7.1: General Concepts and Criteria

Taylor

University of Virginia

Definitions

A **point estimator** of a parameter θ is a single number that can be regarded as a sensible value for θ (e.g $\bar{X}$ is an estimator for μ when your data are iid normal).

A **point estimate** is a particular realization for a point estimator (e.g. after you observe data $\bar{x}$ comes out to be 3.98).

This situation is analagous to the difference between X and x that we talked about before.

Example 7.2 and 7.3 from the book

You want to estimate μ of a normal distribution. You can use the mean, the median, the midrange, or a trimmed mean.

You want to estimate σ^2 of a normal distribution. You can use $S^2 = \frac{\sum_i (X_i - \bar{X})^2}{n-1}$ or $\hat{\sigma}^2 = \frac{\sum_i (X_i - \bar{X})^2}{n}$

Definition

Usually we have a few options for estimators $\hat{\theta}$. How do we compare them?

bias of an estimator $\hat{\theta}$ estimating the parameter θ, written Bias($\hat{\theta}$), is defined as

$$E\left[\hat{\theta} - \theta\right]$$

mean square error of an estimator $\hat{\theta}$ estimating the parameter θ, written MSE($\hat{\theta}$), is defined as

$$E\left[(\hat{\theta} - \theta)^2\right]$$

Note

Note: $(\hat{\theta} - \theta)$, $|\hat{\theta} - \theta|$ or $(\hat{\theta} - \theta)^2$ are all random variables. That's why we take the average.

A result

Here's a useful result. Verify the third equality.

$$\begin{aligned} E\left[(\hat\theta-\theta)^2\right] &= E\left[(\hat\theta - E(\hat\theta) + E(\hat\theta) - \theta)^2\right] \\ &= E\left[\left(\hat\theta - E(\hat\theta)\right)^2 + \left(E(\hat\theta) - \theta\right)^2 + 2\left(\hat\theta - E(\hat\theta)\right)\left(E(\hat\theta) - \theta\right)\right] \\ &= E\left[(\hat\theta - E\hat\theta)^2\right] + E\left[(\theta - E\hat\theta)^2\right] + 0 \\ &= E\left[(\hat\theta - E\hat\theta)^2\right] + (\theta - E\hat\theta)^2 \end{aligned}$$

So $\mathrm{MSE}(\hat\theta) = V(\hat\theta) + \left[\mathrm{Bias}(\hat\theta)\right]^2$

Example 7.4 on page 335

We want to estimate the population proportion of successes (p). Let's take an estimator $\hat{p}$ to be the empirical proportion of successes. More specifically, let X be the number of successes out of n things. We know $X \sim \text{Binomial}(n, p)$. Then $\hat{p} = \frac{X}{n}$. What's its bias? What is its MSE?

Example 7.4 on page 335

We want to estimate the population proportion of successes (p). Let's take an estimator $\hat{p}$ to be the empirical proportion of successes. More specifically, let X be the number of successes out of n things. We know $X \sim \text{Binomial}(n, p)$. Then $\hat{p} = \frac{X}{n}$. What's its bias? What is its MSE?

$$E[\hat{p}] = E\left[\frac{X}{n}\right] = \frac{EX}{n} = \frac{np}{n} = p$$

Example 7.4 on page 335

We want to estimate the population proportion of successes (p). Let's take an estimator $\hat{p}$ to be the empirical proportion of successes. More specifically, let X be the number of successes out of n things. We know $X \sim \text{Binomial}(n, p)$. Then $\hat{p} = \frac{X}{n}$. What's its bias? What is its MSE?

$$E[\hat{p}] = E\left[\frac{X}{n}\right] = \frac{EX}{n} = \frac{np}{n} = p$$

$$E[(\hat{p} - p)^2] = E[(\hat{p} - E[\hat{p}])^2] = \text{Var}(\hat{p})$$

and

$$\text{Var}(\hat{p}) = \text{Var}\left(\frac{X}{n}\right) = \frac{1}{n^2}\text{Var}(X) = \frac{np(1-p)}{n^2} = p(1-p)/n$$

Example 7.4 on page 335

Let $X_1, \ldots, X_n \overset{iid}{\sim}$ Poisson(λ). Let's evaluate the estimator $\hat{\lambda} = \bar{X}$. What's its bias? What is its MSE?

Example 7.4 on page 335

Let $X_1, \ldots, X_n \overset{iid}{\sim} \text{Poisson}(\lambda)$. Let's evaluate the estimator $\hat{\lambda} = \bar{X}$. What's its bias? What is its MSE?

$$E[\hat{\lambda}] = E\left[\frac{\sum X_i}{n}\right] = \lambda$$

Example 7.4 on page 335

Let $X_1, \ldots, X_n \overset{iid}{\sim}$ Poisson(λ). Let's evaluate the estimator $\hat{\lambda} = \bar{X}$. What's its bias? What is its MSE?

$$E[\hat{\lambda}] = E\left[\frac{\sum X_i}{n}\right] = \lambda$$

$$E[(\hat{\lambda} - \lambda)^2] = E[(\hat{\lambda} - E[\hat{\lambda}])^2] = \text{Var}(\hat{\lambda})$$

and

$$\text{Var}(\hat{\lambda}) = \text{Var}\left(\frac{\sum_i X_i}{n}\right) = \frac{1}{n^2}\sum_i \text{Var}(X_i) = \frac{n\lambda}{n^2} = \lambda/n$$

Definitions

In our last example, the mean of the estimator was the thing we were trying to estimate. This is not always the case. When this happens, it has a name.

A point estimator $\hat{\theta}$ for θ is an **unbiased estimator** if

$$E[\hat{\theta}] = \theta$$

for all possible θ. We can also write this as

$$E[\hat{\theta} - \theta] = 0$$

Example

Let's say we're getting ready to observe some data $X_1, \ldots, X_n \overset{i.i.d.}{\sim} \mathcal{N}(\mu, \sigma^2)$. Say we want to estimate σ^2 with $\hat{\sigma}^2 = \frac{\sum_i (X_i - \bar{X})^2}{n}$. Is this estimator unbiased?

We're going to use that variance shortcut formula a lot: $E[X^2] = V[X] + (E[X])^2$.

Also we'll use the fact that $\sum_i (x_i - \bar{x})^2 = \sum_i x_i^2 - n\bar{x}^2$ (we used this last chapter talking about the sampling distribution of $(n-1)S^2/\sigma^2$)

Example 7.5 on page 338

$$\begin{aligned}
E[\hat{\sigma}^2] &= E\left[\frac{\sum_i (X_i - \bar{X})^2}{n}\right] \\
&= \frac{1}{n} E\left[\sum_i X_i^2 - n\bar{X}^2\right] \\
&= \frac{1}{n}\left[\sum_i E[X_i^2] - nE[\bar{X}^2]\right] \\
&= \frac{1}{n}\sum_i E[X_i^2] - E[\bar{X}^2] \\
&= \frac{1}{n}\sum_i \left[V[X_i] + (E[X_i])^2\right] - \left[V[\bar{X}] + (E[\bar{X}])^2\right] \\
&= \frac{1}{n}\sum_i \left[\sigma^2 + \mu^2\right] - \left[\frac{\sigma^2}{n} + \mu^2\right]
\end{aligned}$$

Example (continued)

$$\begin{aligned}
\ldots &= \frac{1}{n}\sum_i \left[\sigma^2 + \mu^2\right] - \left[\frac{\sigma^2}{n} + \mu^2\right] \\
&= \frac{1}{n} n(\sigma^2 + \mu^2) - \frac{\sigma^2}{n} - \mu^2 \\
&= \sigma^2 - \frac{\sigma^2}{n} \\
&= \sigma^2(1 - 1/n) \\
&= \sigma^2 \frac{n-1}{n} \\
&\neq \sigma^2
\end{aligned}$$

Definition

Suppose we took all the unbiased estimators that exist, $\{\hat{\theta}_1, \hat{\theta}_2, \ldots\}$. Each of these has a variance, since it's a random variable. Sometimes their variance is a function of what they're trying to estimate (θ).

Among alll estimators of θ that are unbiased, the one with the minimum variance for all θ, is called the **minimum variance unbiased estimator**. We'll talk about this more in later sections.

Note: sometimes it's called the uniformly minimum variance unbiased estimator to emphasize the fact that the variance is usually a function in θ, and our choice needs to have the smallest variance for all of these θ that you can plug in

Definition

The **standard error** of an estimator $\hat{\theta}$ is its standard deviation.

$$\text{SE}(\hat{\theta}) = \sqrt{V[\hat{\theta}]}$$

Sometimes the SE is a function of unknown parameters. The **estimated standard error** arises when we plug in estimates for the stuff we don't know for sure.

Example 7.10 on page 344

Suppose $X_1, \ldots, X_n \overset{i.i.d.}{\sim} \mathcal{N}(\mu, \sigma^2)$. Suppose further that we know σ^2, but we don't know μ. Let's estimate μ with $\hat{\mu} = \bar{X}$.

$$V[\bar{X}] = V\left[\frac{\sum_i X_i}{n}\right] = \frac{1}{n^2}\sum_i V[X_i] = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

So $\text{SE}(\bar{X}) = \frac{\sigma}{\sqrt{n}}$. If we didn't know σ, we could plug in s, the sample variance for it. This would give us an estimated standard error of $\frac{s}{\sqrt{n}}$.

The bootstrap

A few things

1. This is a quick intro
2. This won't be covered on the test
3. It's more computational than we're accustomed to
4. It's a good last resort tool for when there's no chance we'll figure out a sampling distribution for $\hat{\theta}$

The bootstrap

Here's the basic idea

1. You have some data $x_1, \ldots, x_n$
2. You want to know the distribution of $\hat{\theta} = f(X_1, \ldots, X_n)$
3. Computing this number from your data will give you a number, but it won't give you any idea what the distribution of $\hat{\theta}$ is like

It basically goes like this

1. sample from $x_1, \ldots, x_n$ **with replacement** a bootstrap sample $x_1^*, \ldots, x_n^*$
2. this sample is the same length, so you're probably going to have some repeats in your sample
3. compute an estimate $\hat{\theta}_1^*$ from this bootstrap sample. Now you have one estimate
4. Do this over and over again, B times. You get $\{\hat{\theta}_1^*, \hat{\theta}_2^*, \hat{\theta}_3^*, \ldots, \hat{\theta}_B^*\}$
5. Now you can do whatever you want with $\{\hat{\theta}_1^*, \hat{\theta}_2^*, \hat{\theta}_3^*, \ldots, \hat{\theta}_B^*\}$. You can compute the mean, variance, standard deviation, or just plot a histogram.