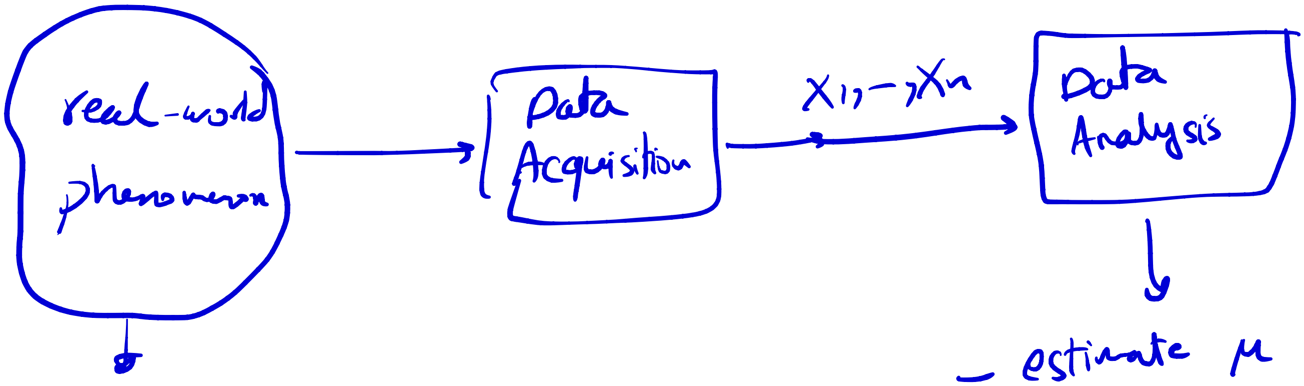
Lecture 5:



$$\begin{cases} \text{population mean: } \mu \\ \text{population variance: } \sigma^2 \end{cases}$$

– estimate μ by

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

Confidence interval:

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

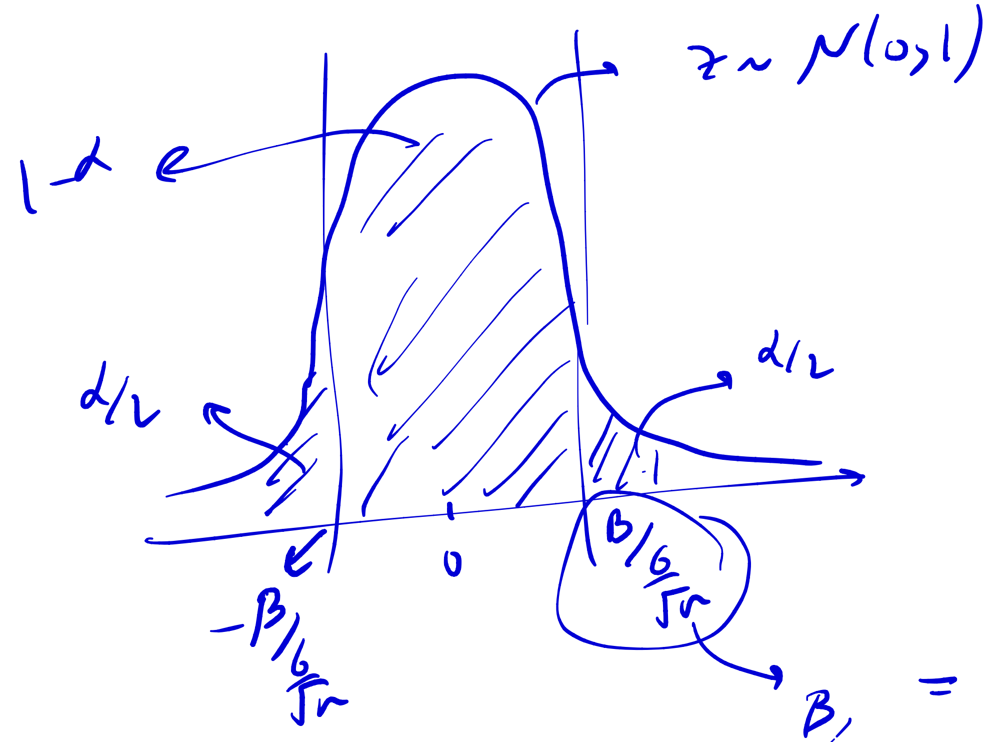
$$\|$$

$$\Pr\{\mu - \beta \le \bar{X} \le \mu + \beta\} \approx 1 - \alpha$$

$$\|$$

CLT → $Z \sim N(0,1)$

$$\Pr\left\{ -\frac{\beta}{(\sigma/\sqrt{n})} \le \frac{\bar{X} - \mu}{(\sigma/\sqrt{n})} \le \frac{\beta}{(\sigma/\sqrt{n})} \right\} \approx 1 - \alpha + \frac{c}{n}$$



$$\frac{B}{\left(\frac{\sigma}{\sqrt{n}}\right)} = z_{\alpha/2}$$

$100(1-\alpha)\%$ Confidence interval,

$$\Pr\left\{\mu \in \left[\bar{X} - \frac{\sigma}{\sqrt{n}} z_{\alpha/2},\ \bar{X} + \frac{\sigma}{\sqrt{n}} z_{\alpha/2}\right]\right\} = 1-\alpha$$

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

↓

An estimator for σ would be: $\sqrt{\hat{\sigma}^2} = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2}$

In summary, we estimate σ from data (and hence there is an estimation error) and plug the resulting estimate into the confidence interval.

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

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

There is an error $|\hat{\sigma} - \sigma|$ which needs to be taken into account.

In the following few pages we will argue that the error $|\hat{\sigma} - \sigma|$ is so small compared to the other values (involve in the confidence interval calculation) so that we can neglect it.

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

$$\|$$

$$\left[\bar{X} - \frac{\hat{\sigma}}{\sqrt{n}} z_{\alpha/2} - \underbrace{\frac{\sigma-\hat{\sigma}}{\sqrt{n}} z_{\alpha/2}}_{\text{error} = \frac{\text{constant}}{n}}\,,\ \bar{X} + \frac{\hat{\sigma}}{\sqrt{n}} + \underbrace{\frac{\sigma-\hat{\sigma}}{\sqrt{n}} z_{\alpha/2}}_{\text{error} = \frac{\text{constant}}{n}}\right]$$

We will show that $|\sigma - \hat{\sigma}|$ is at most $\frac{c}{\sqrt{n}}$ and hence $\frac{|\sigma-\hat{\sigma}|}{\sqrt{n}} = \frac{c}{n}$.

$$\Downarrow$$

width of the interval:

$$\frac{2\sigma}{\sqrt{n}} z_{\alpha/2} + \overbrace{\text{error caused by replacing } \sigma \text{ with } \hat{\sigma}}^{\text{this error is } \frac{\text{constant}}{n}}$$

Since $\frac{\text{constant}}{n} \underset{n \text{ large}}{\ll} \frac{2\sigma}{\sqrt{n}} z_{\alpha/2}$, we can neglect the error resulted by replacing σ with $\hat{\sigma}$.

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

$$E[\hat{\sigma}^2] = \frac{n-1}{n} \sigma^2 \quad (\Rightarrow \text{biased})$$

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

The following derivations are optional and won't be asked in the exams/homeworks

note that in the formula for confidence interval, we are using

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

We will now show:

$$|\hat{\sigma}_{new} - \sigma| = \frac{c}{\sqrt{n}} \Longrightarrow \frac{|\hat{\sigma} - \sigma|}{\sqrt{n}} = \frac{c}{n}$$

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

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

$\hookrightarrow$ unbiased $E[\hat{\sigma}^2_{new}] = \sigma^2$

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

$\neq$ unbiased for $\hat{\sigma}$

if $E[X] = \alpha$

$\cancel{\Downarrow}$

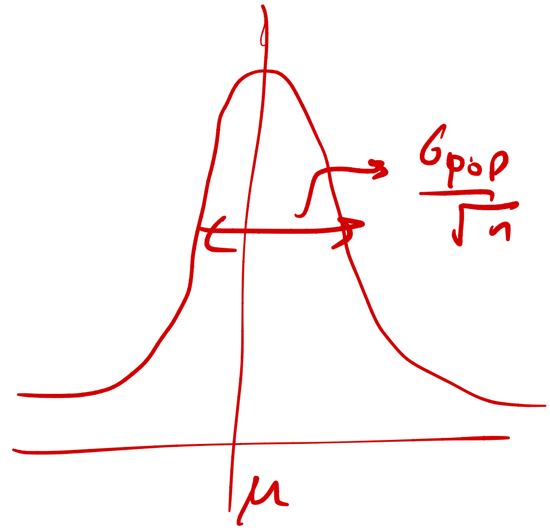
$E[\sqrt{X}] = \sqrt{\alpha}$

$$\hat{\sigma}_{new} = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2} = \underbrace{\sqrt{\frac{1}{n-1} \sum (x_i - \mu)^2 + \frac{c'}{\sqrt{n}}}}_{①}$$

$$|\bar{x} - \mu| = \frac{c}{\sqrt{n}}$$

$$\bar{X} \overset{CLT}{=} \mu + \frac{\sigma_{pop}}{\sqrt{n}} Z$$

$$\bar{X} - \mu \approx \frac{c}{\sqrt{n}}$$



$$\underbrace{\left| \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)^2}\; - \sigma^2 \right| \overset{CLT}{=} \frac{c''}{\sqrt{n}} \quad ②$$

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

Let : $y_i = (x_i - \mu)^2$

then $\left| \frac{1}{n} \sum_{i=1}^{n} y_i - E[y_i] \right| \overset{CLT}{\approx} \frac{c''}{\sqrt{n}}$

---

① : $\underline{\hat{\sigma}_{new}} = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \mu)^2} + \frac{c'}{\sqrt{n}}$

$$\overset{②}{=} \sqrt{\sigma^2 + \frac{c''}{\sqrt{n}}} + c'/\sqrt{n} = \sigma + \frac{c}{\sqrt{n}}$$

$$\frac{1}{n-1} = \frac{1}{n} + \frac{c'''}{n}$$

$$\sigma_{new} = \sqrt{\underbrace{\frac{1}{n-1} \sum (x_i-\mu)^2 + \frac{c'}{\sqrt{n}}}} = \sqrt{\left(\frac{n}{n-1}\right) \underbrace{\boxed{\frac{1}{n} \sum_{i=1}^{n} (x_i-\mu)^2}}_{\sigma^2 + \frac{c''}{\sqrt{n}}}}$$

$$\textcircled{2}$$

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

$$= 1 - 1/n - \underbrace{\frac{1}{n(n-1)}}_{\frac{1}{n^2}}$$

$$= \sqrt{\left(\frac{n}{n-1}\right)\left(\sigma^2 + \frac{c''}{\sqrt{n}}\right)}$$

$$= \sqrt{\sigma^2 + \frac{c''}{\sqrt{n}} + \frac{c'''}{n}}$$

$$= \sqrt{\sigma^2 + \frac{D}{\sqrt{n}}}$$

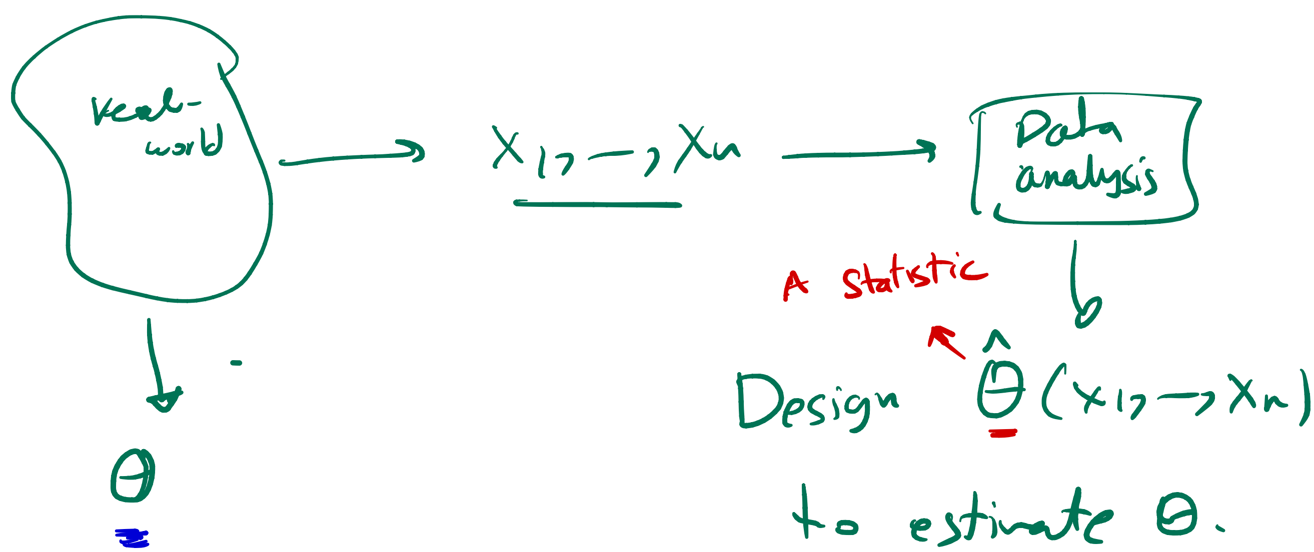
$$= \sigma + \frac{D'}{\sqrt{n}}$$

$$\Rightarrow \boxed{\hat{\sigma}_{new} = \sigma + D'/\sqrt{n}}$$

$$\sigma_{mo} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2} \longrightarrow \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

$$\boxed{\begin{array}{ll} \sqrt{x+\delta} & \text{where } \delta \text{ is small} \\ & \delta \ll x \\ \approx \sqrt{x} + c\cdot\delta + o(\delta^2) & \end{array}}$$

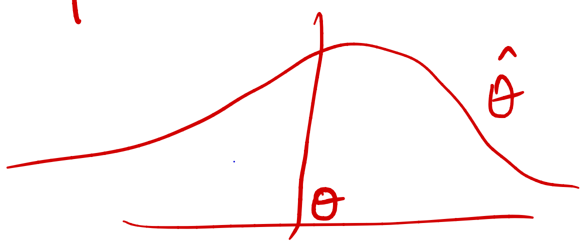
Taylor



<u>Point estimation:</u> A point estimate of a parameter θ is a single number

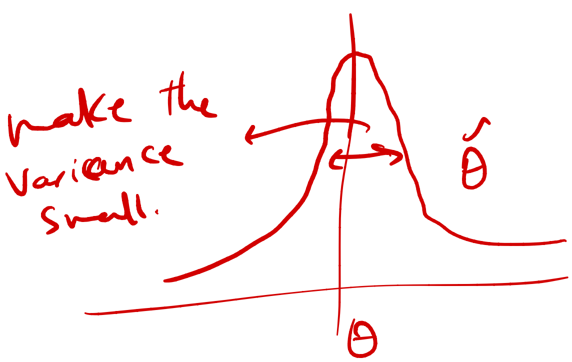
obtained by selecting/designing a suitable statistic and computing its value from the sample data. The selected statistic is called the point estimator for θ.

① We prefer $\hat{\theta}$ to be unbiased.



$E[\hat{\theta}(x_1, \ldots, x_n)] = \theta$

② We prefer $\hat{\theta}$ (if unbiased) to have the minimum possible variance.



Systematic approaches for parameter estimation:

We will introduce two general approaches:

(1) The Method of moments

(2) The method of maximum-likelihood.

Basic assumption: Assume that we know that the data is generated according to a pdf whose parameters are unknown.

$$X_1, \dots, X_n \overset{iid}{\sim} \underbrace{f(x \mid \overbrace{\theta_1, \theta_2, \dots, \theta_m}^{\text{unknown}})}_{\substack{\text{PDF of the} \\ \text{distribution that} \\ \text{generates the dat}}}$$

known

essentially what we're doing is to assume a "parametric" model that generates the data.

Example: $X_1, \dots, X_n \overset{iid}{\sim} N(\underset{\text{Unknown}}{\mu}, \underset{\text{Unknown}}{\sigma^2})$

PDF: $\dfrac{1}{\sqrt{2\pi}\,\sigma}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

$= \underset{\text{known}}{f}(x \mid \underbrace{\overset{\theta_1}{\mu}, \overset{\theta_2}{\sigma}}_{\text{parameters are unknown}})$

Example: $X_1, \dots, X_n \overset{iid}{\sim} f(x \mid \theta)$

$= \underbrace{\theta\, e^{-\lambda\theta}\, \mathbb{1}\{x \geq 0\}}_{\text{exponential distribution}}$

$X_1, \dots, X_n \sim f(x \mid \theta_1, \dots, \theta_m)$

Goal: Design: $\hat{\theta}_1(X_1, \dots, X_n) \xrightarrow{\text{estimator for}} \theta_1$

$\vdots$

$\hat{\theta}_m(X_1, \dots, X_n) \xrightarrow{\text{estimator for}} \theta_m$

The method of moments:

Definition: The k-th moment of a random variable X is defined as:

$$\mu_k = E[X^k]$$

e.g. the first moment is the mean:

$$\mu_1 = E[X]$$

Question: Given an iid sample $X_1, \ldots, X_n$, from a distribution with k-th moment μ_k, design an unbiased estimator for μ_k (from the sample):

$$X_1, \ldots, X_n \longrightarrow \hat{\mu}_k$$

$\hat{\mu}_1 \longrightarrow$ estimator for the mean $\mu_1 = E[X]$

$$\frac{X_1 + \cdots + X_n}{n}$$

$$\hat{\mu}_k \longrightarrow \mu_k = \underbrace{E[X^k]}$$

unbiased estimator $\hat{\mu}_k = \dfrac{X_1^k + X_2^k + \cdots + X_n^k}{n}$

$$E[\hat{\mu}_k] = E\left[\frac{X_1^k + \cdots + X_n^k}{n}\right]$$

$$\overset{iid}{=} E[X_1^k] = \mu_k .$$