Chapter 6 Point Estimation:

6.1 Often in statistics, we want to estimate parameters of a population distribution.

For example, we may want to estimate p, the true proportion of certain electrical components which are defective. Say we get a sample of 25 components, and 4 are defective. We use our sample to estimate p. $\text{Bin}(25, p)$

For the estimator, we write $\hat{p}$, "p hat".

$$\hat{p} = .16 = \frac{4}{25}$$

This is an example of a point estimate.

A <u>point estimate</u> for a parameter θ ("theta") is a single number that is a sensible value for θ. It is obtained by selecting a suitable statistic and computing its value from the sample data. The selected statistic is called the <u>point estimator</u> of θ.

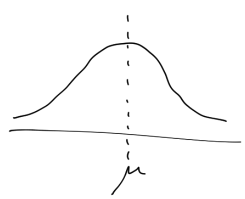
<u>Example 1</u>: We want to study the breakdown voltage for pieces of epoxy resin. We believe the break down voltages are normally distributed*, but we don't know the mean μ or variance σ^2. We get a sample $x_1, \ldots, x_n$. What are some possible point estimators for μ? What are some point estimators for σ^2?

<u>Point Estimators for μ</u>:

sample mean $\overline{X}$, $\bar{x} = \dfrac{x_1 + x_2 + \ldots + x_n}{n}$

sample median $\tilde{X}$

⋮



<u>Point Estimators for σ^2</u>:

$Var(X) = E[X^2] - E[X]^2$,

$$\hat{\sigma}^2 = \frac{1}{n}\sum x_i^2 + \bar{x}^2$$

$$\hat{\sigma}^2 = S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

<u>Unbiased Estimators</u>: A point estimator $\hat{\theta}$ is an <u>unbiased estimator</u> of θ if $E[\hat{\theta}] = \theta$ for every possible value of θ. If $\hat{\theta}$ is not unbiased, $E[\hat{\theta}] - \theta$ is called the <u>bias</u> of $\hat{\theta}$.

Example 2: Let $X \sim Bin(n,p)$. Say we know the value of n, but not p.

We estimate $\hat{p} = \frac{X}{n}$. Is $\hat{p}$ an unbiased estimator of p?

$$E[\hat{p}] = E\left[\frac{X}{n}\right] = \frac{1}{n} E[X] = \frac{1}{n} \cdot np = p$$

lookup mean of
Bin(n,p) distribution

So $\hat{p}$ is an unbiased estimator for p.

Example 3: Say certain reaction times are uniformly distributed from 0 to Θ.

We collect a sample $X_1, \dots, X_n$ and estimate $\hat{\Theta} = \max(X_1, \dots, X_n)$.

Explain how we can know $\hat{\Theta}$ is biased without doing calculations.

All of data $X_1, \dots, X_n$ is less than Θ

so $\hat{\Theta} = \max(X_1, \dots, X_n)$ is also less than Θ.

So $E[\hat{\Theta}] < \Theta$.

Principle of Unbiased Estimation: When choosing among several different estimators for Θ, select one that is unbiased.

Proposition: Let $X_1, \dots, X_n$ be a random sample from a distribution with mean μ and variance σ^2. Then the estimator

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

is unbiased for estimating σ^2. $E[\hat{\sigma}^2] = \sigma^2$.

Also $\overline{X}$ is an unbiased estimator for μ. $E[\overline{X}] = \mu$.

Example 4: Suppose we have samples of the growth of two types of trees over 1 year. $X_1, X_2, \ldots, X_5$ are measurements of the growth of 5 trees of the first type, with mean μ_1 and variance σ^2. $Y_1, \ldots, Y_7$ are measurements of 7 trees of the second type with mean μ_2 and the same variance σ^2.

Let $\overline{X} = \dfrac{X_1 + \ldots + X_5}{5}$ and $\overline{Y} = \dfrac{Y_1 + \ldots + Y_7}{7}$.

Assume $X_1, \ldots, X_5, Y_1, \ldots, Y_7$ are all independent.

Let S_1^2 be the sample variance of the X_i's and S_2^2 be the sample variance of the Y_i's.

a) Show that $\overline{X} - \overline{Y}$ is an unbiased estimator of $\mu_1 - \mu_2$.

b) For which value of k is $\hat{\sigma}^2 = k(S_1^2 + S_2^2)$ an unbiased estimator for σ^2?

a) $E[\overline{X} - \overline{Y}] = E[\overline{X}] - E[\overline{Y}] = \mu_1 - \mu_2$. So $\overline{X} - \overline{Y}$ is unbiased for estimating $\mu_1 - \mu_2$

b) $E[k(S_1^2 + S_2^2)] = k(E[S_1^2 + S_2^2]) = k(\underbrace{E[S_1^2]}_{\sigma^2} + \underbrace{E[S_2^2]}_{\sigma^2}) = k(\sigma^2 + \sigma^2)$

$= 2k\sigma^2$

Want this to be σ^2. $2k = 1$, $\boxed{k = \frac{1}{2}}$.

The Standard Error of an Estimator $\hat{\theta}$ is its standard deviation $\sigma_{\hat{\theta}} = \sqrt{Var(\hat{\theta})}$. This represents a typical deviation between the estimate and the value of θ.

Example 5: Find the standard error of $\hat{\theta} = \overline{X} - \overline{Y}$ from Example 4.

$$Var(\overline{X} - \overline{Y}) = Var(\overline{X}) + Var(-\overline{Y}) = Var(\overline{X}) + (-1)^2 Var(\overline{Y})$$

independent

$$= Var(\overline{X}) + Var(\overline{Y})$$

$$= \frac{\sigma^2}{5} + \frac{\sigma^2}{7} = \frac{12}{35}\sigma^2$$

$$SE(\overline{X} - \overline{Y}) = \sqrt{Var(\overline{X} - \overline{Y})} = \sqrt{\frac{12}{35}\sigma^2} = \boxed{\sqrt{\frac{12}{35}}\,\sigma}$$