28 2025-04-02 | Week 11 | Lecture 27

28.1 Property of point estimators: Bias

Section 6.1

Let $\theta \in \mathbb{R}$ be an unknown parameter. Suppose that $\widehat{\theta}$ is a point estimator of θ. We say that $\widehat{\theta}$ is *unbiased* if

$$\mathbb{E}\left[\widehat{\theta}\right] = \theta.$$

The *bias* of $\widehat{\theta}$ is the number $\mathbb{E}\left[\widehat{\theta}\right] - \theta$.

For clarity, recall that as an estimator, $\widehat{\theta}$ is a function $\widehat{\theta} = f(X_1, \ldots, X_n)$. And so when we write

$$\mathbb{E}\left[\widehat{\theta}\right]$$

we really mean

$$\mathbb{E}\left[f(X_1, \ldots, X_n)\right]$$

In the broccoli example, the estimator we used was the *sample mean:*

$$\widehat{\theta}(X_1, \ldots, X_n) = \frac{X_1 + \ldots + X_n}{n}.$$

This estimator is unbiased because

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

Example 85 (A biased estimator)**.** 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. For n rolls, the data consists of n random variables

$$X_1, \ldots, X_n$$

where X_i is the i^{th} dice roll. One natural estimator is

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

This estimator is called the *running maximum*. For example, if we roll the hidden dice four times and get the following numbers for $X_1, \ldots, X_4$:

$$3, 11, 7, \text{ and } 4,$$

then our estimate would be that the dice has 11 sides.

If I roll the dice enough times, I will eventually roll the highest number; it follows from this that

$$\lim_{n \to \infty} \widehat{\theta} = \theta.$$

But we expect it to be biased (since it will "usually" be less than the true number of sides, and will never be more).

To make this precise, we can use the law of total expectation. Let $p = \mathbb{P}\left[\widehat{\theta} = \theta\right]$, so that $1-p = \mathbb{P}\left[\hat{\theta} < \theta\right]$.

$$\begin{aligned}
\mathbb{E}\left[\widehat{\theta}\right] &= \mathbb{E}\left[\widehat{\theta} \mid \widehat{\theta} = \theta\right] p + \mathbb{E}\left[\widehat{\theta} \mid \widehat{\theta} < \theta\right](1-p) \\
&= \mathbb{E}\left[\theta \mid \widehat{\theta} = \theta\right] p + \mathbb{E}\left[\widehat{\theta} \mid \widehat{\theta} < \theta\right](1-p) \\
&< \mathbb{E}\left[\theta \mid \widehat{\theta} = \theta\right] p + \mathbb{E}\left[\theta \mid \widehat{\theta} < \theta\right](1-p) \\
&= \theta p + \theta(1-p) \\
&= \theta.
\end{aligned}$$

End of Example 85. □

Example 86. Here's another method to estimate the number of sides θ of the hidden dice.

Let Y be the dice roll. We know that since the dice has θ sides, it has expected value

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

So maybe let's take the sample average, which in this case is

$$\frac{X_1 + \ldots + X_4}{4} = \frac{22}{4}$$

and see what value of θ would give us the closest average to this number.

That is, let us take as our estimate $\widehat{\theta}$ the value satisfying the equation

$$\frac{\widehat{\theta} + 1}{2} = \frac{22}{4}. \tag{34}$$

Solving for $\widehat{\theta}$, we get $\widehat{\theta} = 10$.

End of Example 86. □

Theorem 87 (Law of Large Numbers)**.** *Let $X_1, X_2 \ldots$ be a sequence of IID random variables. Such that $\mu := \mathbb{E}[X_1]$ is finite, $\sigma^2 := \mathrm{Var}(X_1)$ is finite, and the fourth moment $\mathbb{E}\left[X_i^4\right] < \infty$. Then*

$$\frac{X_1, \ldots, X_n}{n} \to \mathbb{E}[X_1] \qquad \text{as } n \to \infty$$

This theorem says that if you flip a coin repeatedly, the proportion of heads will converge to 1/2.