Ch 7. Point Estimation —– Modules

30. Unbiased Estimation

31. Maximum Likelihood Estimation and Method of Moments

32. Sampling Distributions

7.30 Unbiased Estimation

Intro to Estimation

Unbiased Estimators

Big Example

Big Example (cont'd)

Mean Squared Error

Intro to Estimation

Definition: A **statistic** is a function of the observations $X_1, \ldots, X_n$, and not explicitly dependent on any unknown parameters.

Examples: $\bar{X}$, S^2.

Statistics are *random variables*. If we take two different samples, we'd expect to get two different values of a statistic.

A statistic is usually used to estimate some unknown **parameter** from the underlying probability distribution of the X_i's.

Let $X_1, \ldots, X_n$ be i.i.d. RV's and let $T(\mathbf{X}) \equiv T(X_1, \ldots, X_n)$ be a statistic based on the X_i's. Suppose we use $T(\mathbf{X})$ to estimate some unknown parameter θ. Then $T(\mathbf{X})$ is called a **point estimator** for θ.

Examples: $\bar{X}$ is usually a point estimator for the mean $\mu = \mathsf{E}[X_i]$, and S^2 is often a point estimator for the variance $\sigma^2 = \mathsf{Var}(X_i)$.

It would be nice if $T(\mathbf{X})$ had certain properties:

* Its expected value should equal the parameter it's trying to estimate, and

* It should have low variance.

Unbiased Estimators

Definition: $T(\mathbf{X})$ is **unbiased** for θ if $\mathsf{E}[T(\mathbf{X})] = \theta$.

Example/Theorem: Suppose $X_1, \ldots, X_n$ are i.i.d. anything with mean μ. Then

$$\mathsf{E}[\bar{X}] \;=\; \mathsf{E}\left[\sum_{i=1}^{n} X_i/n\right] \;=\; \mathsf{E}[X_i] \;=\; \mu.$$

Thus, $\bar{X}$ is always unbiased for μ. This is why $\bar{X}$ is called the **sample mean**.

Example/Theorem: Suppose $X_1, \ldots, X_n$ are i.i.d. anything with mean μ and variance σ^2. Then

$$\mathsf{E}[S^2] = \mathsf{E}\left[\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}\right] = \mathsf{Var}(X_i) = \sigma^2.$$

Thus, S^2 is always unbiased for σ^2. This is why S^2 is called the **sample variance**.

Proof: First, some algebra gives

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$$

$$= \frac{\sum_{i=1}^n (X_i^2 - 2\bar{X}X_i + \bar{X}^2)}{n-1}$$

$$= \frac{\sum_{i=1}^n X_i^2 - 2\bar{X}\sum_{i=1}^n X_i + n\bar{X}^2)}{n-1}$$

$$= \frac{\sum_{i=1}^n X_i^2 - 2n\bar{X}^2 + n\bar{X}^2)}{n-1}$$

$$= \frac{\sum_{i=1}^n X_i^2 - n\bar{X}^2}{n-1}$$

Since $\mathsf{E}[X_1] = \mathsf{E}[\bar{X}]$ and $\mathsf{Var}(\bar{X}) = \mathsf{Var}(X_1)/n = \sigma^2/n$,

$$\begin{aligned}
\mathsf{E}[S^2] &= \mathsf{E}\left[\frac{\sum_{i=1}^n X_i^2 - n\bar{X}^2}{n-1}\right] = \frac{\sum_{i=1}^n \mathsf{E}[X_i^2] - n\mathsf{E}[\bar{X}^2]}{n-1} \\
&= \frac{n}{n-1}\Big(\mathsf{E}[X_1^2] - \mathsf{E}[\bar{X}^2]\Big) \\
&= \frac{n}{n-1}\Big(\mathsf{Var}(X_1) + (\mathsf{E}[X_1])^2 - \mathsf{Var}(\bar{X}) - (\mathsf{E}[\bar{X}])^2\Big) \\
&= \frac{n}{n-1}(\sigma^2 - \sigma^2/n) = \sigma^2. \quad \text{Done.}
\end{aligned}$$

Remark: S is *not* unbiased for the standard dev σ.

Big Example: Suppose that $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \text{U}(0, \theta)$, i.e., the p.d.f. is $f(x) = 1/\theta$, $0 < x < \theta$.

We'll look at three unbiased estimators:

$$Y_1 = 2\bar{X}$$

$$Y_2 = \tfrac{n+1}{n} \max_{1 \le i \le n} X_i$$

$$Y_3 = \begin{cases} 12\bar{X} & \text{w.p. } 1/2 \\ -8\bar{X} & \text{w.p. } 1/2 \end{cases}$$

Easy Estimator: $Y_1 = 2\bar{X}$.

Proof (it's unbiased): $\mathsf{E}[Y_1] = 2\mathsf{E}[\bar{X}] = 2\mathsf{E}[X_i] = \theta$.

Harder Estimator: $Y_2 = \frac{n+1}{n} \max_{1 \le i \le n} X_i$.

Why might this estimator for θ make sense?

Proof: $\mathsf{E}[Y_2] = \frac{n+1}{n}\mathsf{E}[\max_i X_i] = \theta$ iff

$$\mathsf{E}[\max X_i] = \frac{n\theta}{n+1} \quad \text{(what we'll show).}$$

First, let's get the c.d.f. of $M \equiv \max_i X_i$:

$$
\begin{aligned}
\Pr(M \le y) \\
&= \Pr(X_1 \le y \text{ and } X_2 \le y \text{ and } \cdots \text{ and } X_n \le y) \\
&= \Pr(X_1 \le y)\Pr(X_2 \le y)\cdots\Pr(X_n \le y) \quad (X_i\text{'s indep}) \\
&= [\Pr(X_1 \le y)]^n \quad (X_i\text{'s indentically distributed}) \\
&= \left[\int_0^y f_{X_1}(x)\,dx\right]^n = \left[\int_0^y 1/\theta\,dx\right]^n = (y/\theta)^n.
\end{aligned}
$$

This implies that the p.d.f. of M is

$$f_M(y) \equiv \frac{d}{dy}(y/\theta)^n = \frac{ny^{n-1}}{\theta^n},$$

and this implies that

$$\mathsf{E}[M] = \int_0^\theta y f_M(y)\,dy = \int_0^\theta \frac{ny^n}{\theta^n} = \frac{n\theta}{n+1}.$$

Whew! This finally shows that $Y_2 = \frac{n+1}{n}\max_{1\le i\le n} X_i$ is unbiased for θ.

Finally, let's look at...

Stupid Estimator:

$$Y_3 = \begin{cases} 12\bar{X} & \text{w.p. } 1/2 \\ -8\bar{X} & \text{w.p. } 1/2 \end{cases}$$

Ha! It's possible to get a *negative* estimate for θ, which is strange since $\theta > 0$!

Proof (it's unbiased):

$$E[Y_3] = 12E[\bar{X}] \cdot \frac{1}{2} - 8E[\bar{X}] \cdot \frac{1}{2} = 2E[\bar{X}] = \theta.$$

Usually, it's *good* for an estimator to be unbiased, but the "stupid" estimator Y_3 shows that unbiased estimators can sometimes be goofy.

Therefore, let's look at some other properties an estimator can have.

For instance, consider the *variance* of an estimator.

Big Example (cont'd): Again suppose that

$$X_1, \ldots, X_n \overset{\text{iid}}{\sim} \mathsf{U}(0, \theta).$$

Both $Y_1 = 2\bar{X}$ and $Y_2 = \frac{n+1}{n}M$ are unbiased for θ.

Let's find $\text{Var}(Y_1)$ and $\text{Var}(Y_2)$.

$$\text{Var}(Y_1) \;=\; 4\text{Var}(\bar{X}) \;=\; \frac{4}{n} \cdot \text{Var}(X_i) \;=\; \frac{4}{n} \cdot \frac{\theta^2}{12} \;=\; \frac{\theta^2}{3n}.$$

$$\begin{aligned}
\text{Var}(Y_2) &= \left(\frac{n+1}{n}\right)^2 \text{Var}(M) \\
&= \left(\frac{n+1}{n}\right)^2 \text{E}[M^2] - \left(\frac{n+1}{n} \cdot \text{E}[M]\right)^2 \\
&= \left(\frac{n+1}{n}\right)^2 \int_0^\theta \frac{ny^{n+1}}{\theta^n}\,dy - \theta^2 \\
&= \theta^2 \cdot \frac{(n+1)^2}{n(n+2)} - \theta^2 = \frac{\theta^2}{n(n+2)}.
\end{aligned}$$

Thus, both Y_1 and Y_2 are unbiased, but Y_2 has *much lower variance* than Y_1.

Mean Squared Error

Definition: The **MSE** of an estimator $T(\mathbf{X})$ of θ is

$$\text{MSE}(T(\mathbf{X})) \equiv \text{E}[(T(\mathbf{X}) - \theta)^2].$$

Before giving an easier interpretation of MSE, define $\text{Bias}(T(\mathbf{X})) \equiv \text{E}[T(\mathbf{X})] - \theta$.

Remark: Easier interpretation of MSE.

$$\begin{aligned}\mathsf{MSE}(T(\mathbf{X})) &= \mathsf{E}[T^2] - 2\theta\mathsf{E}[T] + \theta^2 \\ &= \mathsf{E}[T^2] - (\mathsf{E}[T])^2 + (\mathsf{E}[T])^2 - 2\theta\mathsf{E}[T] + \theta^2 \\ &= \mathsf{Var}(T) + (\underbrace{\mathsf{E}[T] - \theta}_{\mathsf{Bias}})^2.\end{aligned}$$

So the MSE combines the bias and variance of an estimator.

The lower the MSE the better. If $T_1(\mathbf{X})$ and $T_2(\mathbf{X})$ are two estimators of θ, we'd usually prefer the one with the lower MSE — even if it happens to have higher bias.

Definition: The **relative efficiency** of $T_2(\mathbf{X})$ to $T_1(\mathbf{X})$ is $\text{MSE}(T_1(\mathbf{X}))/\text{MSE}(T_2(\mathbf{X}))$. If this quantity is < 1, then we'd prefer $T_1(\mathbf{X})$.

Example: $X_1, \ldots, X_n \stackrel{\text{iid}}{\sim} \text{U}(0, \theta)$.

Two estimators: $Y_1 = 2\bar{X}$ and $Y_2 = \frac{n+1}{n} \max_i X_i$.

Showed previously that $\text{E}[Y_1] = \text{E}[Y_2] = \theta$ (so both estimators are unbiased).

Also, $\text{Var}(Y_1) = \frac{\theta^2}{3n}$ and $\text{Var}(Y_2) = \frac{\theta^2}{n(n+2)}$.

Thus, $\text{MSE}(Y_1) = \frac{\theta^2}{3n}$ and $\text{MSE}(Y_2) = \frac{\theta^2}{n(n+2)}$, so Y_2 is better.