26 2026-04-08 | Week 12 | Lecture 26

26.1 Point estimation

Chapter 6.1

Recall a **population** is a large body of data that is the target of our interest, and a **statistic** is any quantity whose value can be computed from a sample.

Often, we have some quantity of interest, called a ***target parameter***, and we want a single "best guess" of some quantity of interest. When this is the case, the we call a statistic an estimator:

Definition 121 (Estimator)**.** An ***estimator*** is a statistic, that is a function $T(X_1, \ldots, X_n)$ of a sample, that is used to approximate a target parameter. An ***estimate*** is the realized value of an estimator (e.g., a number) that is obtained when the sample is actually taken.

In words, an estimator is a rule, often expressed as a formula, that tells us how to calculate the value of an estimate based on the measurements contained in a sample.

Example 122 (Broccoli Poll)**.** Suppose the fraction of people in Hawaii who like broccoli is p. This is an unknown **target parameter**. We don't know what p is, but we would like to estimate it.In principle, we could ask everyone in the state, but that would be impractical. Instead, we take a random sample: we choose randomly n individuals and ask each of them whether they like broccoli or not. Let

$$\widehat{p} := \text{ fraction of our sample who like broccoli}$$

Suppose we sample $n = 100$ individuals and 20 of them like broccoli. Then

$$\widehat{p} = \frac{20}{100} = .2$$

The **sample** in this example consists of n random variables $X_1, \ldots, X_n$, where

$$X_i = \begin{cases} 1 & : \text{ i-th sampled person likes broccoli} \\ 0 & : \text{ they don't} \end{cases}$$

Our **point estimate** is $\widehat{p} = 0.2$ and the **point estimator** is the function $\widehat{p} = \frac{X_1 + \ldots + X_n}{n}$, which is a **sample mean**.

Taking a step back, we can think of our estimate as

$$\widehat{p} = p + \text{error of estimation}$$

It is perhaps clear that if we increase the sample size n, then the estimation error should decrease, so that

$$\widehat{p} \to p \text{ as } n \to \infty.$$

End of Example 122. □

Definition 123 (Bias and standard error)**.** Let $\widehat{\theta}$ be a point estimator of a target parameter θ. The ***bias*** of $\widehat{\theta}$ is the number $\text{Bias}(\widehat{\theta}) = \mathbb{E}[\widehat{\theta}] - \theta$. If $\text{Bias}(\widehat{\theta}) = 0$, we say that $\widehat{\theta}$ is ***unbiased***, otherwise we say it is ***biased***.

The ***standard error*** of $\widehat{\theta}$, denoted $\sigma_{\widehat{\theta}}$, is its standard deviation:

$$\sigma_{\widehat{\theta}} = \sqrt{\text{Var}(\widehat{\theta})}$$

Example 124 (Broccoli poll continued)**.** In the broccoli example, the estimator we used was the ***sample mean:***

$$\widehat{p} = \frac{X_1 + \ldots + X_n}{n}.$$

Claim 1: $\widehat{p}$ is unbiased.

Proof of Claim 1.

$$\begin{aligned}\mathbb{E}\left[\widehat{p}\right] &= \mathbb{E}\left[\frac{X_1 + \ldots + X_n}{n}\right] \\ &= \frac{1}{n}\left(\mathbb{E}\left[X_1\right] + \ldots + \mathbb{E}\left[X_n\right]\right) \\ &= \frac{1}{n}(np) \\ &= p.\end{aligned}$$

□ Claim

Claim 2: $\sigma_{\widehat{p}} = \sqrt{\frac{p(1-p)}{n}}$.

Proof of Claim 2. First observe that for every $i \in \{1, \ldots, n\}$, we have

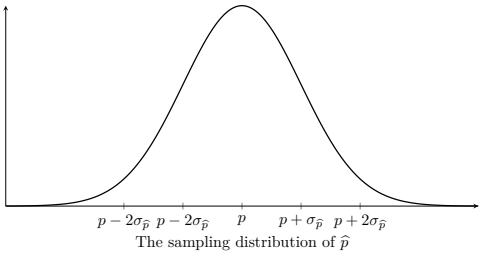
$$\mathrm{Var}(X_i) = \mathbb{E}\left[X_i^2\right] - \left(\mathbb{E}\left[X_i\right]\right)^2 = p - p^2 = p(1-p).$$

Using Theorem 66,

$$\begin{aligned}\mathrm{Var}(\widehat{p}) &= \mathrm{Var}\left(\frac{X_1 + \ldots + X_n}{n}\right) \\ &= \frac{1}{n^2}\left(\mathrm{Var}(X_1 + \ldots + X_n)\right) \\ &= \frac{1}{n^2}\left(\mathrm{Var}(X_1) + \ldots + \mathrm{Var}(X_n)\right) && \text{by independence} \\ &= \frac{1}{n^2}\left(np(1-p)\right) \\ &= \frac{p(1-p)}{n}\end{aligned}$$

Taking square roots implies the statement of the claim. □ Claim

Finally, since $\widehat{p}$ is a sample mean, the central limit theorem implies that its distribution is approximately normal:



The sampling distribution of $\widehat{p}$

As n grows, the value of $\sigma_{\widehat{p}} = \sqrt{\frac{p(1-p)}{n}}$ gets smaller, so the above curve (which is centered around the true parameter value p) gets narrower.

End of Example 124. □

Example 125 (Estimating the sides of a dice)**.** Suppose I roll a dice behind a screen, shouting out the number rolled each time. Your job is to estimate the number of sides of the dice, which we will call θ.

Given n rolls, the sample consists of the n random variables

$$X_1, \ldots, X_n$$

where X_i is the i^{th} dice roll. We'll give two examples of estimators:

- Our first example is called the *running maximum*, and it is defined as

$$\widehat{\theta} = \max(X_1, \ldots, X_n).$$

For example, if I roll the dice four times and obtain $3, 11, 7$, and 4, then your estimate would be that the dice has 11 sides.

This sounds silly, but with enough data this estimator will give you the correct answer with high probability. This is because, if I roll the dice enough times, then I will eventually roll the highest number; hence,

$$\lim_{n\to\infty} \widehat{\theta}(X_1, \ldots, X_n) = \theta.$$

One drawback is that this estimator is biased because $\mathbb{E}[\widehat{\theta}] < \theta$. This is because the estimator will "usually" be less than the true number of sides, and will never be more—so on average, $\widehat{\theta}$ will be smaller than θ.

- Our second example is what is called the *method of moments estimator*. In this setting, it's defined as

$$\widehat{\theta} = 2\left(\frac{X_1 + \ldots + X_n}{n}\right) - 1.$$

Let's demystify this formula. Suppose X a random variable representing one roll of the dice. We know that since the dice has θ sides, X has expected value

$$\mathbb{E}[X] = \frac{1}{\theta}(1 + 2 + \ldots + \theta) = \frac{\theta + 1}{2}.$$

Rearranging, we get

$$\theta = 2\mathbb{E}[X] - 1. \tag{17}$$

We also know that by the law of large numbers, $\mathbb{E}[X] \approx \frac{X_1+\ldots+X_n}{n}$. Combining this with Eq. (17) implies

$$\theta \approx 2\left(\frac{X_1 + \ldots + X_n}{n}\right) - 1.$$

The right hand side is what we've chosen to be our $\widehat{\theta}$, and hopefuly this calculation makes it seem like a reasonable choice.

This estimator is unbiased. To see why, observe that

$$\begin{aligned}
\mathbb{E}[\widehat{\theta}] &= \mathbb{E}\left[2\left(\frac{X_1 + \ldots + X_n}{n}\right) - 1\right] \\
&= 2\mathbb{E}\left[\frac{X_1 + \ldots + X_n}{n}\right] - 1 \\
&= 2\mathbb{E}[X] - 1 \\
&= \theta && \text{by Eq. (17).}
\end{aligned}$$

In our example, with dice rolls $3, 11, 7$, and 4, our estimator $\widehat{\theta}$ returns the estimate

$$2\left(\frac{3 + 11 + 7 + 4}{4}\right) - 1 = 2\left(\frac{25}{4}\right) - 1 = 11.5.$$

So your estimate would be that the dice has 11.5 sides.

End of Example 125. □

Given two unbiased estimators, then *ceteris paribus*, we would typically prefer the estimator which has lower variance.