Tchebysheff's Inequality

March 18, 2020

1 Bounding Probabilities

Up to this point one of the main questions we have been interested in is: What is the probability a random variable, X, is between a and b for two (possibly infinite) numbers a and b.

So far we have always been given the pdf of X, so to answer this question we compute the integral:

$$\mathbb{P}(a \leq X \leq b) = \int_a^b f_X(x)dx$$

We will see that there are many times when we don't know the pdf the random variable, and other times where even if it is known, it is difficult to do computations with. In these cases we might still have partial information about X, which we can use to give upper or lower bounds on the probability it takes a value in some set.

In many cases the expectation and variance can give these bounds, and furthermore are easier to compute than the entire pdf. One of the bounds is known as Tchebysheff's Inequality, which we now state:

THEOREM 1.1 (Tchebysheff's Inequality). *Let Y be a random variable with mean μ and variance σ^2. Then for any $k > 0$,*

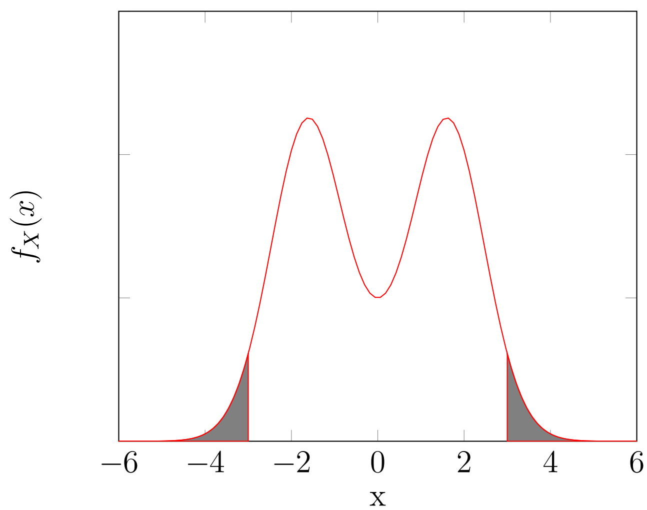
$$\mathbb{P}(|Y - \mu| < k\sigma) \geq 1 - 1/k^2 \text{ or } \mathbb{P}(|Y - \mu| \geq k\sigma) \leq 1/k^2$$

Some remarks on the theorem:

- The two statements are equivalent, because the events we are computing the probabilities of are complements.
- This is a useful Theorem because it is very general, all that we need is for σ^2 to be finite.
- The book has chosen a creative spelling of Tchebysheff, Chebyshev is more common.

Below is a picture to have in mind. I have plotted the pdf of random variable, X, with mean 0 and standard deviation 1.

If we want to know the probability X is more than 3 standard deviations away from its mean, we would integrate the shaded region fo the picture. If for some reason we didn't know complete information on the pdf or the integral is too hard, Tchebysheff's inequality gives up an upper bound. We know the area of the shaded region under the curve is less that $1/3^2 = 1/9$.



EXAMPLE 1.2. *Let Y be a gamma random variable with parameters $\alpha = 10$ and $\beta = 3$. Use Tchebysheff's inequality to give an upper bound on the probability $Y > 50$.*

Solution: From the gamma distribution we know $\mathbb{E}[Y] = 10 * 3 = 30$ and $\mathbb{V}[Y] = 10 * 3^2$, or $\sigma = \sqrt{10} * 3$.

Now we consider the quantity we want and manipulate it into a form where we can apply Tchebysheff's inequality.

$$\mathbb{P}(Y > 50) = \mathbb{P}(Y - 30 > 50 - 30) \leq \mathbb{P}(|Y - 30| > 20)$$

In the last step we used that the event $Y - 30 > 20$ is a subset of the event $|Y - 30| > 20$, this is not terribly efficient, but suffices for our purposes. Now we write 20 as a factor of $\sigma = 3\sqrt{10}$, namely $20 = 3\sqrt{10} * \frac{20}{3\sqrt{10}}$, so we can now apply Tchebysheff's inequality with $k = \frac{20}{3\sqrt{10}}$ and conclude

$$\mathbb{P}(|Y - 30| > 20) \leq \frac{1}{\left(\frac{20}{3\sqrt{10}}\right)^2} = \frac{9 * 10}{20^2} = .225.$$

We'll see more on this inequality in later chapters. The main point for us now is that Tchebysheff's inequality tells us that if the variance is small then a random variable is close to its expectation, and if the variance is large, then it is not unlikely for a random variable to be far away from its expectation.