Lecture 34 (Gr 25-27)

14 November 2022 10:34

Chapter 9

Estimation problems

Statistical inference → You make predictions about population from samples.

Classical method
→ You just rely on the sample data.
↓
We will focus on classical method.

Bayesian Approach
Sample data + some information about the distribution of population (old data)

- The problems for statistical inference can be classified into two types:-

Estimation problem

Example

Elections → A party, B Party

60%

Party A wants to predict what %age of people will vote for them?

→ You want to estimate something. (Chapter 9)

Testing of Hypothesis

- Somebody is claiming that Party A will win the Loksabha elections. (Hypothesis)
- On the basis of your sample data you will reject or accept the hypothesis.

Chapter 10

Estimation problems

① θ Population parameters (μ & σ)

② $\hat{\theta}$ → sample statistics.

Example of sample statistics

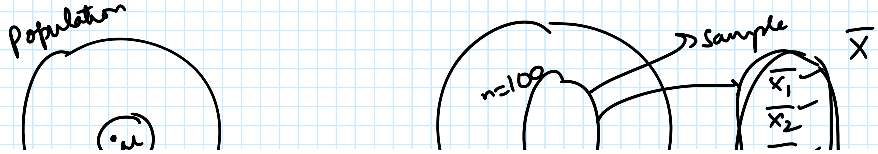
$\bar{X}$ ✓

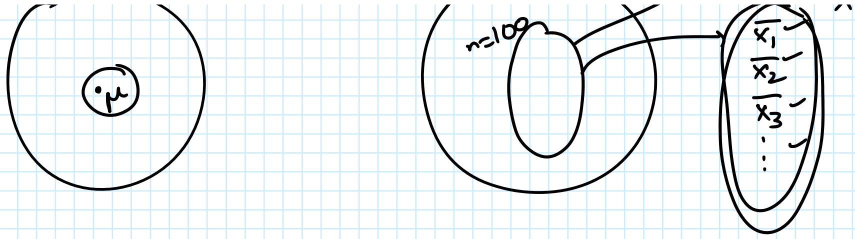
S^2 ✓

t ✓ $= \frac{(n-1)S^2}{\sigma^2}$

Definition

- **Point estimation** of a **population parameter** θ, is **a single value** $\hat{\theta}$ of **Sample statistic** $\hat{\theta}$.





- Question is how do we choose $\hat{\Theta}$ (single value) or $\hat{\Theta}$ (R.V.). There are some expectation from $\hat{\Theta}$.

① Unbiasedness :- A statistic $\hat{\Theta}$ is said to be unbiased estimator of the population parameter θ if

$$E(\hat{\theta}) = \theta$$

H.W Show that S^2 is an unbiased estimator of the parameter σ^2.

Here you have to show $E(S^2) = \sigma^2$

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

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \rightarrow \text{Not unbiased}$$

Do it yourself.

② Suppose you have $\hat{\Theta}_1$ and $\hat{\Theta}_2$ sample statistics for population parameter θ both unbiased, then which one to choose?

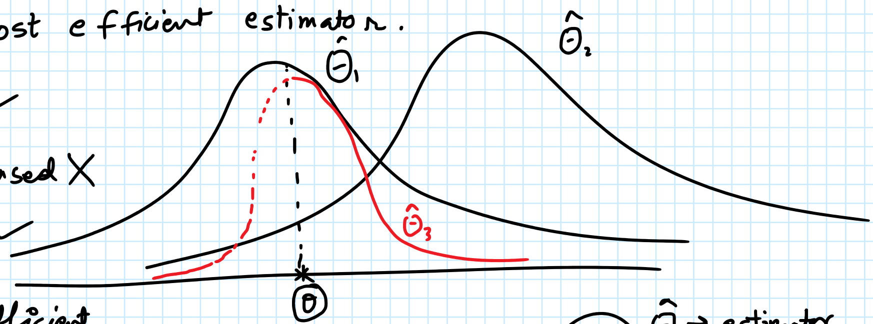
The one with less variance should be chosen.

Defintion If we consider all possible unbiased estimators of some population parameter θ, the one with the smallest variance is called the most efficient estimator.

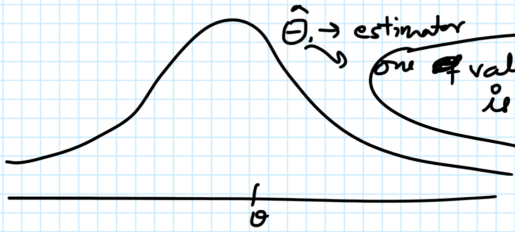
$\hat{\Theta}_1 \rightarrow$ unbiased ✓

$\hat{\Theta}_2 \rightarrow$ Not unbiased ✗

$\hat{\Theta}_3 \rightarrow$ unbiased ✓



$\hat{\Theta}_3$ is most efficient.



$\hat{\Theta} \rightarrow$ estimator

one value of it is called point estimator. $\rightarrow$ mostly $= E(\hat{\Theta})$

- Since point estimators are difficult to achieve. we rely more on interval estimates.

$\hat{\Theta}$ $\hat{\theta}$

- Since point estimators are difficult to achieve. we rely more on interval estimates.
- What is interval estimation?

$\hat{\Theta}$ $\hat{\theta}$

$[\hat{\Theta}_L \ \hat{\Theta}_U]$

Suppose θ is population parameter & you find numbers $\hat{\Theta}_L$ & $\hat{\Theta}_U$ s.t.

$$P\left(\theta \in (\hat{\Theta}_L, \hat{\Theta}_U)\right) = 1-\alpha$$

then the interval $(\hat{\Theta}_L \ \hat{\Theta}_U)$ is called a $(1-\alpha)\,100\%$ Confidence interval.

Example If you find out that

$$P\left(\mu \in (-1, 3)\right) = 97\%$$

then $(-1, 3)$ is 97% confidence interval.

- 97% confidence interval $(-100, 100)$ X
- 93% " " $(-3, 3)$
- 90% " " $(-2.5, 3)$

Both can work depending on what you are looking for.

How to find confidence intervals for μ (population mean)

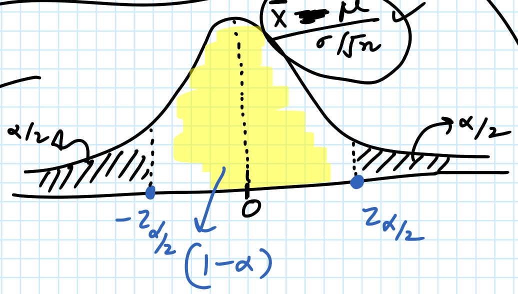
$\theta = \mu$

$\hat{\Theta} \rightsquigarrow \bar{X}$ (sampling distribution of mean)

} CLT

$\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$

(This is true for all n if original population is normally distributed otherwise for $n \geq 30$).



$$* \quad P\left(\hat{\Theta}_L < \mu < \hat{\Theta}_U\right) = 1-\alpha$$

$$-z_{\alpha/2} < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < z_{\alpha/2}$$

where $z_{\alpha/2}$ will be found from the table of standard normal distribution

$$-z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \bar{X}-\mu < z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

$$z_{\alpha/2}\frac{\sigma}{\sqrt{n}} > \mu - \bar{X} > -z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

$$\bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}} > \mu > \bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

$$P\Big(\mu \in \Big(\underbrace{\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}}_{\hat{\Theta}_L},\ \underbrace{\bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}}_{\hat{\Theta}_U}\Big)\Big) = 1-\alpha$$

For μ, the $(1-\alpha)100\%$ confidence interval is

$$\left[\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}},\ \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$