

# Lecture 9: Estimation

---

Lecturer: Jie Fu

EEL 3850

## Motivating problem

$X$  : height of 3rd students in population

$$\mathbb{E}(X) = \mu_X$$

- Estimating the average height of 3<sup>rd</sup> grade students in schools

Data sample 1:

53.49	51.59	53.94	56.57	51.30	51.30	Average = 51.51 inches
56.74	54.30	50.59	53.63	50.61	50.60	
52.73	46.26	46.83	50.31	48.96	52.94	
49.28	47.76	56.40	51.32	52.20	47.73	$6 \times 4 = 24$ students.

Data sample 2:

52.33	48.55	53.13	50.20	51.12	50.19	Average = 52.56 inches
57.56	51.96	48.83	54.47	48.34	52.63	
46.12	48.02	52.59	54.22	52.51	51.65	
51.10	47.56	49.84	50.62	55.17	53.03	

$X_i$ : i<sup>th</sup> 3<sup>rd</sup> grader's height.

$$\mu_X = \mathbb{E}(X_i) \quad \forall i$$

$$\mu_n = \mu_{24} = \frac{1}{24} \sum_{i=1}^{24} X_i$$

# Classical inference

---

- unknown parameter  $\theta$  as a **deterministic** (not random!) but unknown quantity.
  - Average height.

$$X_i \quad \mu_x = \text{Average height.}$$

"constant" that we  
don't know.

- The observation  $X$  is random and its distribution
  - $p_X(x; \theta)$  if  $X$  is discrete
  - $f_X(x; \theta)$  if  $X$  is continuous
  - depends on the value of  $\theta$ .

$$X_i \sim N(\mu_x, \sigma_x^2) : \theta := (\mu_x, \sigma_x)$$

parameters used to define the distribution.

$$B \sim \text{Bernoulli}(p). \quad \theta := p$$

# Classical inference

---

$$\theta \rightarrow f_X(x; \theta) \xrightarrow{x_1, x_2, \dots, x_n} \text{Estimator} \rightarrow \hat{\theta}$$

discrete RV.  
 continuous RV.  
 complex

estimated value for  $\theta$

- Given observations  $X = (X_1, \dots, X_n)$ , an estimator  $\hat{\Theta} = g(X)$  is function of  $X$ .
- Thus,  $\hat{\Theta}$  is a random variable.
- Let  $n$  be the number of observations, the mean and variance of  $\hat{\Theta}_n$  are denoted  $E_{\theta}[\hat{\Theta}_n]$  and  $\text{var}_{\theta}[\hat{\Theta}_n]$ , respectively.

$X$ : mean , variance , [ interval ] that contains the estimate mean

## Terminology regarding estimators

output of estimator

Random variable (e.g.  $M_n$ )

- The underlying parameter  $\theta$  to be estimated is a **constant**.
- Estimation error:  $\tilde{\Theta}_n = \widehat{\Theta}_n - \theta$ .
- Bias of the estimator:  $b_\theta(\widehat{\Theta}_n) = E_\theta[\widehat{\Theta}_n] - \theta$ , is the expected value of the estimation error.

$$\mathbb{E}(\tilde{\Theta}_n)$$

e.g.  $\mathbb{E}(M_n) - M_x = 0$

## Estimation of the Mean

unknown RV.  $X$ ;  $\mu_X$  mean

- Suppose that the observations  $X_1, \dots, X_n$  are i.i.d., with an "sample Average" unknown common mean  $\mu_X$ .
- $M_n = \hat{\mu}_X = \frac{1}{n} \sum_{i=1}^n X_i$  is unbiased estimator
  - for any  $n$ , the expected value of the average is equal to the true mean.

```
heights= np.array([121.92, 132.64, 113.31, 97.20, 140.94, 139.04, 115.98, 128.27, 121.84, 97.73])
```

The average height estimate

```
average = np.mean(heights)= 122.185
```



## Properties of the Estimator of the mean

---

- $\hat{\mu}_x = \frac{1}{n} \sum_{i=1}^n X_i$  is unbiased estimator
  - for any n, the expected value of the average is equal to the true mean.

$$\begin{aligned}
 \text{bias}(\hat{\mu}_x) &= E(\hat{\mu}_x - \mu_x) \\
 &= E(\hat{\mu}_x) - \underbrace{E(\mu_x)}_{\text{constant.}} \\
 &= E(\hat{\mu}_x) - \mu_x = \mu_x - \mu_x = 0
 \end{aligned}$$

$$\hat{\mu}_x \sim N(\mu_x, \frac{\sigma_x^2}{n})$$

Average error  $\rightarrow 0$

Estimating the variance  $X$ .

$$\text{Var}(X) = \mathbb{E}((X - \mu_X)^2)$$

Let  $\sigma_X^2$  denote the variance of the random variables. Then there are two cases that should be considered for estimating the variance.

**Known mean:** If the mean of the random variables,  $\mu_X$ , is known. Let the sample-variance estimator for this case be defined by

$$\mathbb{E}(g(x)) = \frac{1}{N} \sum_{i=1}^N g(x_i)$$

$$\hat{\sigma}_X^2 = \frac{1}{N} \sum_{i=1}^N (X_i - \underbrace{\mu_X}_{\text{Constant}})^2.$$

Sample Approximation

$$\mathbb{E}((X - \mu_X)^2) \approx \frac{1}{N} \sum_{i=1}^N (x_i - \mu_X)^2$$

Sample set  $\{x_1, x_2, \dots, x_N\}$ .

## Estimating the variance

---

- Let's first determine if the sample variance estimator is biased when the true mean is known: (*experiment validate*)

$$\text{bias}(\hat{\sigma}_x^2) = \overline{E}(\hat{\sigma}_x^2 - \sigma_x^2) = 0$$

error

$$\overline{E}(\hat{\sigma}_x^2) = \overline{E}\left(\frac{1}{N} \sum_{i=1}^N (\bar{x}_i - \mu_x)^2\right)$$

proof.

$$\longrightarrow \sigma_x^2$$

## Estimating the variance

---

**Known mean:** If the mean of the random variables,  $\mu_X$ , is known. Let the sample-variance estimator for this case be defined by

$$\hat{\sigma}_X^2 = \frac{1}{N} \sum_{i=1}^N (X_i - \mu_X)^2.$$

**Unknown mean:** it is natural to replace  $\mu_X$  with its sample estimate  $\hat{\mu}_X$ :

$$\hat{\sigma}_X^2 = \frac{1}{N} \sum_{i=1}^N (X_i - \hat{\mu}_X)^2.$$

See the ADDITIONAL PDF for a proof why the variance estimator above is biased.

## Estimating the variance

---

- Let's first determine if the sample variance estimator is biased when we replace the true mean with its sample estimate: (*experiment validate*)

## Estimating the variance

---

**Known mean:** If the mean of the random variables,  $\mu_X$ , is known. Let the sample-variance estimator for this case be defined by

$$\hat{\sigma}_X^2 = \frac{1}{N} \sum_{i=1}^N (X_i - \mu_X)^2.$$

**Unknown mean:** unbiased estimator:

$$\hat{\sigma}_X^2 = \frac{1}{N-1} \sum_{i=1}^N (X_i - \hat{\mu}_X)^2.$$

The change in denominator is often referred to as a \*degrees-of-freedom (dof) correction\*.

## Example

---

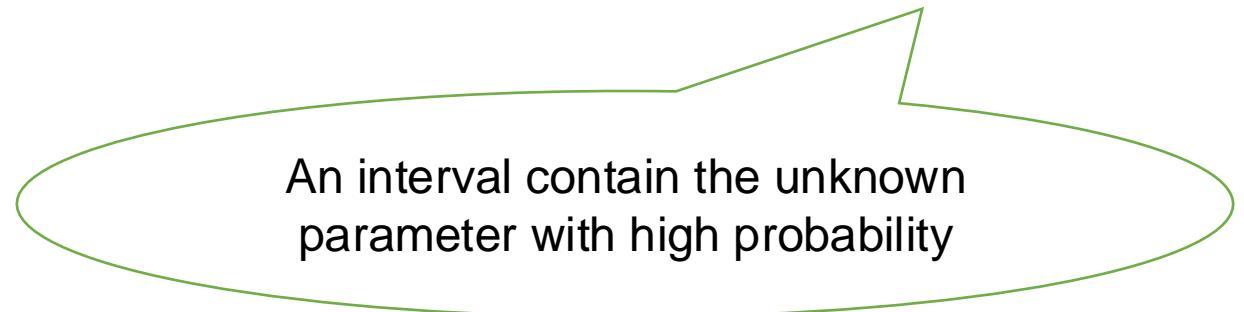
- Suppose we have a sample of student scores from an exam, and we want to estimate the population mean score.
- Sample data: `array([84, 78, 71, 74, 60, 76, 50, 86, 67, 82, 89, 93, 79, 72, 78, 76, 71, 85, 86, 52, 61, 63, 92, 71, 80, 60, 76, 81, 57, 88])`
- Total 30 samples
- Using the same sample data, we want to estimate the population variance.

## Point estimate vs interval estimate

---

- Instead of estimating a single value, **an interval estimate is also used:**
- **For an unknown parameter**

$$\theta \rightarrow f_X(x; \theta) \xrightarrow{x_1, x_2, \dots, x_n} \text{Interval Estimator} \rightarrow [\hat{\theta}^-, \hat{\theta}^+]$$



## Confidence intervals (CIs)

---

- The value of an estimator may not be informative enough
- Let us first fix a desired confidence level,  $1 - \alpha$ , where  $\alpha$  is typically a small number.
- We then replace the point estimator  $\widehat{\Theta}_n$  by a lower estimator  $\widehat{\Theta}_n^-$  and an upper estimator  $\widehat{\Theta}_n^+$ , s.t.

$$P(\widehat{\Theta}_n^- \leq \theta \leq \widehat{\Theta}_n^+) \geq 1 - \alpha$$

for every possible value of  $\theta$ .

- We call  $[\widehat{\Theta}_n^-, \widehat{\Theta}_n^+]$  a  $(1 - \alpha)$  confidence interval.

## Confidence intervals (CIs)

---

- $\hat{\mu}_X = \frac{1}{n} \sum_{i=1}^n X_i$  estimated mean. , point estimator.
- Recall: the observations  $X_1, \dots, X_n$  are i.i.d., with an unknown common mean  $\mu_X$

$$\hat{\mu}_X \sim \mathcal{N}\left(\mu_X, \frac{\sigma^2}{n}\right)$$

- Recall CLT:

We call  $[\hat{\mu}_X^-, \hat{\mu}_X^+]$  a  $(1 - \alpha)$  confidence interval if
   
*lower bound*
*upper bound.*

$$P(\hat{\mu}_X^- \leq \mu_X \leq \hat{\mu}_X^+) > 1 - \alpha$$

$\alpha = 0.05$   
 $(1 - \alpha) = 95\%$

## Confidence intervals (CIs)

$$? P(\hat{u}_x > u_x + d) \leq \frac{\alpha}{2} = 0.025$$

Standardization:

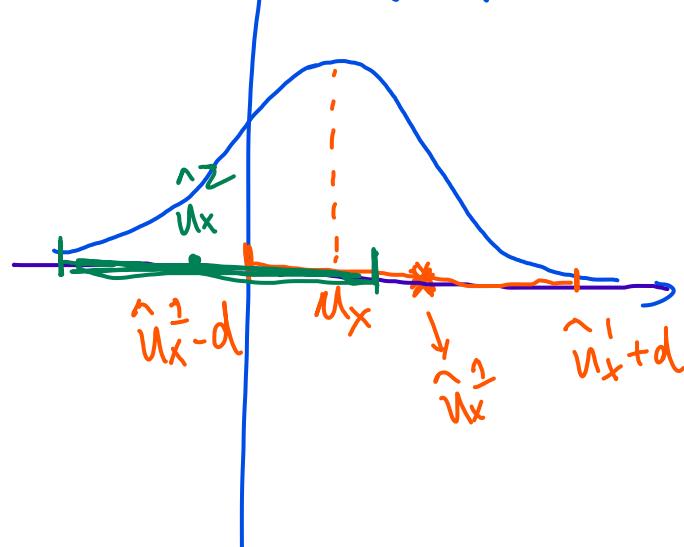
$$P\left(\frac{\hat{u}_x - u_x}{\sqrt{\sigma^2/N}} > \frac{d}{\sqrt{\sigma^2/N}}\right) \leq \frac{\alpha}{2}$$

- Suppose  $\alpha = 0.05$

- Let's compute the 95% confidence interval about the mean of unknown RV using the samples.

point estimator

pdf of  $\hat{u}_x$



$$P(\hat{u}_x > u_x + d) \leq \frac{\alpha}{2}$$

<https://www.mathsisfun.com/data/standard-normal-distribution-table.html>

$$\frac{\hat{u}_x - u_x}{\sqrt{\sigma^2/N}} \sim N(0, 1) \quad P(G > ?) \leq 0.025$$

$$\hat{u}_x = \frac{1}{N} \sum_{i=1}^N x_i \rightarrow \text{interval estimator } [\hat{u}_x - d, \hat{u}_x + d]$$

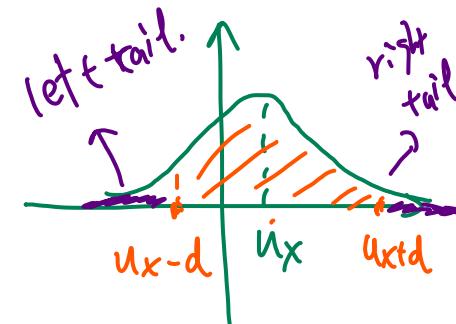
$$\hat{u}_x \sim N(u_x, \frac{\sigma^2}{N})$$

Take a sample from  $\hat{u}_x$ ,  $\hat{u}_x^1$

$$P(\hat{u}_x - d \leq u_x \leq \hat{u}_x + d) \geq 1 - \alpha$$

$$\Leftrightarrow P(\hat{u}_x \leq u_x + d \text{ and } \hat{u}_x \geq u_x - d) \geq 1 - \alpha$$

$$\Leftrightarrow P(u_x - d \leq \hat{u}_x \leq u_x + d) \geq 1 - \alpha$$



# Confidence interval for mean estimate with unknown variance

---

- Recall if the variance is unknown, we have an unbiased estimate for the variance

**Unknown mean:** unbiased estimator:

$$\hat{\sigma}_X^2 = \frac{1}{N-1} \sum_{i=1}^N (X_i - \hat{\mu}_X)^2.$$

$$\frac{\hat{\mu}_X - \mu_X}{\sigma_X / \sqrt{n}}$$

has a Student's  $t$ -distribution with  $\nu = n - 1$  degrees of freedom (dof)  $t_\nu$ .  
(Like a Gaussian, but more spread out!)

# Summary

---

1. Point Estimation for Mean with prior knowledge of the population variance

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n x_i$$

2. Standard Error of the Mean with known population variance (not an estimate but can be computed based on the property of variance.)

$$Var = \frac{\sigma^2}{n}$$

$$SE = \frac{\sigma}{\sqrt{n}}$$

2-2. known mean . estimate variance

$$\hat{\sigma}^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{M}_x)^2$$

3. Point Estimation for Mean without prior knowledge of the population variance

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n x_i$$

4. Point Estimation for Variance without prior knowledge of the population variance

mean

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

↳ estimated mean

## summary

---

- Formula to compute confidence interval

$$1 - \alpha \quad \alpha = 0.02$$

For a population mean  $\mu$  when the sample size is large ( $n \geq 30$ ), the confidence interval is given by:

$$\left( \hat{\mu}_X - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \quad \hat{\mu}_X + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right) \quad (1)$$

where:

- $\hat{\mu}_X$  = sample mean
- $\sigma$  = population standard deviation (or sample standard deviation if unknown)
- $n$  = sample size
- $z_{\alpha/2}$  = critical value from the standard normal table for a given confidence level

$$P(G > z_{\alpha/2}) \leq \frac{\alpha}{2}$$

$\downarrow$   
standard Gaussian

$$z_{\alpha/2} = z_{0.01} = G_{.1sf}(0,1)$$