Lecture 4

$$\begin{cases} X_1, X_2, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2) \\ \bar{X} = \dfrac{X_1 + \cdots + X_n}{n} \sim \underline{N}\left(\mu, \dfrac{\sigma^2}{n}\right) \end{cases}$$

$$\Pr\{\mu \in [\bar{X} - \beta, \bar{X} + \beta]\} = \underline{1-\alpha}$$

$$\beta = \frac{\sigma}{\sqrt{n}} \cdot z_{\alpha/2}$$

General case: $X_1, \ldots, X_n \overset{iid}{\sim} \text{dist}(\mu, \sigma^2)$

$$\bar{X} \overset{CLT}{\sim} \underline{N}\left(\mu, \sigma^2/n\right) + \left(\begin{array}{c}\text{negligible} \\ \text{erro}\end{array}\right)$$

All the derivation from the Gaussian case carries over, and we obtain:

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

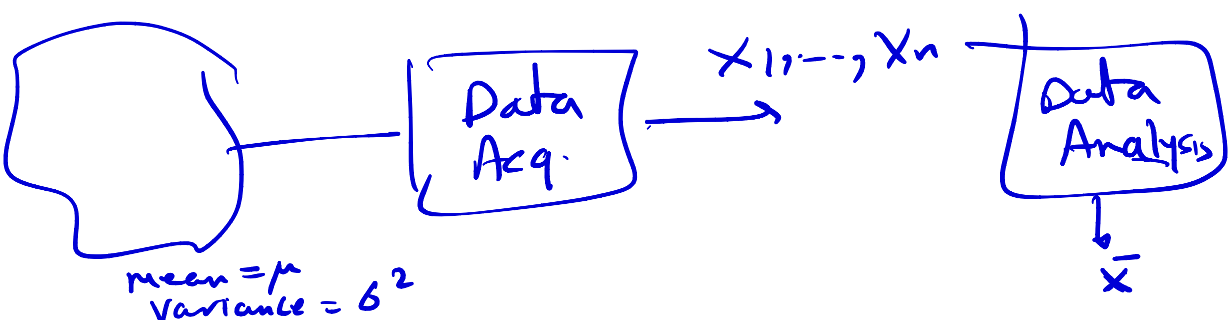
$$= 1-\alpha + \cancel{\frac{\text{Constant}}{n}} \rightarrow \text{neglect}$$

Definition:

A $100(1-\alpha)\%$ Confidence interval for the mean-estimation problem, when $X_i \sim \text{dist}(\mu, \sigma^2)$, is given as

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

where $\bar{X} \triangleq \frac{X_1 + \cdots + X_n}{n}$ is the Sample mean.



width of the confidence interval

$$\frac{2\sigma}{\sqrt{n}} z_{\alpha/2}$$

Question:

Assume that we'd like the width of the $100\cdot(1-\alpha)\%$ confidence interval to at most a given value w (e.g. $w=0.01$).

How many samples do we need?

$$\Pr\{\mu \in [\bar{X} - \underbrace{w/2}_{0.01}, \bar{X} + \underbrace{w/2}_{0.01}]\} \underset{\geq}{=} \underbrace{1-\alpha}_{0.95}$$

$$n \geq ?$$

$$\longrightarrow \quad w \overset{\leq}{=} \frac{2\sigma}{\sqrt{n}} z_{\alpha/2} \implies n \overset{\geq}{=} \left(\frac{2\sigma z_{\alpha/2}}{w}\right)^2$$

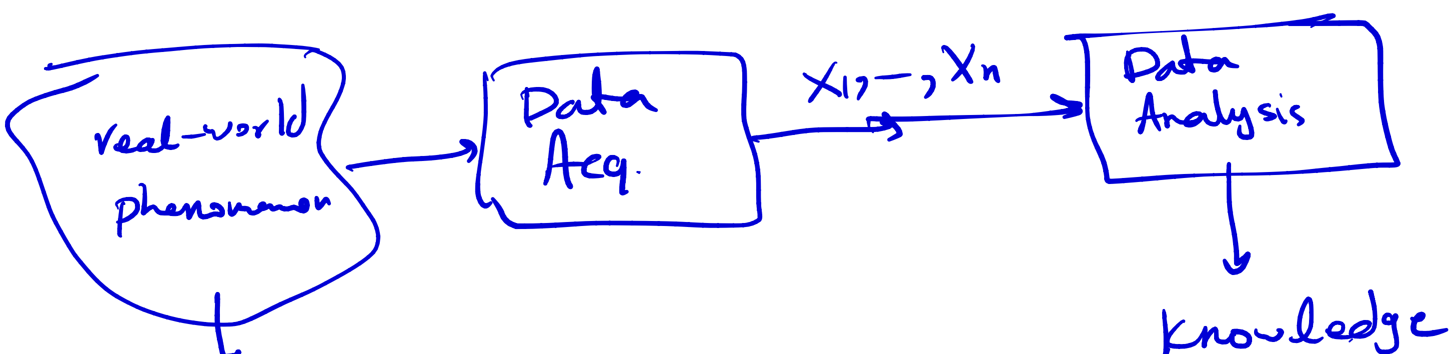
There is still a PROBLEM:

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

Problem → We don't have the value of σ

Solution → Estimate it from data.

Point estimation:



θ
(unknown)

Our task:

Design a statistic $\hat{\theta}(X_1, \dots, X_n)$ that estimates θ.

Example:

θ = population mean in the previous application that we've talked about

$$\underbrace{\theta = E[X_i]}_{\triangleq \mu} \longrightarrow \hat{\theta}(X_1, \dots, X_n) = \frac{X_1 + \dots + X_n}{n}$$

Another example:

θ : Variance of the population

$$E[(X_i - \mu)^2] \triangleq \sigma^2$$

$$\mu = E[X_i]$$

Let's now come up with an estimator for the case where $\theta = \sigma^2$.

$$\sigma^2 = E[(X - \mu)^2]$$

$$\bar{X}$$

$$\hat{\theta}(\underbrace{X_1, \dots, X_n}) \Rightarrow \text{close to } \sigma^2$$

estimated from data

$$\sigma^2 = E\big[(X - \underset{\sim}{\bar{X}})^2\big]$$

$X_1, \dots, X_n \overset{iid}{\sim} X$

$$\hat{\sigma}^2 = \frac{(X_1-\bar{X})^2 + \cdots + (X_n-\bar{X})^2}{n}$$

Is $\hat{\sigma}^2$ a good estimator?

- If $n \to \infty$, do we have $\hat{\sigma}^2 \xrightarrow[n\to\infty]{} \sigma^2$?

Yes: when $n \to \infty$ $\longrightarrow$ $\bar{X} \to \mu$

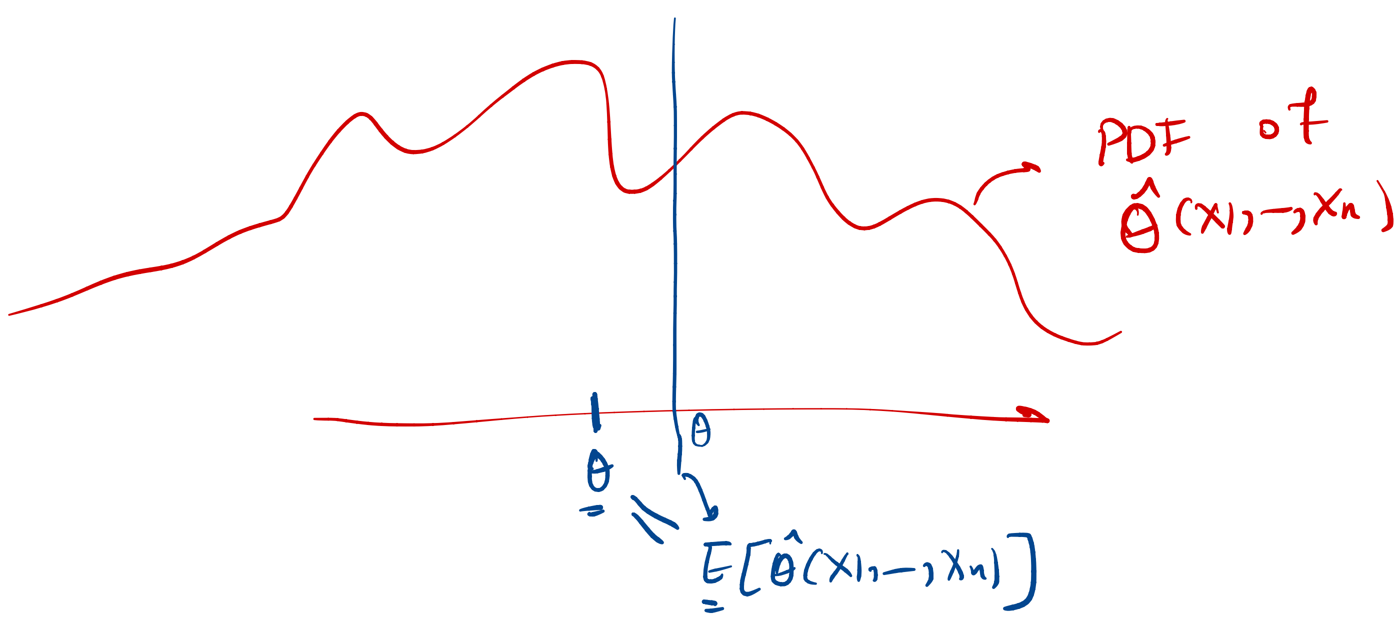
So when $n \to \infty$: $\underbrace{\dfrac{(X_1-\mu)^2 + \cdots + (X_n-\mu)^2}{n}}$

$$\xrightarrow[n\to\infty]{\text{Law of Large numbers}} E\big[(X-\mu)^2\big] = \sigma^2$$

Definition: An estimator $\hat{\theta}(X_1, \ldots, X_n)$ of a parameter θ is called unbiased if

$$E[\hat{\theta}(X_1, \ldots, X_n)] = \theta \quad (\text{for all } \theta \text{ and all } n)$$

The value $E[\hat{\theta}(X_1, \ldots, X_n)] - \theta$ is called the "bias" of the estimation.



Question : is $\hat{\sigma}^2$ an unbiased estimator?

$$\hat{\sigma}^2 = \frac{(X_1 - \bar{X})^2 + \cdots + (X_n - \bar{X})^2}{n}$$

$$\sigma^2 = E[(X-\mu)^2]$$

↳ assumed that μ is known and consider the following estimator for σ^2:

$$\hat{\sigma}^2_{simple} = \frac{(X_1 - \mu)^2 + \cdots + (X_n - \mu)^2}{n}$$

↳ $\hat{\sigma}^2_{simple}$ is an unbiased estimator (just compute the E)

but our original estimator was:

$$\hat{\sigma}^2 = \frac{(X_1 - \bar{X})^2 + \cdots + (X_n - \bar{X})^2}{n}$$

because we're also estimating μ (via $\bar{X}$) the overall estimator becomes biased.

The "right" way to answer the question of unbiasedness of $\hat{\sigma}^2$ is to "compute" the expectation:

$$E\left[\hat{\sigma}^2(X_1, \ldots, X_n)\right] \overset{?}{=} \sigma^2$$

$$E\left[\hat{\sigma}^2(X_1,\dots,X_n)\right]$$

$$= E\left[\frac{(X_1-\bar{X})^2+\dots+(X_n-\bar{X})^2}{n}\right]$$

$$= E\left[\frac{X_1^2+\bar{X}^2-2X_1\bar{X}+\dots+X_n^2+\bar{X}^2-2X_n\bar{X}}{n}\right]$$

$$= E\left[\frac{\sum_{i=1}^{n} X_i^2}{n}\right] + \underbrace{E\left[\frac{n\bar{X}^2}{n}\right]-2E\left[\bar{X}\overbrace{\left(\frac{X_1+\dots+X_n}{n}\right)}^{\bar{X}}\right]}_{E[\bar{X}^2]-2E[\bar{X}^2]}$$

$$= \underbrace{E\left[\frac{\sum X_i^2}{n}\right]}_{\substack{X_i\text{'s are iid} \\ E[X_1^2]=\dots=E[X_n^2]}} - E[\bar{X}^2]$$

$$= \underbrace{E[X_1^2]}_{} - \underbrace{E[\bar{X}^2]}_{}$$

$$= (\sigma^2 + \mu^2) - \left(\sigma^2/n + \mu^2\right)$$

$$= \sigma^2(1 - 1/n) \neq \sigma^2$$

$$\Downarrow$$

$\hat{\sigma}^2$ is biased!

recall:

$E[X_1] = \mu$

$E[(X_1-\mu)^2] = \sigma^2$

$$E[(X-\mu)^2]$$

$$= E[X^2] - \mu^2$$

$$\Rightarrow E[X^2] = \mu^2 + \sigma^2$$

$$Var(Y)$$

$$= E[Y^2] - (E[Y])^2 \longrightarrow$$

let $Y := \bar{X}$

$$\underbrace{Var(\bar{X})}_{\sigma^2/n} = E[\bar{X}^2] - (E[\bar{X}])^2$$

$$\sigma^2/n = E[\bar{X}^2] - \mu^2$$

As we computed $E[\hat{\sigma}^2] = \sigma^2(1-1/n)$.

An easy way to make the estimator $\hat{\sigma}^2$ unbiased is to rescale it by $(1-1/n)$:

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

Hence the following estimator is an unbiased estimator for σ^2:

$$\frac{\hat{\sigma}^2}{(1-1/n)} = \frac{1}{(1-1/n)} \cdot \frac{(x_1-\bar{x})^2 + \cdots + (x_n-\bar{x})^2}{n}$$

$$:= \frac{(x_1-\bar{x})^2 + \cdots + (x_n-\bar{x})^2}{n-1}$$

An important principle that we learned from the above example is:

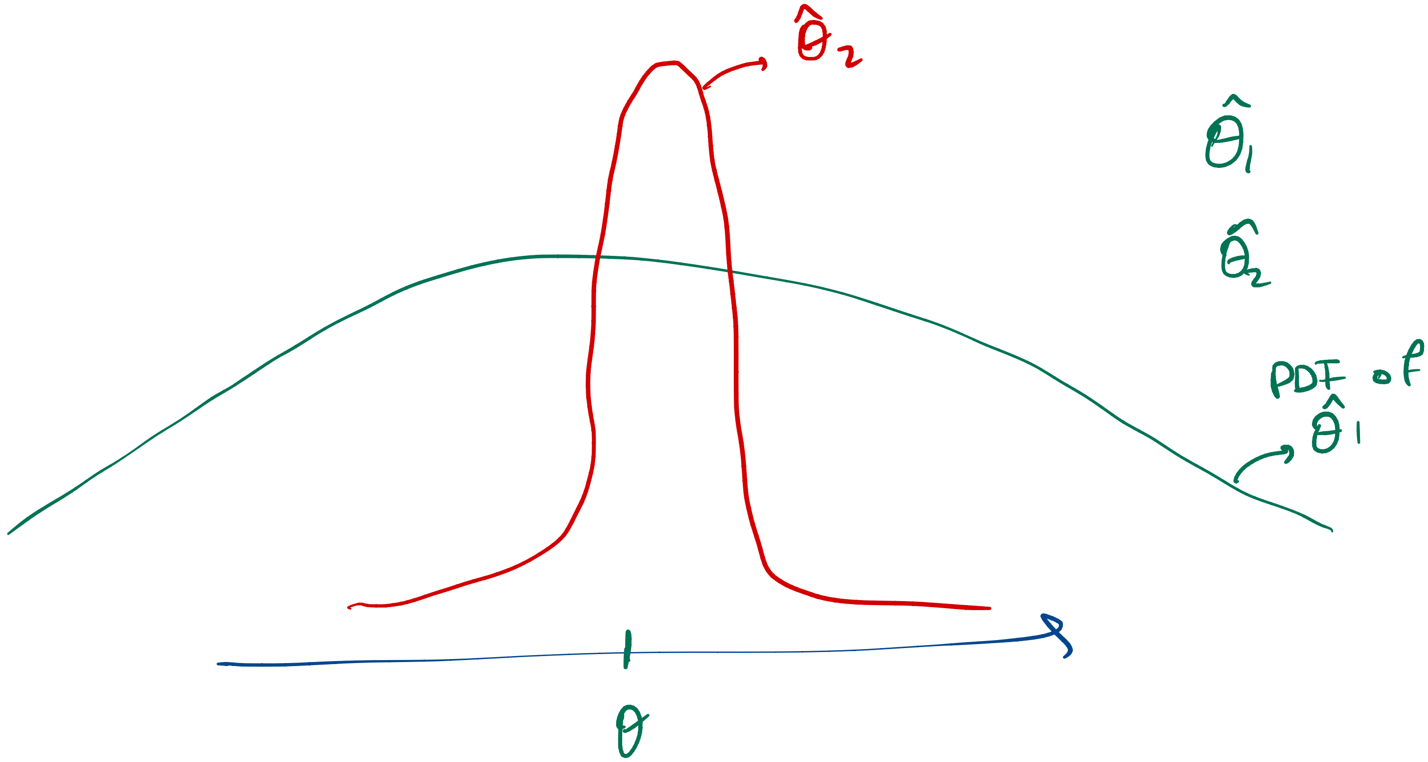
The Principle of Unbiased Estimation:

Among all the estimators of a parameter θ, we usually prefer the ones that are unbiased.

Another important principle is:

The principle of Minimum Variance Unbiase Estimation:
(MVUE):

Among all the possible unbiased estimators for a parameter θ, we prefer the one with minimum variance. The resulting estimator is called the MVUE estimator.



$$\underbrace{E\left[(X-\mu)^2\right]}_{\sigma^2} \xleftarrow{n\to\infty} \frac{(X_1-\mu)^2+\cdots+(X_n-\mu)^2}{n}$$

$$= \sigma^2$$

$$\boxed{\bar{X} = \frac{X_1+\cdots+X_n}{n} = \mu}$$

X_1