3Y03 - 3J04 PROBABILITY & STATISTICS FOR (CIVIL) ENGINEERING

C01 - Lecture 22

Last time How 'good' is an estimator $\hat{\Theta}$ of θ?

BIAS $= E(\hat{\Theta}) - \theta$. $\boxed{\text{If } E(\hat{\Theta}) = \theta,\ \hat{\Theta} \text{ is unbiased.}}$

If $\hat{\Theta}_1$ and $\hat{\Theta}_2$ are both unbiased estimators of θ, the 'better' estimator is the one with the smaller variance $V(\hat{\Theta}_i)$.

$\longrightarrow$ MVUE (Minimum Variance Unbiased Estimator) has smallest.

The standard error of $\hat{\Theta}$ (an unbiased estimator of θ) is $\sigma_{\hat{\Theta}} = \sqrt{V(\hat{\Theta})}$.

Example Suppose we are sampling from a normal distribution mean μ and variance σ^2. Then we know $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$

The standard error of $\bar{X}$ is $\sigma_{\bar{X}} = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}}$.

What if we don't know σ?

We can estimate it with S (even though biased)

& then the estimated standard error of $\bar{X}$ is $\hat{\sigma}_{\bar{X}} = \frac{S}{\sqrt{n}}$.

In general the estimated standard error of an estimator $\hat{\Theta}$ (for θ) is $\hat{\sigma}_{\hat{\Theta}}$ (or $= se(\hat{\Theta})$) & is obtained by substituting into formula for $\sigma_{\hat{\Theta}}$ estimates for any parameters appearing that are unknown.

Biased Estimators

Here we have the mean square error (MSE) of $\hat{\Theta}$, which incorporates the bias of $\hat{\Theta}$:

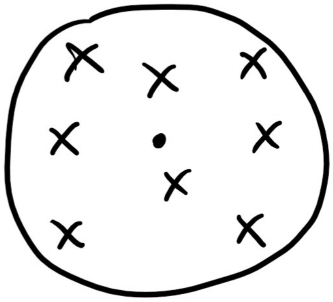
Have a go at working this out!

$$MSE(\hat{\Theta}) = E\left((\hat{\Theta} - \theta)^2\right) = \ldots = V(\hat{\Theta}) + (\text{bias})^2$$

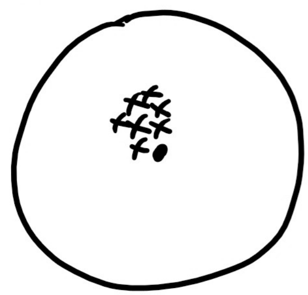
Generalizes standard error^2:

$$MSE(\hat{\Theta}) = V(\hat{\Theta}) = \sigma^2_{\hat{\Theta}} \quad \text{if } \hat{\Theta} \text{ is unbiased.}$$

Can use MSE to compare any type of estimators: (unbiased or biased)



$E(\hat{\Theta})$ = bullseye
wide spread
(high variance)



$E(\hat{\Theta})$ – near bullseye
small spread
(low variance)

We might well prefer to have this biased estimator of bullseye.

More formally:

We may well find that $MSE(\hat{\Theta}_1) < MSE(\hat{\Theta}_2)$ even though $\hat{\Theta}_1$ biased, $\hat{\Theta}_2$ unbiased

More precisely the relative efficiency of $\hat{\Theta}_1$ and $\hat{\Theta}_2$ is

Smaller MSE = more efficient

$$\frac{MSE(\hat{\Theta}_1)}{MSE(\hat{\Theta}_2)} \longrightarrow \text{if} \begin{cases} < 1 & \hat{\Theta}_1 \text{ is more efficient} \\ > 1 & \hat{\Theta}_2 \text{ is more efficient} \end{cases}$$

An optimal estimator for θ is $\hat{\Theta}$ with minimal MSE (for θ).

8 Confidence Intervals

We choose $\hat{\Theta}$, estimator for θ.
Given data set $\{x_1, \ldots, x_n\}$, we use $\hat{\Theta}$ to get an estimate $\hat{\theta}$ (a #) for θ.
How close to true value of θ is $\hat{\theta}$? How confident are we in our estimate?

If we repeat experiment lots of times, we'd get lots of estimates for θ & could build up an idea of a plausible interval in which true value θ is likely to lie.

We can work all of this out with just one data set:

Definition A $(1-\alpha)$ (or $100(1-\alpha)\%$) confidence interval (C.I.) for θ is an interval

$$\left(L(x_1, \ldots, x_n),\ R(x_1, \ldots, x_n) \right) \text{ such that}$$

$$P(L \leq \theta \leq R) = 1-\alpha.$$

small

Typically $1-\alpha = 0.9, 0.95, 0.99$ (or $100(1-\alpha)\% =$ $90\%, 95\%, 99\%$)

The C.I. is an interval estimator for θ.

How to find such an interval?

Depends on underlying distribution & on information already known (or not known)

C.I. for Mean of Normal Population — when Variance is known

i.e. we know the value of σ^2 (hence of σ).

- unrealistic
- gives basic ideas

We use $\bar{X}$ to estimate μ.

We know $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$ if population $\sim N(\mu, \sigma^2)$

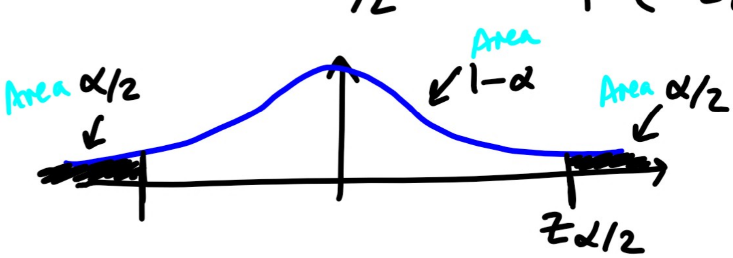
or $Z = \dfrac{\bar{X} - \mu}{(\sigma/\sqrt{n})} \sim N(0, 1)$ if you prefer

The obvious way to find L, R with

$$P(L \leq \mu \leq R) = 1 - \alpha \text{ is to}$$

make (L, R) symmetric (with μ in the middle)

i.e. want $z_{\alpha/2}$ with $P(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}) = 1 - \alpha$



So $P(Z \leq z_{\alpha/2}) = 1 - \frac{\alpha}{2}$

i.e. $P\left(-z_{\alpha/2} \le \dfrac{\bar{X}-\mu}{(\sigma/\sqrt{n})} \le z_{\alpha/2}\right) = 1-\alpha$

$\vdots$

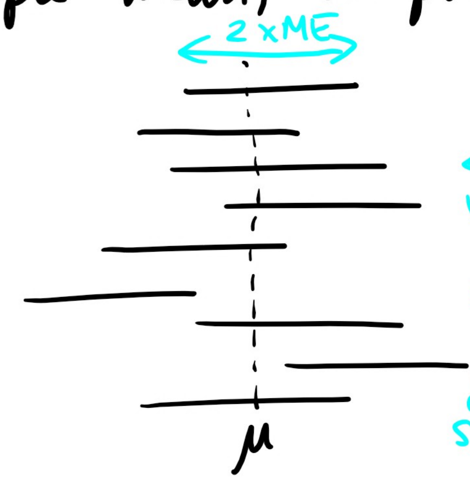
i.e. $P\Big(\underbrace{\bar{X} - z_{\alpha/2}\left(\tfrac{\sigma}{\sqrt{n}}\right)}_{L(X_1,\dots,X_n)} \le \mu \le \underbrace{\bar{X} + z_{\alpha/2}\left(\tfrac{\sigma}{\sqrt{n}}\right)}_{R(X_1,\dots,X_n)}\Big) = 1-\alpha$

So a $100(1-\alpha)\%$ confidence interval for μ is

$$\left(\bar{x} - z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right),\ \bar{x} + z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)\right)$$ where

$\bar{x}$ is sample mean, computed using the data.

If one such confidence were found for each new data set, the idea is that μ would lie in $100(1-\alpha)\%$ of the intervals generated in this way.



ME = Margin of Error $= z_{\frac{\alpha}{2}}\left(\dfrac{\sigma}{\sqrt{n}}\right)$

Width of $100(1-\alpha)\%$ C.I. always 2×ME – does not depend on data in this situation

The amount that the $100(1-\alpha)\%$ confidence interval extends either side of $\bar{x}$.

μ lies in the above confidence interval for $100(1-\alpha)\%$ of expeiments/data sets.

<u>Example</u> Find a 93% C.I. for the mean μ of a Normally distributed population (standard deviation $\sigma = 5.7$) with sample size $n = 10$.

Solution $\quad ME = z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right) = z_{\alpha/2}\left(\frac{5.7}{\sqrt{10}}\right)$

$100(1-\alpha) = 93$ so $1-\alpha = 0.93 \Rightarrow \alpha = 0.07$

So $\frac{\alpha}{2} = 0.035$.

To find $z_{\alpha/2}$ we use $P(Z \le z_{\alpha/2}) = 1 - \frac{\alpha}{2} = 0.965$.

Use Normal Table in reverse : 1.81

(Or you could interpolate between 1.81 and 1.82.)

So $ME = (1.81) \times \left(\frac{5.7}{\sqrt{10}}\right) =$ a #. 3.26