible bound. Fortunately, we can say a bit more about a probability distribution if we know it's variance σ^2 as well as it's expected value.

Proposition 5 (Chebyshev's Inequality)**.** *Let X be any random variable with finite expected value and variance. Then for every positive real number a,*

$$P(|X - E(X)| \geq a) \leq \frac{\mathrm{Var}(X)}{a^2}.$$

There is a direct proof of this inequality in Grinstead and Snell (p. 305) but we can also prove it using Markov's inequality!

Proof. Let $Y = (X - E(X))^2$. Then Y is a non-negative valued random variable with expected value $E(Y) = \text{Var}(X)$. By Markov's inequality,

$$P(Y \geq a^2) \leq \frac{E(Y)}{a} = \frac{\text{Var}(X)}{a^2}.$$

But notice that the event $Y \geq a^2$ is the same as $|X - E(X)| \geq a$, so we conclude that

$$P(|X - E(X)| \geq a) \leq \frac{\text{Var}(X)}{a^2}.$$

□

Chebyshev's inequality gives a bound on the probability that X is far from it's expected value. If we set $a = k\sigma$, where σ is the standard deviation, then the inequality takes the form

$$P(|X - \mu)| \geq k\sigma) \leq \frac{\text{Var}(X)}{k^2\sigma^2} = \frac{1}{k^2}.$$

Example 6. Suppose a fair coin is flipped 100 times. Find a bound on the probability that the number of times the coin lands on heads is at least 60 or at most 40.

Let X be the number of times the coin lands on heads. We know X has a binomial distribution with expected value 50 and variance $100 \cdot 0.5 \cdot 0.5 = 25$. By Chebyshev, we have

$$P(X < 40 \cup X > 60) = P(|X - \mu| \geq 10) \leq \frac{25}{10^2} = \frac{1}{4}.$$

The actual probability of this happening is close to 5%.

Example 7. While in principle Chebyshev's inequality asks about distance from the mean in either direction, it can still be used to give a bound on how often a random variable can take large values, and will usually give much better bounds than Markov's inequality. Let's revisit Example 3 in which we toss a weighted coin with probability of landing heads 20%. Doing this 20 times, Markov's inequality gives a bound of $\frac{1}{4}$ on the probability that at least 16 flips result in heads. Using Chebyshev's inequality,

$$\begin{aligned}
P(X \geq 16) &= P(0 \leq X \leq 16) \\
&= P(-8 \leq X \leq 16) \quad \text{(since } X \text{ can't be negative)} \\
&= P(|X - 4| \geq 12) \\
&\leq \frac{\text{Var}(X)}{12^2} \\
&= \frac{20 \cdot 0.2 \cdot 0.8}{144} \\
&= \frac{3.2}{144} = \frac{1}{45}
\end{aligned}$$

This is a much better bound than given by Markov's inequality, but still far from the actual probability.

Exercise 8. A biased coin lands heads with probability $\frac{1}{10}$. This coin is flipped 200 times. Use Markov's inequality to give an upper bound on the probability that the coin lands heads at least 120 times. Improve this bound using Chebyshev's inequality.

Exercise 9. The average height of a raccoon is 10 inches.

1. Given an upper bound on the probability that a certain raccoon is at least 15 inches tall.
2. The standard deviation this height distribution is 2 inches. Find a lower bound on the probability that a certain raccoon is between 5 and 15 inches tall.
3. Now assume this distribution is normal. Use a normal CDF table to repeat the calculation from part (b). How close was your bound to the true probability?

Exercise 10. Like we did in Example 4 for Markov's inequality, prove that Chebyshev's inequality is tight: find a probability distribution for X and a value a such that $P(|X - E(X)| \geq a) = \frac{\mathrm{Var}(X)}{a^2}$. (Hint: This random variable will take only three values.)