Lecture 26

Todays topic is estimation. That is how do we estimate a parameter based on sample data

Notation

θ – general symbol to represent unknown parameter

$\hat{\theta}$ – symbol representing an estimate of θ

Some parameter examples we have already encountered along with an estimate.

Parameter	Estimate
μ	$\bar{X} = \frac{\sum X_i}{n}$
σ^2	$s^2 = \frac{\sum (X_i - \bar{X})^2}{n-1}$
σ	$s = \sqrt{\frac{\sum (X_i - \bar{X})^2}{n-1}}$
p	$\hat{p} = \frac{X}{n}$

Types of estimates

- Point estimate – a single number
- Interval estimate – a range of numbers

Point estimate

A point estimate of a parameter θ is a single number $\hat{\theta}$, that could be viewed as a sensible value for θ, as computed from a given sample of data.

<u>Note</u> we use estimator to refer to the method and estimate to refer to the value.

<u>Note</u> There is often more than one reasonable point estimator.

eg consider normal data. We could estimate μ using the sample mean $\bar{X}$ or using the sample median $\tilde{X}$ or to be robust we might use a trimmed mean $\bar{X}_{tr}$. We could estimate σ^2 using

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} \quad \text{or} \quad \hat{\sigma}^2 = \frac{\sum (x_i - \bar{x})^2}{n}$$

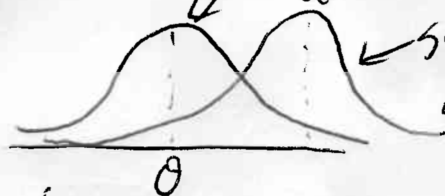
How do we choose between estimators?

Bias - is the estimator on target or not on average.
want this small or non existant.

Variance - how variable is the estimator, is it likely that we will be far away from the target even if we are on target on average. want this small also.

Unbiased: $E(\hat{\theta}) = \theta$ (ie on average we are on target)

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



when choosing between estimators choose the one which is unbiased.

eg let $X_1, X_2, \ldots X_n$ be a SRS from a dist with mean μ and variance σ^2.

which of

$$s^2 = \frac{\sum (X_i - \overline{X})^2}{n-1} \quad \text{and} \quad \hat{\sigma}^2 = \frac{\sum (X_i - \overline{X})^2}{n-1}$$

is unbiased?

$$E[s^2] = E\left[\frac{1}{n-1} \sum (X_i - \overline{X})^2\right]$$

$$= \frac{1}{n-1} E\left[\sum X_i^2 - \frac{1}{n}(\sum X_i)^2\right]$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n} E[X_i^2] - \frac{1}{n} E((\sum X_i)^2)\right]$$

recall that $E[X^2] = Var(X)^2 + (E[X])^2$
$= \sigma^2 + \mu^2$

$$= \frac{1}{n-1}\left[\sum(\sigma^2+\mu^2) - \frac{1}{n}\left[Var(\sum X_i) + [E(\sum X)]^2\right]\right]$$

$$= \frac{1}{n-1}\left[(n\sigma^2 + n\mu^2) - \frac{1}{n}(n\sigma^2 + n^2\mu^2)\right]$$

$$= \frac{1}{n-1}\left[n\sigma^2 + n\mu^2 - \sigma^2 - n\mu^2 \right)$$

$$= \frac{(n-1)\sigma^2}{n-1} = \sigma^2$$

So $E[S^2] = \sigma^2$ and S^2 is unbiased estimator of σ^2.

Note that $\frac{(n-1)}{n} S^2 = \frac{(n-1)}{n}\frac{1}{n-1}\sum(X_i - \bar{X})^2$

$$= \frac{1}{n}\sum(X_i - \bar{X})^2 = \hat{\sigma}^2$$

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

$$= \frac{n-1}{n}\sigma^2$$

is not unbiased.

How about estimators for μ?

$$E[\bar{X}] = E\left[\frac{\sum X_i}{n}\right] = \frac{1}{n} E\left[\sum X_i\right] = \frac{1}{n}\sum E[X_i]$$

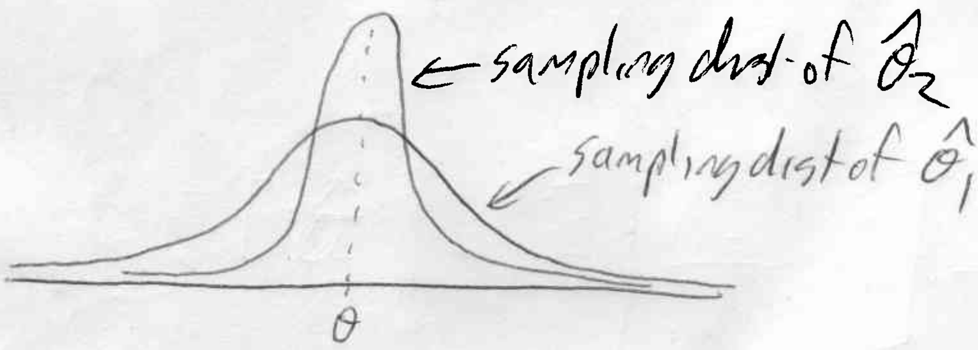
$$= \frac{1}{n} n\mu = \mu$$

So the sample mean $\bar{X}$ is unbiased estimate of μ.

(note it can also be shown that $\tilde{X}$ the sample median and trimmed means are unbiased if we assume that the distribution is symmetric).

MVUE — Minimum Variance Unbiased Estimator

— choose between two or more unbiased estimators $\hat{\theta}_1, \hat{\theta}_2 \ldots$ of θ by ~~taking~~ one with smallest variance



Note It can be shown that $\bar{X}$ is MVUE for μ for samples from $N(\mu, \sigma)$ distribution.

Standard error

The standard error (SE) is an estimate of the standard deviation of an estimator

eg

$$SE(\bar{X}) = \frac{\sigma}{\sqrt{n}} \quad (\text{if } \sigma \text{ is known}$$

$$SE(\bar{X}) = \frac{S}{\sqrt{n}} \quad (\text{if } \sigma \text{ is unknown})$$

It is quite common to report both the estimate and estimated SE together (rather than the estimate alone).

Some Examples

1) of n_1 randomly selected male smokers, X_1 smoked filter cigarettes. of n_2 randomly selected females X_2 smoked filter cigarettes. Let p_1 and p_2 denote probabilities of smoking filtered

cigarettes, for males and females respectively. Note this implies $X_1 \sim Bin(n_1, p_1)$ and $X_2 \sim Bin(n_2, p_2)$

(a) Show that $(X_1/n_1) - (X_2/n_2)$ is unbiased estimator of $p_1 - p_2$

$$E\left[\frac{X_1}{n_1} - \frac{X_2}{n_2}\right] = E\left[\frac{X_1}{n_1}\right] - E\left[\frac{X_2}{n_2}\right]$$

$$= \frac{1}{n_1} E[X_1] - \frac{1}{n_2} E[X_2]$$

$$= \frac{1}{n_1} n_1 p_1 - \frac{1}{n_2} n_2 p_2$$

$$= p_1 - p_2$$

(b) what is the SE of this estimator?

$$Var\left(\frac{X_1}{n_1} - \frac{X_2}{n_2}\right) = Var\left(\frac{X_1}{n_1}\right) + Var\left(\frac{X_2}{n_2}\right)$$

$$= \frac{1}{n_1^2} Var(X_1) + \frac{1}{n_2^2} Var(X_2)$$

$$= \frac{1}{n_1^2} n_1 p_1 (1-p_1) + \frac{1}{n_2^2} n_2 p_2 (1-p_2)$$

$$= \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}$$

so $SE\left(\frac{X_1}{n_1} - \frac{X_2}{n_2}\right) = \sqrt{\frac{p(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$

Note in practise we would estimate $SE\left(\frac{X_1}{n_1} - \frac{X_2}{n_2}\right)$

with $\hat{p}_1 = \frac{x_1}{n_1}$ $\quad \hat{p}_2 = \frac{x_2}{n_2}$

and $SE\left(\frac{X_1}{n_1} - \frac{X_2}{n_2}\right) = \sqrt{\frac{\frac{x_1}{n_1}\left(1-\frac{x_1}{n_1}\right)}{n_1} + \frac{\frac{x_2}{n_2}\left(1-\frac{x_2}{n_2}\right)}{n_2}}$