

4.4: The Law of Large Numbers

If we want our sample statistics to be much closer to the population parameters, what can we do about it?

The answer is to collect more data. Larger samples are a much better approximation to the true population distribution than smaller samples. I feel a bit silly saying this, but the thing I want you to take away from this is that large samples generally give you better information. It does feel a bit obvious that more data will give you better answers.

The question is, why is this so? Not surprisingly, this intuition that we all share turns out to be correct, and statisticians refer to it as the **law of large numbers**. The law of large numbers is a mathematical law that applies to many different sample statistics, but the simplest way to think about it is as a law about averages. The sample mean is the most obvious example of a statistic that relies on averaging (because that's what the mean is... an average), so let's look at that. When applied to the sample mean, what the law of large numbers states is that as the sample gets larger, the sample mean tends to get closer to the true population mean. Or, to say it a little bit more precisely, as the sample size “approaches” infinity (written as $N \rightarrow \infty$) the sample mean approaches the population mean ($\bar{X} \rightarrow \mu$).

I don't intend to subject you to a proof that the law of large numbers is true, but it's one of the most important tools for statistical theory. The law of large numbers is the thing we can use to justify our belief that collecting more and more data will eventually lead us to the truth. For any particular data set, the sample statistics that we calculate from it will be wrong, but the law of large numbers tells us that if we keep collecting more data those sample statistics will tend to get closer and closer to the true population parameters.

It's okay if this isn't obvious to you. I'm just mentioning it because it will start to feel obvious, and explains lots about why statistical formulas are the way they are. I mean, we've already run across that in how the formula for the standard deviation of a

sample ($\sqrt{\frac{\sum (X - \bar{X})^2}{N - 1}}$) is different from the formula for the standard deviation of the population ($\sqrt{\frac{\sum (X - \bar{X})^2}{N}}$).

Contributors and Attributions

- [Danielle Navarro](#) ([University of New South Wales](#))
- [Dr. MO](#) ([Taft College](#))

This page titled [4.4: The Law of Large Numbers](#) is shared under a [CC BY-SA 4.0](#) license and was authored, remixed, and/or curated by [Danielle Navarro](#).