LECTURE: POINT ESTIMATORS (II)

1. Mean-Squared Error

Recall:

Let $\hat{\theta}$ be an estimator for θ, then

$$\text{Bias}(\hat{\theta}) = E(\hat{\theta}) - \theta$$

This is one measure of the "goodness" of $\hat{\theta}$. A perhaps better one is the Mean Square Error:

Definition:

$$\text{MSE}(\hat{\theta}) = E[(\hat{\theta} - \theta)^2]$$

This is the average square distance from $\hat{\theta}$ to θ

Magic MSE Formula:

$$\text{MSE}(\hat{\theta}) = [\text{Bias}(\hat{\theta})]^2 + \text{Var}(\hat{\theta})$$

This is because:

$$
\begin{aligned}
\text{MSE}(\hat{\theta}) &= E[(\hat{\theta} - \theta)^2] \\
&= E[(\hat{\theta} - E(\hat{\theta})) + (E(\hat{\theta}) - \theta)]^2 \\
&= E[(\hat{\theta} - E(\hat{\theta}))^2] + 2E[(\hat{\theta} - E(\hat{\theta}))(\underbrace{E(\hat{\theta}) - \theta}_{\text{constant}})] + E[\underbrace{(E(\hat{\theta}) - \theta)^2}_{\text{constant}}] \\
&= \text{Var}(\hat{\theta}) + 2(E(\hat{\theta}) - \theta)E[(\hat{\theta} - E(\hat{\theta}))] + (E(\hat{\theta}) - \theta)^2 \\
&= \text{Var}(\hat{\theta}) + 2(E(\hat{\theta}) - \theta)\left(\cancel{E(\hat{\theta})} - \cancel{E(\hat{\theta})}\right) + [\text{Bias}(\hat{\theta})]^2 \\
&= \text{Var}(\hat{\theta}) + [\text{Bias}(\hat{\theta})]^2
\end{aligned}
$$

Low bias is a good quality for an estimator, but ideally we would also like to have low variance because we want the $\hat{\theta}$ to be close to θ

In general, there is a trade-off between bias and variance. For a given MSE, if we wish $\hat{\theta}$ to have lower bias, then we must accept a higher variance, and vice-versa.

Example 1:

(a) Find the MSE of the sample mean $\bar{Y}$

Since Bias $\left(\bar{Y}\right) = 0$ by the Magic MSE formula, MSE $\left(\bar{Y}\right) =$ Var $\left(\bar{Y}\right)$

We've also shown that $\text{Var}(\bar{Y}) = \frac{\sigma^2}{n}$ hence $\text{MSE}(\bar{Y}) = \frac{\sigma^2}{n}$

Note: $\text{MSE}(\bar{Y})$ goes to 0 as $n \to \infty$. This makes intuitive sense that a larger sample provides a better estimator for the population mean.

(b) Find the MSE of the sample proportion $\hat{p} = \frac{Y}{n}$

Recall $Y \sim \text{Binomial}(n, p)$

Once again $\text{Bias}(\hat{p}) = 0$ hence $\text{MSE}(\hat{p}) = \text{Var}(\hat{p})$

Since the variance of a binomial random variable is $np(1-p)$, we get

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

2. Differences

Sometimes we're in the difference between two populations

Example 2:

Who gets more sleep: First-years or Seniors?

In this case, the parameter of interest is the **difference** in the mean amount of sleep between first-years and seniors.

Let μ_1 and σ_1^2 be the mean and var of the amount of sleep of first-years

Let μ_2 and σ_2^2 be the mean and var of the amount of sleep of seniors

Then our parameter of interest is $p = \mu_1 - \mu_2$

An estimator for p is $\hat{p} = \bar{Y}_1 - \bar{Y}_2$

Where $\bar{Y}_1$ is the sample mean of n_1 first-year students and $\bar{Y}_2$ is the sample mean of n_2 seniors.

By taking expected values, we can show that $\bar{Y}_1 - \bar{Y}_2$ is unbiased.

Moreover, by independence and $\operatorname{Var}(aX+b) = a^2 \operatorname{Var}(aX)$ we get

$$\operatorname{Var}(\bar{Y}_1 - \bar{Y}_2) = \operatorname{Var}(\bar{Y}_1) + \operatorname{Var}(-\bar{Y}_2) = \operatorname{Var}(\bar{Y}_1) + \operatorname{Var}(\bar{Y}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

Example 3:

Suppose we are interested the preference for Chocolate Ice Cream in hot vs cold areas in the US

Here "hot" means living in a city whose average temperature is $\geq 80F$

Let p_1 is the proportion of Chocolate supporters in hot areas

Let p_2 is the proportion of Chocolate supporters in cold areas

Then the parameter of interest is $p = p_1 - p_2$, and an estimator of p is

$$\hat{p} = \frac{Y_1}{n_1} - \frac{Y_2}{n_2}$$

Where we survey n_1 people in a hot area and n_2 people in a cold area.

The expected value of $\hat{p}$ is $p_1 - p_2$, so this estimator is unbiased, and

$$\operatorname{Var}(\hat{p}_1 - \hat{p}_2) = \operatorname{Var}(\hat{p}_1) + \operatorname{Var}(-\hat{p}_2) = \operatorname{Var}(\hat{p}_1) + \operatorname{Var}(\hat{p}_2) = \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}$$

Summary: We summarize these common estimators for population mean and proportion in the following table:

Parameter of Interest	Sample Size	Estimator	Expected Value	Variance
μ	n	$\bar{Y}$	μ	$\frac{\sigma^2}{n}$
p	n	$\hat{p} = \frac{Y}{n}$	p	$\frac{p(1-p)}{n}$
$\mu_1 - \mu_2$	n_1 and n_2	$\bar{Y}_1 - \bar{Y}_2$	$\mu_1 - \mu_2$	$\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$
$p_1 - p_2$	n_1 and n_2	$\hat{p}_1 - \hat{p}_2$	$p_1 - p_2$	$\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}$

Note: The standard deviation of an estimator is sometimes called the **standard error**

3. Sample Variance

Recall: Sample Variance

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

Question: Why the $n-1$?

In fact, let $\bar{S}$ be the same definition but with n

Definition: Sample Variance at Home

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

We will now show that $\bar{S}^2$ is biased whereas S^2 is unbiased! This explains why we use the definition with $n-1$

Claim:

$\bar{S}^2$ is biased

STEP 1: Find a nice formula for $\sum_{i=1}^{n}(Y_i - \bar{Y})^2$

$$\begin{aligned}\sum_{i=1}^{n}(Y_i - \bar{Y})^2 &= \sum_{i=1}^{n}(Y_i^2 - 2Y_i\bar{Y} + \bar{Y}^2) \\ &= \sum_{i=1}^{n} Y_i^2 - 2\bar{Y}\sum_{i=1}^{n} Y_i + n\left(\bar{Y}^2\right) \quad (\bar{Y} \text{ is constant }) \\ &= \sum_{i=1}^{n} Y_i^2 - 2n\bar{Y}^2 + n\bar{Y}^2 \quad (\text{ def of } \bar{Y}) \\ &= \left(\sum_{i=1}^{n} Y_i^2\right) - n\bar{Y}^2\end{aligned}$$

STEP 2: Take expected values:

$$E\left[\sum_{i=1}^{n}(Y_i - \bar{Y})^2\right] = E\left(\sum_{i=1}^{n} Y_i^2\right) - nE(\bar{Y}^2) = \left(\sum_{i=1}^{n} E(Y_i^2)\right) - nE(\bar{Y}^2)$$

STEP 3: By the Magic Variance formula

$$E(Y_i^2) = \text{Var}(Y_i) + [E(Y_i)]^2 = \sigma^2 + \mu^2$$

$$\text{Similarly } E(\bar{Y}^2) = \text{Var}(\bar{Y}) + [E(\bar{Y})]^2 = \frac{\sigma^2}{n} + \mu^2$$

STEP 4: We then plug this into **STEP 2**

$$E\left[\sum_{i=1}^{n}(Y_i - \bar{Y})^2\right] = \left(\sum_{i=1}^{n}(\sigma^2 + \mu^2)\right) - n\left(\frac{\sigma^2}{n} + \mu^2\right) = n\sigma^2 + \cancel{n\mu^2} - \sigma^2 - \cancel{n\mu^2}$$

$$=(n-1)\sigma^2$$

STEP 5: Grand Finale

$$E(\bar{S}^2) = E\left[\frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})^2\right] = \frac{1}{n}E\left[\sum_{i=1}^{n}(Y_i - \bar{Y})^2\right] = \frac{1}{n}(n-1)\sigma^2 = \left(\frac{n-1}{n}\right)\sigma^2$$

Since $E(\bar{S}^2) \neq \sigma^2$, we get that $\bar{S}^2$ is biased □

Question: How to fix this?

Use $n-1$ instead of n to get S^2

Claim:

S^2 is unbiased

STEP 1: On the one hand, by our calculation above, we have

$$E\left[\left(\frac{n}{n-1}\right)\bar{S}^2\right] = \left(\frac{n}{n-1}\right)E(\bar{S}^2) = \left(\frac{n}{n-1}\right)\left(\frac{n-1}{n}\right)\sigma^2 = \sigma^2$$

So $\left(\frac{n}{n-1}\right)\bar{S}^2$ is unbiased

STEP 2: On the other hand, this is precisely S^2 because

$$\left(\frac{n}{n-1}\right)\bar{S}^2 = \frac{n}{n-1}\left(\frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})^2\right) = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2 = S^2$$

So S^2 is unbiased □