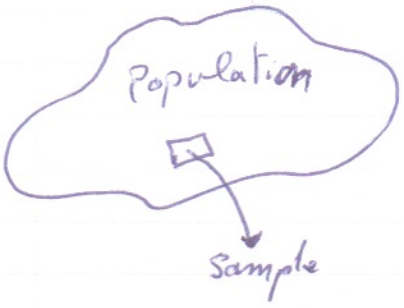
# MATH 357.

## Hon. Statistics

# Math 357: Hon. Statistics

## Introduction:

The main goal of statistics:
make inference (an educated guess) about a population based on a sample drawn from that population.




Examples:

- Say something about how long people with Alzheimer disease live from onset of their disease, based on a sample of people with Alzheimer.
- Say something about the proportion of canadians that are in favor of capital punishment, based on a sample of 1000 canadians.
- Say something about the reliability of certain aircraft components based on tests carried out on a sample of 100 such components.

We need to be more precise by what we mean by a population

First, we shall mean a population of real numbers (or vectors)

Second, when we talk of describing a population, we shall mean describing the probability distribution $F$ of that population (of these possible values — real nb / vectors)

There are 2 main branches of statistics:

(1) → Parametric inference

(2) → Non-parametric / distribution free inference

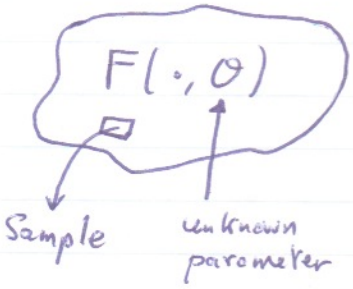
→ Under (1), we begin by assuming that F belongs to some so-called parametric family of distributions. (e.g. F comes from a normal dist or Exp dist, or Poisson)

What is unknown is one or more <u>parameters</u> that tie down the specific distribution in the family (e.g. $F \sim N(\mu, 3)$ where $\mu$ is unknown → inference about $\mu$

or $F \sim N(\mu, \sigma^2)$ where the vector $(\mu, \sigma^2)$ is unknown → inference about $\Theta := (\mu, \sigma^2)$

or $F \sim$ Bernoulli$(p)$ where $p$ is unknown → inference about $p$)

→ under (2), we do <u>not</u> start out by making any assumption about the distribution F, ie it is <u>not</u> assumed to belong to any particular family.




Our inference will be entirely parametric.

# Order Statistics

## I Definition & Main theorem:

**Order Statistics** = let $X_1, \dots, X_n$ be IID r.v. drawn from some distribution $F$:

- the order statistics of $(X_k)_{k=1}^n$ is the sample $Y_1 \le Y_2 \le \dots \le Y_n$,
- $Y_i$ = $i^{th}$ order statistic $\rightsquigarrow \forall \omega \in \Omega : Y_i(\omega) = \min(X_1(\omega), \dots, X_n(\omega))$

$$\begin{cases} Y_1 = \min(X_1, \dots, X_n) \\ Y_2 = 2^{nd}\min(X_1, \dots, X_n) \\ \vdots \\ Y_n = \max(X_1, \dots, X_n) \end{cases}$$

**Note** = If the $X_i$'s are (absolutely) continuous with PDF $f_X$, then with probability 1: $Y_1 < Y_2 < \dots < Y_n$

($\Leftrightarrow$ $Y_i = Y_{i+1} \Leftrightarrow X_i = X_{i+1}$ → happens with prob 0 ($\Leftrightarrow$ $X_i$'s are IID r.v.)

**Main theorem:** $X_1, \dots, X_n$ continuous IID r.v. with common PDF $f_X$ (i.e. same $\forall X_i$ c/ IID r.v.):

Then: $$f_{Y_1,\dots,Y_n}(y_1, \dots, y_n) = \begin{cases} n! \prod_{i=1}^n f_X(y_i), & \text{for } y_1 < y_2 < \dots < y_n \\ 0, & \text{elsewhere} \end{cases}$$

**Note** = ⚠ the $Y_i$'s are clearly NOT IID.

## II Heuristic derivation: Multinomial distribution setup

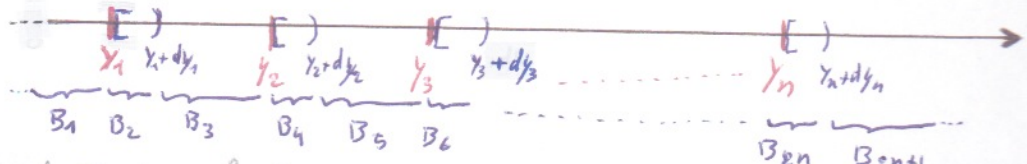
**Multinomial distribution**: • $n$ independant trials, each of which can result in one of $k$ possible outcomes.

• $P_i$ = prob (outcome $i$ at a trial) → constant for every trial

• $X_i$ = # of trials resulting in outcome $i$

$$\Rightarrow \begin{cases} P(X_1 = n_1, \dots, X_k = n_k) = \dfrac{n!}{n_1! \dots n_k!} P_1^{n_1} \dots P_k^{n_k} \\ \sum_{i=1}^k n_i = n \quad \& \quad \sum_{i=1}^k P_i = 1 \end{cases}$$

**Heuristic Proof** =




Divide the horizontal axis into $2n+1$ bins

**Note**: each $X_i$ can only fall in one of these bins

- $X_i$'s independant → outcomes of $X_i$ are indep.
- $X_i$'s identically dist → prob of falling in bin $j$ is constant $\forall j$

$\Rightarrow$ Multinomial Setup!

Joint PDF:

$$f(y_1, \dots, y_n) = \lim_{dy_1, \dots, dy_n \to 0} \left[ \frac{P(Y_1 \in [y_1, y_1+dy_1), \dots, Y_n \in [y_n, y_n+dy_n))}{dy_1 \cdot dy_2 \cdot \dots \cdot dy_n} \right]$$

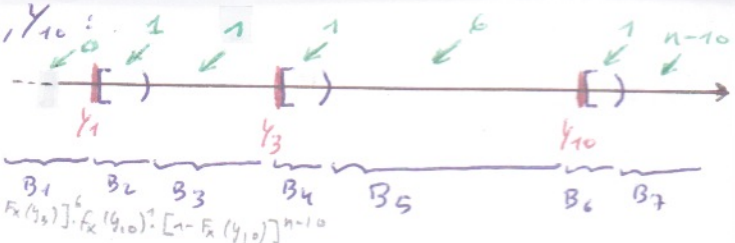
$$= \lim_{dy_1, \dots, dy_n \to 0} \left[ \frac{P(0\, X_i \text{ in } B_1; 1\, X_i \text{ in } B_2; 0\, X_i \text{ in } B_3; \dots; 1\, X_i \text{ in } B_{2n}; 0\, X_i \text{ in } B_{2n+1})}{dy_1\, dy_2 \dots dy_n} \right]$$

$$= \lim_{dy_1, \dots, dy_n \to 0} \left[ \frac{n!}{0!\,1!\,0!\,1! \dots 1!\,0!} \cdot \frac{F_X(y_1)^0 \cdot (f_X(y_1)dy_1)^1 \cdot (F_X(y_2) - F_X(y_1))^0 \cdot \dots \cdot (f_X(y_n)dy_n)^1 (1 - F_X(y_n))^0}{dy_1\, dy_2 \dots dy_n} \right]$$

$$= n! \prod_{i=1}^n f_X(y_i)$$

$$\Rightarrow f(y_1, \dots, y_n) = \begin{cases} n! \prod_{i=1}^n f_X(y_i), & \text{for } y_1 < y_2 < \dots < y_n \\ 0, & \text{otherwise} \end{cases}$$

Ex: Given $X_1, \dots, X_n$, find the joint PDF of $Y_1, Y_3, Y_{10}$:




$$f(y_1,y_3,y_{10}) = \frac{P(Y_1 \le y_1, Y_3 \le y_3, Y_{10} \le y_{10})}{dy_1\, dy_3\, dy_{10}}$$

$$= \frac{n!}{0!\,1!\,1!\,1!\,6!\,1!\,(n-10)!} F_x(y_1)^0 \cdot f_x(y_1)^1 \cdot [F_x(y_3)-F_x(y_1)]^1 \cdot f_x(y_3)^1 \cdot [F_x(y_{10})-F_x(y_3)]^6 \cdot f_x(y_{10})^1 \cdot [1-F_x(y_{10})]^{n-10}$$

$$\Rightarrow f(y_1,y_3,y_{10}) = \begin{cases} \frac{n!}{6!\,(n-10)!} f_x(y_1) f_x(y_3) f_x(y_{10}) \cdot [F_x(y_3)-F_x(y_1)] \cdot [F_x(y_{10})-F_x(y_3)]^6 \cdot [1-F_x(y_{10})]^{n-10}, & \text{if } y_1 < y_3 < y_{10} \\ 0, & \text{otherwise} \end{cases}$$

## (III) Rigorous derivation: n-dim transformation from the $X_i$'s to $Y_i$'s

**Idea** = Let $\tilde{X} = (X_1, \dots, X_n)$, $\tilde{Y}_n = (Y_1, \dots, Y_n)$:

make a n-dim transfo $\tilde{g}: \tilde{X} \longrightarrow \tilde{Y}$, $\tilde{g}(\tilde{x}) = \tilde{y}$ where $g_i(\tilde{x}) = y_i$

**Case 1**: If $\tilde{g}$ is 1-1: $\boxed{f_{\tilde{Y}}(\tilde{y}) = f_{\tilde{X}}\left(g_1^{-1}(\tilde{y}), \dots, g_n^{-1}(\tilde{y})\right) \cdot |J|}$, $J = \det\left(\frac{\partial x_i}{\partial y_j}\right)$

**Case 2**: If $\tilde{g}$ is not 1-1 on the support R of $f_{\tilde{X}}$, but $\tilde{g}$ is 1-1 on subregions $R_\ell$ of R

$\tilde{g}_\ell$ := restriction of $\tilde{g}$ to $R_\ell$.

$J_\ell := \det\left(\frac{\partial x_i}{\partial y_j}\Big|_{x_i \in R_\ell}\right)$

$\Rightarrow \boxed{f_{\tilde{Y}}(\tilde{y}) = \sum_{\ell=1}^{L} f_{\tilde{X}}\left(\tilde{g}_\ell(\tilde{y})\right) \cdot |J_\ell|}$

$\begin{cases} R = \bigcup_{\ell=1}^{L} R_\ell \\ R_\ell \cap R_{\ell'} = \phi \text{ for } \ell \neq \ell' \end{cases}$

**Return to proof**:

- the transfo $\begin{cases} y_1 = \min(x_1, \dots, x_n) \\ \vdots \\ y_n = \max(x_1, \dots, x_n) \end{cases}$ is NOT 1-1 (∵ Permuting the $x_i$'s leave the $y_j$'s unchanged.
- Define the region $R_\ell$ := the particular permutation of $x_i$'s such that: $\begin{cases} y_1 = x_{\ell,1} \\ \vdots \\ y_n = x_{\ell,n} \end{cases}$ ⇒ the transfo restricted to $R_\ell$ is 1-1!

  → So $J_\ell = \det\begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix}$ for this particular permutation.

  For any other permutation: $|J_\ell| = 1$ (∵ all rows have one 1 + all the rest is 0)

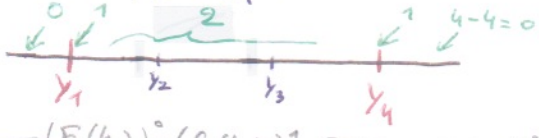
  → L = # of possible regions = # of possible permutate of the $x_i$'s $\Rightarrow L = n!$
- So: $f_{\tilde{Y}}(\tilde{y}) = \sum_{\ell=1}^{n!} f_{\tilde{X}}(y_1, \dots, y_n) \cdot 1 \underset{\text{IID}}{=} \sum_{\ell=1}^{n!} \prod_{i=1}^{n} f_X(y_i) \Rightarrow f_{\tilde{Y}}(\tilde{y}) = \begin{cases} n! \prod_{i=1}^{n} f_X(y_i), & y_1 < y_2 < \dots < y_n \\ 0, & \text{otherwise.} \end{cases}$

## (IV) Application:

Q: $\{X_i\}_{i=1}^{4}$ are IID exp(β) r.v. with PDF: $f_X(x) = \begin{cases} \frac{1}{\beta} e^{-\frac{x}{\beta}}, & x \ge 0 \\ 0, & x < 0 \end{cases}$

Find the joint PDF of $Y_1$ & $Y_4$:

A1: Multinomial way:




$$f_{Y_1,Y_4}(y_1,y_4) = \frac{4!}{0!\,1!\,2!\,1!\,0!} (F_x(y_1))^0 \cdot (f_x(y_1))^1 \cdot [F_x(y_4)-F_x(y_1)]^2 \cdot (f_x(y_4))^1 \cdot [1-F_x(y_4)]^0$$

$$\Rightarrow f_{Y_1,Y_4}(y_1,y_4) = \begin{cases} \frac{2}{\beta^2} e^{-\frac{y_1+y_4}{\beta}} \cdot \left[e^{-\frac{y_1}{\beta}} - e^{-\frac{y_4}{\beta}}\right]^2, & y_1 < y_4 \\ 0, & \text{elsewhere} \end{cases}$$

A2 = Integrate over the others:

$$f_{Y_1,Y_4}(y_1,y_4) = \begin{cases} 4! \int_{y_1}^{y_4} dy_2 \int_{y_2}^{y_4} dy_3 \frac{e^{-y_1/\beta}}{\beta} \cdot \frac{e^{-y_2/\beta}}{\beta} \cdot \frac{e^{-y_3/\beta}}{\beta} \cdot \frac{e^{-y_4/\beta}}{\beta}, & y_1 < y_4 \\ 0, & \text{elsewhere} \end{cases}$$

(∵ $y_1 < y_2 < y_3 < y_4$ ⇒ $y_1 < y_2 < y_4$, $y_2 < y_3 < y_4$)


↳ Same result as above, but painful!

# Sampling distributions :

## (1) Sample Mean & Sample Variance :

**Sample Mean** = $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

**Sample Variance** = $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$

} where $X_i$'s are **IID**

**Sampling distribution:** The distribution of functions of the observations $X_1, \dots, X_n$.

**Prop :** $X_1, \dots, X_n$ IID $N(\mu, \sigma^2)$ r.v. : $\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$ (*) use MGF!

Note : exact relation, not a limit!

**Prop :** $X_1, \dots, X_n$ IID r.v. with mean $\mu$ & variance $\sigma^2$ :

- $\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \xrightarrow{D} Z \sim N(0,1)$ as $n \to \infty$ (*) use CLT! (Convergence in distribution : the CDF converges to CDF of limit $F_Z$ at every continuity point $x$ of $F_Z$)
- $\frac{\overline{X} - \mu}{S/\sqrt{n}} \xrightarrow{D} Z \sim N(0,1)$ as $n \to \infty$ (*) Slutsky : $\left.\begin{matrix} X_n \xrightarrow{D} X \\ Y_n \xrightarrow{P} a \in \mathbb{R} \end{matrix}\right\} \Rightarrow Y_n X_n \xrightarrow{D} aX$

$S \xrightarrow{P} \sigma \Rightarrow \begin{cases} Y_n = S/\sigma \xrightarrow{P} 1 \\ Z_n = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \xrightarrow{D} N(0,1) \end{cases}$

**Prop :** $X_1, \dots, X_n$ IID $Ber(p)$ r.v. : $X := \sum_{i=1}^{n} X_i$ , $\hat{p} := \frac{X}{n}$ (↳ $\begin{cases} \mu = p \\ \sigma^2 = p(1-p) \end{cases}$)

Note : $X$ = # of success $\Rightarrow$ Binomial distrib°

- $\frac{\frac{X}{n} - p}{\sqrt{p(1-p)/n}} \xrightarrow{D} Z \sim N(0,1)$ as $n \to \infty$ (*) result 2.a with $\mu = p$, $\sigma^2 = \sqrt{p(p-1)}$

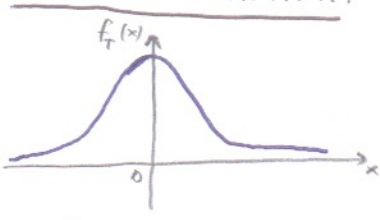
$\Rightarrow$ Normal approximat° to Binomial distrib

- $\frac{\frac{X}{n} - p}{\sqrt{\hat{p}(1-\hat{p})/n}} \xrightarrow{D} Z \sim N(0,1)$ as $n \to \infty$ (*) Slutsky : $\hat{p} = \frac{X}{n} \xrightarrow{P} p \Rightarrow \begin{cases} Y_n = \frac{\hat{p}(1-\hat{p})}{p(1-p)} \xrightarrow{P} 1 \\ Z_n = \frac{\frac{X}{n} - p}{\sqrt{\hat{p}(1-\hat{p})/n}} \xrightarrow{D} N(0,1) \end{cases}$

**Prop :** $X_1, \dots, X_n$ IID $N(\mu, \sigma^2)$ r.v. : $\frac{\overline{X} - \mu}{S/\sqrt{n}} \sim t_\nu$ $\Rightarrow$ (Student) $t$ distribution with $\nu$ degrees of freedom

Proof = Next Section !

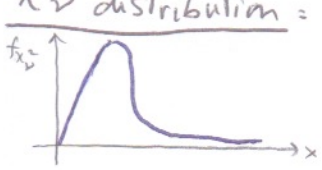
**Student $t$ distribution :**




- $f_{T_\nu}(x) = \frac{\Gamma(\frac{1+\nu}{2})}{\sqrt{\pi\nu}\,\Gamma(\frac{\nu}{2})} \cdot \left(1 + \frac{x^2}{\nu^2}\right)^{-\frac{1+\nu}{2}}$ $\forall x \in \mathbb{R}$
- Symmetric about zero $\forall \nu$
- $Y \sim t_1$ has a Cauchy distribution
- $Y \sim t_\nu$ does not have a $\nu^{th}$ moment : $Y$ has moments of order $r < \nu \Rightarrow E(|Y|^\nu) = \infty$

$\longrightarrow f_Y(y)$ has a heavier tail than the normal distrib° (The Normal dist has all its moments)

**$\chi^2_\nu$ distribution :**




- $f_{\chi^2_\nu}(x) = \frac{1}{2^{\frac{\nu}{2}}\Gamma(\frac{\nu}{2})} x^{\frac{\nu}{2}-1} e^{-\frac{x}{2}}$ $\forall x \geq 0$ $\Rightarrow$ Gamma distrib° : $\begin{cases} \alpha = \nu/2 \\ \beta = 2 \end{cases}$
- $M_{\chi^2_\nu}(t) = \frac{1}{(1-2t)^{\nu/2}}$ , $\forall t < \frac{1}{2}$ $\longrightarrow M_{Gamma(\alpha,\beta)}(t) = \frac{1}{(1-\beta t)^\alpha}$ $\forall t < \frac{1}{\beta}$
- $f_{\chi^2_\nu}(x)$ is NOT Symmetric about zero !

**Prop :** $X_1, \dots, X_n$ independent $\chi^2_{\nu_i}$ r.v. : $\sum_{i=1}^{n} X_i \sim \chi^2_{\sum_{i=1}^{n} \nu_i}$ (*) Use MGF !

# (11) Student T distribution :

**Theorem** : $X_1, \dots, X_n$ IID $N(\mu, \sigma^2)$ r.v. : $\boxed{T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}}$

**Lemma 1** : Let $\begin{cases} Y \sim \chi^2_\nu \\ Z \sim N(0,1) \end{cases}$ : then if $Y \perp Z$ : $\boxed{\frac{Z}{\sqrt{Y/\nu}} \sim t_\nu}$

$\frac{N(0,1)}{\sqrt{\chi^2_\nu/\nu}} \sim t_\nu$

**Proof** : See HW 1

**Lemma 2** : $X_1, \dots, X_n$ IID $N(\mu, \sigma^2)$ r.v. : Then $\boxed{\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}}$ when $\bar{X} \perp S^2$

**Proof** : • $\frac{1}{\sigma^2}\sum_{i=1}^n (X_i - \mu)^2 = \frac{1}{\sigma^2}\sum_{i=1}^n (X_i - \bar{X} + \bar{X} - \mu)^2 = \frac{1}{\sigma^2}\sum_{i=1}^n (X_i - \bar{X})^2 + \left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}\right)^2$

**Recall** : $Z \sim N(0,1) \Rightarrow Z^2 \sim \chi^2_1$

So $\underbrace{\sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma}\right)^2}_{\sim N(0,1)} \sim \chi^2_n \Rightarrow \underbrace{\sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma}\right)^2}_{W \sim \chi^2_n} = \underbrace{\frac{(n-1)S^2}{\sigma^2}}_{Y \sim ?} + \underbrace{\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}\right)^2}_{U \sim \chi^2_1}$

$\sim \chi^2_1$ (as $X_i$'s indep)

• So $W = Y + U$ : Lemma 3 $\Rightarrow \bar{X} \perp S^2 \Rightarrow Y \perp U$

So $M_W(t) = M_{Y+U}(t) \underset{Y \perp U}{=} M_Y(t) M_U(t) \Rightarrow \frac{1}{(1-2t)^{\frac{n}{2}}} = M_Y(t) \cdot \frac{1}{(1-2t)^{\frac{1}{2}}} \Rightarrow M_Y(t) = \frac{1}{(1-2t)^{\frac{n-1}{2}}}$

$\Rightarrow Y = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$

**Lemma 3** : $X_1, \dots, X_n$ IID $N(\mu, \sigma^2)$ r.v. : Then $\boxed{\bar{X} \perp S^2}$ → Characterizes the Normal dist (only true for $N(\mu, \sigma^2)$)

**Proof** : Idea : Write $S^2$ as a funct° of $(X_2 - \bar{X}, \dots, X_n - \bar{X})$ ⇒ Show : $\bar{X} \perp (X_2 - \bar{X}, \dots, X_n - \bar{X})$

WLOG : $\mu = 0$, $\sigma^2 = 1$ (*) reScale the Normal dist

• $S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n-1}\left((X_1 - \bar{X})^2 + \sum_{i=2}^n (X_i - \bar{X})^2\right) \underset{\sum_{i=1}^n (X_i - \bar{X}) = 0}{=} \frac{1}{n-1}\left[\sum_{i=2}^n (X_i - \bar{X})\right] + \frac{1}{n-1}\sum_{i=2}^n (X_i - \bar{X})^2$

• Transfo : $\begin{cases} Y_1 = \bar{X} \\ Y_i = X_i - \bar{X} \quad \forall 2 \le i \le n \end{cases} \Rightarrow \begin{cases} y_1 = \bar{x} \\ y_i = x_i - \bar{x} \end{cases} \Rightarrow \begin{cases} x_1 = n y_1 - \sum_{i=2}^n (y_1 + y_i) \\ x_i = y_1 + y_i \end{cases}$

$\Rightarrow \begin{cases} x_1 = y_1 - \sum_{i=2}^n y_i \\ x_i = y_1 + y_i \end{cases} \forall 2 \le i \le n$ is a 1-1 Transfo $\Rightarrow |J| = \left|\det \begin{pmatrix} 1 & -1 & -1 & \cdots & -1 \\ 1 & 1 & 0 & \cdots & 0 \\ 1 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \cdots & 1 \end{pmatrix}\right| = \left|\det \begin{pmatrix} n & 0 & 0 & \cdots & 0 \\ 1 & 1 & 0 & \cdots & 0 \\ 1 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \cdots & 1 \end{pmatrix}\right|$

$\Rightarrow \boxed{|J| = n}$

**Recall** : $X \perp Y$ r.v. $\Rightarrow X \perp g(Y)$

• So : $f_{Y_1, \dots, Y_n}(y_1, \dots, y_n) = f_{X_1, \dots, X_n}(x_1, \dots, x_n) \cdot |J|$

$\underset{[X_i \text{ are IID}]}{=} \left(\prod_{i=1}^n f_X(x_i)\right) \cdot n = \frac{n}{(2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2}\sum_{i=1}^n x_i^2} \underset{\text{recall}}{=} c \cdot e^{-\frac{1}{2}\left(x_1^2 + \sum_2^n x_i^2\right)} = c\, e^{-\frac{1}{2}\left[y_1 - \sum_2^n y_i\right)^2 + \sum_2^n (y_1 - y_i)^2\right]}$

$= c \underbrace{e^{-\frac{1}{2} n y_1^2}}\ \underbrace{e^{-\frac{1}{2}\left[\left(\sum_2^n y_i\right)^2 + \sum_2^n y_i^2\right]}}$

$\Rightarrow f_{Y_1, \dots, Y_n}(y_1, \dots, y_n) = c\, g_1(y_1)\, g_2(y_2, \dots, y_n) \Rightarrow Y_1 \perp (Y_2, \dots, Y_n) \Rightarrow \bar{X} \perp (X_2 - \bar{X}, \dots, X_n - \bar{X}) \Rightarrow \bar{X} \perp S^2$.

**Proof** : $T = \frac{\bar{X} - \mu}{S/\sqrt{n}} = \frac{(\bar{X} - \mu)/(\sigma/\sqrt{n})}{\sqrt{\frac{(n-1)S^2}{\sigma^2}/n-1}} \sim \frac{Z}{\sqrt{Y/\nu}} \underset{\text{Lemma 1}}{\sim} t_\nu$ where : $\begin{cases} Z \sim N(0,1) & ((*)\ X \sim N(\mu, \sigma^2)) \\ Y \sim \chi^2_{n-1} & ((*)\ \text{Lemma 2 + 3}) \end{cases}$

# III F-distribution:

**Thm:** $(X_1, \dots, X_{n_1}) \perp (Y_1, \dots, Y_{n_2})$ r.v. with $\begin{cases} X_i\text{'s} & \text{IID } N(\mu_1, \sigma^2) \\ Y_i\text{'s} & \text{IID } N(\mu_2, \sigma^2) \end{cases}$ Note: Same $\sigma^2$!

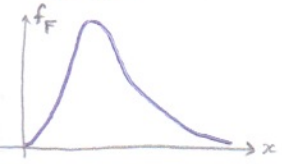
Then: $\frac{S_1^2}{S_2^2} \sim F_{n_1-1, n_2-1}$ with: $\begin{cases} S_1^2 = \frac{1}{n_1-1} \sum_1^{n_1} (X_i - \bar{X}) \\ S_2^2 = \frac{1}{n_2-1} \sum_1^{n_2} (Y_i - \bar{Y}) \end{cases}$

**Proof:** $\frac{S_1^2}{S_2^2} = \frac{(n_1-1)S_1^2}{(n_1-1)\sigma^2} \Big/ \frac{(n_2-1)S_2^2}{(n_2-1)\sigma^2} \sim \frac{\chi^2_{n_1-1}/n_1-1}{\chi^2_{n_2-1}/n_2-1}$ (indep!) $\sim F_{n_1-1, n_2-1}$ (↑ Lemma)

(ps) $X_i \perp Y_j \Rightarrow S_1^2 \perp S_2^2 \Rightarrow \frac{(n_1-1)S_1^2}{\sigma^2} \perp \frac{(n_2-1)S_2^2}{\sigma^2}$

**F-distribution** = $F_{\nu_1, \nu_2}$ : F distribution with degrees of freedom $\nu_1$ & $\nu_2$.

**Prof:**




- $f_{F_{\nu_1,\nu_2}}(x) = \begin{cases} \frac{\Gamma\left(\frac{\nu_1+\nu_2}{2}\right)}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)} \cdot \left(\frac{\nu_1}{\nu_2}\right)^{\frac{\nu_1}{2}} \cdot \frac{x^{\frac{\nu_1}{2}-1}}{\left(1+y\frac{\nu_1}{\nu_2}\right)^{\frac{\nu_1+\nu_2}{2}}}, & 0 \le x < \infty \\ 0, & x < 0 \end{cases}$
- $Y \sim F_{\nu_1,\nu_2}$ has a Support over $[0, \infty)$
- PDF of $F_{\nu_1,\nu_2}$ is skewed to the right

**Lemma** = If $\begin{cases} U \sim \chi^2_{\nu_1} \\ V \sim \chi^2_{\nu_2} \end{cases}$ and $U \perp V$, then $\boxed{\frac{U/\nu_1}{V/\nu_2} \sim F_{\nu_1,\nu_2}}$

$\frac{\chi^2_{\nu_1}/\nu_1}{\chi^2_{\nu_2}/\nu_2} \sim F_{\nu_1,\nu_2}$

**Proof** = HW 1

# STATISTICAL INFERENCE:

Assume that we have observations $X_1, \dots, X_n$ drawn from some distribution $F(\cdot; \theta)$, where $\theta$ is an unknown parameter:

2 main problems:

1. Estimate $\theta$: make an educated guess about $\theta$.
2. Hypothesis testing: carry out a test of hypothesis about $\theta$.

# Terminology :

**Idea :** $X_1, X_2, \dots$ —EXPERIMENT→ $x_1, x_2, \dots$

Prior to Experiment (R.V.) — After the Experiment (fixed numbers → Data)

- Prior to the "experiment" (i.e. before observing our data) : the quantities $X_1, X_2, \dots$ are R.V.'s → we don't know yet what values they will assume!
- After the "experiment" : the realized values of $X_1, X_2, \dots$ are denoted by $x_1, x_2, \dots$ → Fixed numbers!

**statistic** = any measurable function $T(X_1, X_2, \dots, X_n)$ Ex : $\bar{X}$, $S^2$, $X_{26}$, 3.62, $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$, $\frac{\bar{X}-\mu}{S/\sqrt{n}}$, ...

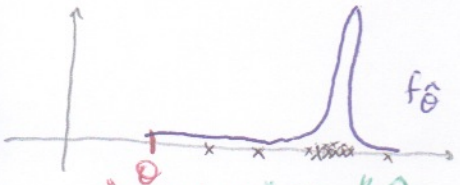
**Estimator** = $\hat{\Theta} = \hat{\Theta}(X_1, \dots, X_n)$ is an Estimator for $\Theta$ if $\hat{\Theta}$ is a statistic that does NOT depend functionally on any unknown parameters
→ $\hat{\Theta}$ is used to estimate $\Theta$!

Note :
- Before the experiment, $\hat{\Theta}(X_1, \dots, X_n)$ is a R.V. (∵ $\hat{\Theta}$ = function of R.V.'s)
  ⇒ $\hat{\Theta}$ has a probability distribution : call it its "(sampling) distribution".
- After the experiment, $\hat{\Theta}(x_1, \dots, x_n)$ is a constant (number or vector).
  ⇒ $\hat{\Theta}(x_1, \dots, x_n)$ is called an ESTIMATE of $\Theta$.

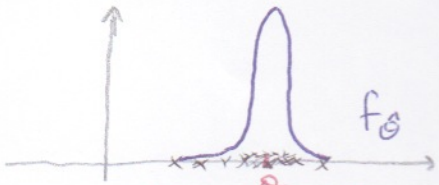
Ex = Given $X_1, \dots, X_n$ :
- $\hat{\Theta} = X_1$ is an estimator for $\mu$, $\sigma^2$, ...
- $\hat{\Theta} = X_1 + 3\mu$ is NOT estimator for $\mu$ (∵ $\hat{\Theta}$ depends on $\mu$)
- $\hat{\Theta} = \frac{X_1 + 2}{\sigma^2}$ is an estimator for $\mu$ only if we know $\sigma^2$.

**"Good" estimators** = We seek recipes that work well on average when we use it!
However, once the experiment is done, we don't know how well we have done:
we can only appeal to the fact that we used a recipe that works well on average.

**Idea** = Desirable to have a (sampling) distribution $f_{\hat{\Theta}}$ of $\hat{\Theta}$ closely concentrated about the parameter $\Theta$ that we want to estimate.




→ Most of the time, the $\hat{\Theta}$ we will get will be far from $\Theta$ because the sampling dist not concentrated about $\Theta$)




↳ sampling dist concentrated about $\Theta$ : most of the time, the $\hat{\Theta}$ we get will be close to $\Theta$.

On summary page = the procedures used are only "good" in the sense that they are "good on average" if the procedure was used repeatedly = we can never say how well we have done on a particular occasion!

# Unbiasedness :

→ one of the criteria of "goodness" that estimators should share!

## ① Definition :

unknown

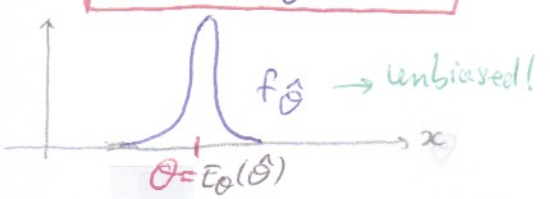
**Unbiased Estimator** = Let $X_1, .., X_n$ be RV's drawn from some distribution with parameter $\theta$

$\hat{\theta}$ is an unbiased estimator for $\theta$ if $\boxed{E_\theta(\hat{\theta}) = \theta, \forall \theta \in (H)}$

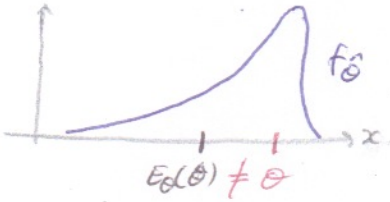
Note : Subscript of $E_\theta$ : use the distrib° of $(X_1, .., X_n)$ at the parameter $\theta$ when calculating the expectat° value

**Parameter Space** $(H)$ : set of all possible values of $\theta$

**Bias** : $\boxed{b(\theta) := E_\theta(\hat{\theta}) - \theta}$




→ unbiased!




→ Biased on one side of $\theta$

Note : usually start with a $\hat{\theta}$ proposal & check if biased : it is often possible to correct for the bias by redefining $\hat{\theta}$.

## ② Examples :

**Ex 1** : $X_1, .., X_n$ IID R.V. : let the mean $\mu$ be unknown

**Prop** = $\hat{\mu} = \overline{X}$ is an unbiased estimator for $\mu$ (**) $E_\mu(\hat{\mu}) = E_\mu(\overline{X}) = \mu \ \forall \mu$ by def of $\mu$

Ex : • $\hat{\mu} = X_1$ ($E_\mu(X_1) = \mu$ by def), $\hat{\mu} = \frac{X_1 + X_2}{2}$ ($E_\mu(\hat{\mu}) = \frac{E(X_1) + E(X_2)}{2} = \frac{\mu + \mu}{2} = \mu$.)

• Biased! : $\hat{\mu} = \overline{X} + 3$ (Correct : $\hat{\mu} \to \hat{\mu} - 3$) : (**) $E_\mu(\overline{X} + 3) = \mu + 3 \neq \mu \ \forall \mu$

$\hat{\mu} = 3$ (**) $E_\mu(\hat{\mu}) = E_\mu(3) = 3 \neq \mu \ \forall \mu \neq 3$)

**Ex 2** : $X_1, .., X_n$ IID R.V. with finite mean $\mu$ and variance $\sigma^2$ :

**Prop** : • $\mu$ known, $\sigma^2$ unknown : $\hat{\sigma}^2 = \frac{\sum_{i=1}^n (X_i - \mu)^2}{n}$ is unbiased for $\sigma^2$

(**) $E_{\sigma^2}(\hat{\sigma}^2) = \frac{1}{n} \sum_{i=1}^n E_{\sigma^2}((X_i - \mu)^2)$

$= \frac{1}{n} \sum_{i=1}^n (\sigma^2 + E(X_i - \mu)^2) = \sigma^2$ ✓

• $\mu$ & $\sigma^2$ unknown : $\hat{\sigma}^2 = S^2 = \frac{\sum_{i=1}^n (X_i - \overline{X})^2}{n-1}$ is unbiased for $\sigma^2$

(**) $S^2 = \frac{1}{n-1} \sum_{i=1}^n [(X_i - \mu) + (\mu - \overline{X})]$

$= \frac{1}{n-1} [\sum_1^n (X_i - \mu)^2 - n(\overline{X} - \mu)^2]$

$E_{\sigma^2}(S^2) = \frac{1}{n-1} [n\sigma^2 - n\frac{\sigma^2}{n}] = \sigma^2$ ✓

(**) $E(\overline{X} - \mu)^2 = \text{Var}(\overline{X}) = \sigma^2/n$

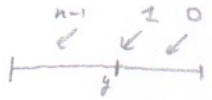
**Ex 3** : $X_1, .., X_n$ IID Bernoulli(p) R.V. : let p be unknown

**Prop** = $\hat{p} = \frac{\sum_{i=1}^n X_i}{n}$ ($X_i$ : $\begin{cases} 0, \text{ prob } p \\ 1, \text{ prob } 1-p \end{cases}$) is unbiased for p (**) $E_p(\hat{p}) = \frac{1}{n} E(\sum X_i) = \frac{1}{n}(np) = p$ ✓

**Ex 4** : $X_1, .., X_n$ IID $U(0, \theta)$ R.V. : let $\theta$ be unknown (So $(H) = \{\theta > 0\}$)

**Prop** : • $\hat{\theta} = 2\overline{X}$ is unbiased for $\theta$ (**) $E_\theta(\hat{\theta}) = 2E_\theta(\overline{X}) = 2\mu = \theta$ ✓

• $\hat{\theta} = 2X_3$ is unbiased for $\theta$ (**) $E_\theta(\hat{\theta}) = 2E_\theta(X_3) = 2\mu = \theta$ ✓ (!! IID !)

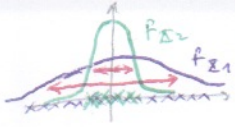



• $\hat{\theta} = Y_n$ ($n^{th}$ order stat) is biased for $\theta$ (**) $E(Y_n) = \int_{-\infty}^{\infty} y \cdot f_{Y_n}(y)\,dy$ where $f_{Y_n}(y) = \frac{n!}{(n-1)!1!0!} \cdot (\frac{y}{\theta})^{n-1} \cdot (\frac{1}{\theta})^1 \cdot (1 - \frac{y}{\theta})^0$

$= \int_0^\theta \frac{n}{\theta^n} y\, y^{n-1}\,dy = \frac{n}{\theta^n} \int_0^\theta y^n dy$

$\Rightarrow E(Y_n) = \frac{n}{n+1}\theta$ → correct it!

$f_{Y_n}(y) = \begin{cases} \frac{n y^{n-1}}{\theta^n}, & 0 < y < \theta \\ 0, & \text{o'wise} \end{cases}$




# Uniform Minimum-Variance Unbiased estimators: UMVUEs!

**Uniform minimum variance unbiased estimators = (UMVUEs)**

Let $\theta$ be an unknown parameter, $\theta \in (H)$:

Then $\hat{\theta}$ is a UMVUE for $\theta$ if:
- $E_\theta(\hat{\theta}) = \theta \quad \forall \theta \in (H)$
- If $\hat{\theta}^*$ is another unbiased estimator for $\theta$: $\mathrm{Var}(\hat{\theta}^*) \geq \mathrm{Var}(\hat{\theta})$ $\forall \theta \in (H)$.

Note: UMVUEs do not always exist!

Prop = If a UMVUE exists, it is unique! ⇝ See HW2!

one way to find UMVUEs (when they exist) (another method later!)

**Cramér-Rao Ineq =**

$\tilde{\Sigma} = (\Sigma_1, \dots, \Sigma_n)$ a vector of r.v. with joint PDF: $F_{\tilde{X}}(\tilde{x};\theta)$, depending on an unknown scalar $\theta \in (H)$. (↳ true also for PMF)

then under certain regularity conditions:

if $E(T(\tilde{\Sigma})) = g(\theta)$, $T$ a statistic, $g$ a differentiable f°

then
$$\mathrm{Var}(T(\tilde{\Sigma})) \geq \frac{(g'(\theta))^2}{E\left[\left(\frac{\partial}{\partial\theta} \ln f_{\tilde{X}}(\tilde{\Sigma},\theta)\right)^2\right]} \overset{\Sigma_i \text{ IID}}{=\!=} \frac{(g'(\theta))^2}{n E\left[\left(\frac{\partial}{\partial\theta} \ln f_{X}(\Sigma,\theta)\right)^2\right]}$$
($f_{X(x,\theta)}$)

Note:
- If $\hat{\theta}$ is an unbiased estimator for $\theta$, then: if $\mathrm{Var}(\theta)$ = C-R lower bound, $\hat{\theta}$ UMVUE!
- If $\hat{\theta}$ is unbiased, $g(\theta) = \theta \Rightarrow g'(\theta) = 1$
- C-R holds for PMFs $P_{\tilde{X}}$ also!

Corr: If $\Sigma_1, \dots, \Sigma_n$ are IID:
$$\mathrm{Var}(T(\tilde{\Sigma})) \geq \frac{(g'(\theta))^2}{-n \cdot E\left[\frac{\partial^2}{\partial\theta^2} \ln f_{\Sigma}(\Sigma,\theta)\right]}$$
(cf) HW2

**Proof: (CR)**

- claim: $E\left[\frac{\partial}{\partial\theta} \ln f_{\tilde{X}}(\tilde{\Sigma},\theta)\right] \overset{\forall\theta\in(H)}{=} 0$ under the reg. cond°

(cf) $E\left[\frac{\partial}{\partial\theta} \ln f_{\tilde{X}}(\tilde{\Sigma},\theta)\right] = \int_{-\infty}^{\infty} \frac{\partial}{\partial\theta} \ln f_{\tilde{X}}(\tilde{x},\theta) \cdot f_{\tilde{X}}(\tilde{x},\theta)\, d\tilde{x} = \int_{-\infty}^{\infty} \frac{f'_{\tilde{X}}(\tilde{x},\theta)}{f_{\tilde{X}}(\tilde{x},\theta)} f_{\tilde{X}}(\tilde{x},\theta)\, d\tilde{x} = \frac{\partial}{\partial\theta} \int_{-\infty}^{\infty} f_{\tilde{X}}(\tilde{x},\theta)\, d\tilde{x} = \frac{\partial}{\partial\theta}(1) = 0$

- claim: $|g'(\theta)| \leq \left[E\left[(T(\tilde{\Sigma}) - g(\theta))^2\right]\right]^{1/2} \left[E\left[\left(\frac{\partial}{\partial\theta} \ln f_{\tilde{X}}(\tilde{\Sigma},\theta)\right)^2\right]\right]^{1/2}$

(cf) $|g'(\theta)| = \left|\frac{\partial}{\partial\theta} g(\theta)\right| = \left|\frac{\partial}{\partial\theta} E(T(\tilde{\Sigma}))\right| = \left|\frac{\partial}{\partial\theta} \int_{-\infty}^{\infty} T(\tilde{x}) \cdot f_{\tilde{X}}(\tilde{x},\theta)\, d\tilde{x}\right| \overset{\text{reg cond°!}}{=} \left|\int_{-\infty}^{\infty} T(\tilde{x}) \cdot \frac{\partial}{\partial\theta} f_{\tilde{X}}(\tilde{x},\theta)\, d\tilde{x}\right|$ (∫ indep of θ (by estimator!))

$\leq \int_{-\infty}^{\infty} |T(\tilde{x})| \cdot \left|\frac{\partial}{\partial\theta} \ln f_{\tilde{X}}(\tilde{x},\theta)\right| \cdot f_{\tilde{X}}(\tilde{x},\theta)\, d\tilde{x} \leq \int_{-\infty}^{\infty} |T(\tilde{x}) - g(\theta)| \cdot \left|\frac{\partial}{\partial\theta} \ln f_{\tilde{X}}(\tilde{x},\theta)\right| \cdot f_{\tilde{X}}(\tilde{x},\theta)\, d\tilde{x}$ (↳ add "zero" by prev. claim!)

(Cauchy-Schwartz) $\leq \left(E\left[(T(\tilde{\Sigma}) - g(\theta))^2\right]\right)^{1/2} \left(E\left[\left(\frac{\partial}{\partial\theta} \ln f_{\tilde{X}}(\tilde{\Sigma},\theta)\right)^2\right]\right)^{1/2}$

- $E(T(\tilde{\Sigma})) = g(\theta) \Rightarrow \mathrm{Var}(T(\tilde{\Sigma})) = E\left[(T(\tilde{\Sigma}) - g(\theta))^2\right]$ ⟶ Square both sides ⇒ Get CR ✓
- By 1st claim: $E\left[\left(\frac{\partial}{\partial\theta} \ln f_{\tilde{X}}(\tilde{\Sigma},\theta)\right)^2\right] \overset{E(\cdot)=0}{=} \mathrm{Var}\left[\frac{\partial}{\partial\theta} \ln f_{\tilde{X}}(\tilde{\Sigma},\theta)\right] \overset{\Sigma_i \text{ indep}}{=\!=} \mathrm{Var}\left[\sum_{i=1}^{n} \frac{\partial}{\partial\theta} \ln f_{X}(\Sigma_i,\theta)\right]$

$= \sum_{i=1}^{n} \mathrm{Var}\left[\frac{\partial}{\partial\theta} \ln f_{X}(\Sigma_i,\theta)\right] = \sum_{i=1}^{n} E\left[\left(\frac{\partial}{\partial\theta} \ln f_{X}(\Sigma_i,\theta)\right)^2\right] = n E\left[\left(\frac{\partial}{\partial\theta} \ln f_{X}(\Sigma,\theta)\right)^2\right]$

($\Sigma_i$ indep; $E(\cdot)=0$)

(cf) $\Sigma_i$ IID ⇒ $f_X(\Sigma_i)$ IID ⇒ $\frac{\partial}{\partial\theta} \ln f_X(\Sigma_i)$ IID ⇒ Var are equal!

## Examples:

- **Bernoulli** = $X_1, \dots, X_n$ IID Bernoulli $(p)$ r.v.'s, $p$ unknown: $\hat{p} = \frac{\bar{X}}{n} = \frac{\sum_{i=1}^{n} X_i}{n}$ UMVUE for $p$!

  (*) **Goal**: $\mathrm{Var}[\hat{p}]$ = CR Lower bound!

  - $\mathrm{Var}[\hat{p}] = \mathrm{Var}\left(\frac{\bar{X}}{n}\right) = \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n}$ (*) Indep $\Rightarrow$ Var $\Sigma$ = $\Sigma$ Var; Id. dist. $\Rightarrow p(1-p)$
  - CR Lowerbound:

$$nE\left[\left(\frac{\partial}{\partial p} \ln P_X(\bar{X}, p)\right)^2\right] = nE\left[\left(\frac{\partial}{\partial p}\ln\left(p^{\bar{X}}(1-p)^{1-\bar{X}}\right)\right)^2\right] = nE\left[\left(\frac{\partial}{\partial p}\left(\bar{X}\ln p + (1-\bar{X})\ln(1-p)\right)\right)^2\right]$$

$$= nE\left[\left(\frac{\bar{X}}{p} - \frac{1-\bar{X}}{1-p}\right)^2\right] = nE\left[\left(\frac{\bar{X}-p}{p(1-p)}\right)^2\right] = \frac{n}{(p(1-p))^2}E\left[(\bar{X}-p)^2\right] = \frac{n}{(p(1-p))^2}\mathrm{Var}(\bar{X}) = \frac{n}{(p(1-p))^2}p(1-p)$$

$$= \frac{n}{p(1-p)} = \frac{1}{\mathrm{Var}\,\hat{p}}$$

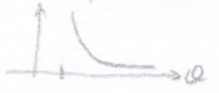
$\Rightarrow$ CR lower bound = $\mathrm{Var}[\hat{p}]$ ✓

- **Normal** = $X_1, \dots, X_n$ IID $N(\mu, \sigma^2)$ r.v.'s:

  - $\mu$ unknown: $\hat{\mu} = \bar{X}$ is UMVUE for $\mu$ (True for $\sigma^2$ known, and for $\sigma^2$ unknown) — HW 2
  - $\sigma^2$ unknown: $\longrightarrow \mu = \mu_0$ known: $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu_0)^2$
    $\longrightarrow \mu$ unknown: $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$
    } UMVUEs for $\sigma^2$

- **Uniform** = $X_1, \dots, X_n$ IID $U(0, \theta)$ r.v.'s, $\theta$ unknown:
  - $\hat{\theta} = 2\bar{X}$ and $\hat{\theta} = \frac{n+1}{n}Y_n$ are unbiased estimators for $\theta$
  - $\hat{\theta} = \frac{n+1}{n}Y_n$ is the UMVUE $\rightsquigarrow \mathrm{Var}(2\bar{X}) > \mathrm{Var}\left(\frac{n+1}{n}Y_n\right)$

  ⚠ Cannot use C-R: regularity cond° fail! (*) range of PDF depends on the param $\theta$.




**Note:**
- Some UMVUEs do not reach the CR lower bound, even when reg cond° satisfied
- UMVUEs depend on the distribution family (Ex: $S^2$ UMVUE for $X_i \sim N(\mu, \sigma^2)$; $S^2$ NOT UMVUE for $\beta^2 = \mathrm{Var}(\bar{X})$, $X_i \sim \mathrm{Exp}(\beta)$)

**Mean-Square Error** = MSE is the "Average" squared distance between $\hat{\theta}$ and $\theta$:

$\hat{\theta}$ an estimator for $\theta$: $\mathrm{MSE}(\hat{\theta}) = E\left[(\hat{\theta} - \theta)^2\right]$

**Thm** = $\mathrm{MSE}(\hat{\theta}) = \mathrm{Var}(\hat{\theta}) + (\mathrm{Bias})^2$ where $\mathrm{Bias} = (E(\hat{\theta}) - \theta)$

**Proof** = $\mathrm{MSE}(\hat{\theta}) = E\left[(\hat{\theta} - \theta)^2\right] = E\left[(\hat{\theta} - E[\hat{\theta}] + E[\hat{\theta}] - \theta)^2\right]$

$$= \mathrm{Var}(\hat{\theta}) + \underbrace{E\left[(\hat{\theta} - E(\hat{\theta}))\cdot(E(\hat{\theta}) - \theta)\right]}_{=0} + \left(E[\hat{\theta} - \theta]\right)^2$$

(*) $= \underbrace{E(\hat{\theta} - \theta)}_{\text{CST}} \cdot \underbrace{\left[E(\hat{\theta} - E(\theta))\right]}_{=0}$

**Note** = $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 \longrightarrow$ Bias = 0; larger MSE!

$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 \longrightarrow$ Bias > 0; smaller MSE!

} → under Normality!

# Consistency :

As sample size $n \to \infty$ : $\hat{\theta} \xrightarrow{P} \theta$

**Consistency** = An estimator $\hat{\theta}$ for $\theta$ is consistent for $\theta$ if : $\hat{\theta} \xrightarrow{P} \theta$ as $n \to \infty$

i.e. : $P(|\hat{\theta} - \theta| > \varepsilon) \to 0$ as $n \to \infty$ $\forall \varepsilon > 0$.

i.e. : $P(|\hat{\theta} - \theta| < \varepsilon) \to 1$ as $n \to \infty$

## Methods for establishing consistency =

1] Try to use the weak law of large nb (WLLN) :

**WLLN** = $X_1, \dots, X_n$ IID r.v. with finite mean $\mu$ : Then $\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} \mu$

**Method** =
- Try to write $\hat{\theta}(\bar{X})$ as $\hat{\theta}(\bar{X})$ for some f° $\hat{\theta}$ of $\bar{X}$.
- Then $g(\bar{X}) \xrightarrow{P} g(\mu)$ (••) $\bar{X} \xrightarrow{P} \mu$
- If $g(\mu) = \theta$, then we are done!

2] If : $\begin{cases} E(\hat{\theta}_n) \to \theta \\ \mathrm{Var}(\hat{\theta}_n) \to 0 \end{cases}$ as $n \to \infty$, then $\hat{\theta}_n \xrightarrow{P} \theta$

3] Directly : use the def of convergence in probability

## Examples =

- $\hat{\mu} = \bar{X}$ is consistent for $\mu$ (••) WLLN
- $S^2$ is consistent for $\sigma^2$ (if $E(X^2) < \infty$) (••) $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} X_i^2 - n\bar{X}^2\right] = \frac{n}{n-1}\left[\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2\right]$
  $X_i$ IID $\Rightarrow X_i^2$ IID, $E(X_i^2) < \infty$ : WLLN $\Rightarrow \frac{1}{n}\sum_{i=1}^{n} X_i^2 \xrightarrow{P} E[X_i^2] = \sigma^2 + \mu^2$
  $\Rightarrow S^2 \to \frac{n}{n+1}[\sigma^2 + \mu^2 - \mu^2] \to \sigma^2$ as $n \to \infty$.
- $X_1, \dots, X_n$ IID Bernoulli(p) : $\hat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i$ Sample proportion (prob of Success)
  $\hat{p}$ is consistent for $p$ (••) WLLN : replace "$\mu$" by "$p$".
- $X_1, \dots, X_n$ IID Exp($\beta$) : $e^{-10/\bar{X}}$ is consistent for $P(X_1 > 10) = e^{-10/\beta}$ (••) WLLN : $\bar{X} \xrightarrow{P} E(X_1) = \beta$
  $x \mapsto e^{-10/x}$ cont $\Rightarrow e^{-10/\bar{X}} \xrightarrow{P} e^{-10/\beta}$
  Note : $\mathrm{Var}(X_i) = \beta^2 \Rightarrow \begin{cases} \bar{X}^2 \xrightarrow{P} \beta^2 \\ S^2 \xrightarrow{P} \beta^2 \end{cases}$ has min variance!
- $X_1, \dots, X_n$ IID $U(0, \theta)$ : $\begin{cases} 2\bar{X} \xrightarrow{P} 2\mu = \theta \\ \frac{n+1}{n} Y_n \xrightarrow{P} \theta \end{cases}$ are both consistent for $\theta$ ! (••) $2\bar{X} \to 2\mu = 2\frac{\theta}{2} = \theta$.

**Proof** : • **direct** : Note : $\frac{n+1}{n} Y_n$ consistent for $\theta \Leftrightarrow Y_n$ consistent for $\theta$. (••) $\frac{n+1}{n} \to 1$

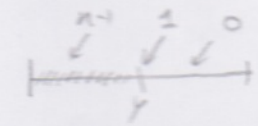
Trick : $P(\max \le x) = P(\text{all } X_i \le x)$, $P(\min \ge x) = P(\text{all } X_i \ge x)$

Given $\varepsilon > 0$ : $P[|Y_n - \theta| > \varepsilon] = P[|\theta - Y_n| > \varepsilon] = P[\theta - Y_n > \varepsilon] = P[Y_n < \theta - \varepsilon] \overset{(*)}{=} (1 - \frac{\varepsilon}{\theta})^n$

(*) $P(Y_n < \theta - \varepsilon) = P(X_1 < \theta - \varepsilon, \dots, X_n < \theta - \varepsilon)$ [$X_i$ IID] $= [P(X_1 < \theta - \varepsilon)]^n = (F_{X_1}(\theta - \varepsilon))^n = (\frac{\theta - \varepsilon}{\theta})^n = (1 - \frac{\varepsilon}{\theta})^n$

$P(Y_n > \theta) = 0$ ; $= 0$ if $\varepsilon > \theta$ ; $\varepsilon < \theta$

- **Alternate** =
  - $E(Y_n) \to \theta$ (••) $E(Y_n) = (\frac{n}{n+1})\theta \to \theta$
  - $\mathrm{Var}(Y_n) \to 0$ (••) $f_{Y_n}(y) = \frac{n!}{1!0!(n-1)!}\left(\frac{y}{\theta}\right)^{n-1} \cdot \frac{1}{\theta} = \frac{n y^{n-1}}{\theta^n}$ for $0 < y < \theta$




$\Rightarrow \mathrm{Var}(Y_n) = E(Y_n^2) - [E(Y_n)]^2 = \int y^2 \frac{n y^{n-1}}{\theta^n} dy - \left(\frac{n}{n+1}\theta\right)^2 = \left(\frac{n}{n+2}\right)\theta^2 - \left(\frac{n}{n+1}\right)^2 \theta \to 0$

# Sufficiency :

→ Once you have the value of a Sufficient statistic, you have extracted all the information contained in the observations about the parameter $\Theta$.

## ① Definition :

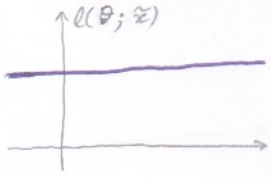
**Likelihood function**: Let $f_{\tilde{X}}(\tilde{x};\Theta)$ ($P_{\tilde{X}}(\tilde{x};\Theta)$) be the PDF of $\tilde{X}$ at $\Theta$ fixed

$\ell(\Theta;\tilde{x})$ is the function of $\Theta$ at $\tilde{x}$ fixed ⇒ "likelihood function"

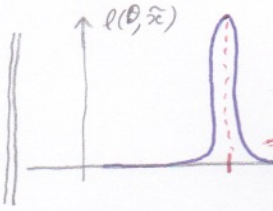
**Note**:
- To compute $\ell(\Theta;\tilde{x})$, you need $f_{\tilde{X}}(\tilde{x};\Theta)$ for $\Theta$ and $\tilde{x}$.
- All the info contained in the data $\tilde{x}$ (fixed) about $\Theta$ is captured by $\ell(\Theta,\tilde{x})$.

→ How the Likelihood changes with $\Theta$!

**Ex**:




⇒ Not very informative about $\Theta$: $\ell(\Theta;\tilde{x})$ varies little at $\Theta$ for any given $\tilde{x}$.




⇒ Informative about $\Theta$: $\ell(\Theta;\tilde{x})$ changes markedly with $\Theta$. gives the values of $\Theta$ that are more plausible to have led to the data $\tilde{x}$.

**Sufficiency**: Let $\tilde{X}$ have joint PDF (PMF) $f_{\tilde{X}}(\tilde{x};\Theta)$

A statistic $T$ is Sufficient for $\Theta$ if $f_{\tilde{X}|T}(\tilde{x};\Theta \mid T(\tilde{X})=t)$ is indep of $\Theta$ $\forall \tilde{x}, \forall t$

→ once you know $t=t(\tilde{x})$ for every $\tilde{x}$, there is no info left in the data about $\Theta$.

**Ex**: $X_1,\dots,X_n$ IID Bernoulli(p) r.v, p unknown : $T(\tilde{X})=\sum_{i=1}^n X_i$ is sufficient for p ($T$ = # of Successes)

Goal: Fix $\tilde{x}$ and $\tilde{x}$ : $P_{\tilde{X}|T}(\tilde{x};p \mid T(\tilde{X})=t)$ indep of p $\forall t$

(def) $P_{\tilde{X}|T}(\tilde{x};p \mid T(\tilde{X})=t) = P(X_1=x_1,\dots,X_n=x_n;p \mid \sum_{i=1}^n X_i = t)$

- if $t \neq \sum_{i=1}^n x_i$, then $\sum_{i=1}^n X_i \neq t \Rightarrow P_{\tilde{X}|T}(\tilde{x};p \mid T(\tilde{X})=t) = 0 \Rightarrow$ indep of p ✓
- if $t = \sum_{i=1}^n x_i$ : $P_{\tilde{X}|T}(\tilde{x};p) = \frac{P(X_1=x_1,\dots,X_n=x_n;p \mid T(\tilde{X})=t)}{P(T(\tilde{X})=t)} = \frac{P(X_1=x_1,\dots,X_n=x_n;p)}{P(\sum_{i=1}^n X_i = t)} = \frac{\text{IID Ber}(p)}{\text{Binomial: t success in n indep trial}}$

$$\Rightarrow P_{\tilde{X}|T}(\tilde{x};p) = \frac{p^t(1-p)^{n-t}}{\binom{n}{t}p^t(1-p)^{n-t}} = \frac{1}{\binom{n}{t}} \quad \text{for } t=0,1,\dots,n \Rightarrow \text{indep of } p \text{ for every } \tilde{x}$$ ✓

**Note**:
- A Sufficient statistic can be a vector $\tilde{T}=(T_1,\dots,T_n)$ Ex: $\tilde{T}=(\sum_{i=1}^n X_i, \sum_{i=1}^n (X_i-\overline{X})^2)$
- Trivial Sufficient statistic : $\tilde{T}(\tilde{X})=\tilde{X}$ ($=(X_1,X_2,\dots,X_n)$) → always Sufficient for $\Theta$.
- Let $T$ a sufficient statistic, $g$ a 1-1 function : $T'=g(T)$ is also Sufficient

(def) given $T=t$ indep of $\Theta$ ⇒ $g(T)=g(t)$ indep of $\Theta$ as $g$ is 1-1.
- Sufficient statistics may not be "good" estimators of $\Theta$ (Ex: $\tilde{X}=\text{Ber}(p)$, p unknown : $T=\sum_{i=1}^n X_i$ & $T'=\frac{1}{n}\sum_{i=1}^n X_i$ Sufficient for p but T' is UMVUE, T is bad) + consistant
- Sufficiency is a family property :

T Sufficient for a family of dist ⇏ T sufficient for another family

Ex: $X_1,\dots,X_n$ IID $U(0,\Theta)$, $\Theta$ unknown

$T(\tilde{X})=\overline{X}$ is NOT Sufficient for $\Theta$ : $f_{\tilde{X}|\sum X_i}(x;\Theta \mid \sum X_i = t) = \frac{f_{\tilde{X}}(\tilde{x};\Theta)}{f_{\sum(X_i)}(t)} = \dots$ depends on $\Theta$ !
- If the data $\tilde{x}=(x_1,\dots,x_n)$ is destroyed and $T(\tilde{x})$ sufficient is known : we can generate data consistent with $\tilde{x}$ but we cannot recover $\tilde{x}$

# (IV) Neymann - Fisher Factorization Theorem:

**N-F Thm:** A statistic $T$ is sufficient $\iff$ we can write $f_{\tilde{X}}(\tilde{x};\theta) = g(T(\tilde{x});\theta)\,h(\tilde{x})$

where: • $g$ depends on $\theta$ and on the observations $\tilde{x}$ only through $t$.
• $h$ is a function of the observations $\tilde{x}$ only ($h$ indep of $\theta$!)

**Proof:** (the proof is done in the discrete case: continuous case true but proof more technical!)

For $\tilde{X}$ discrete:

$\Rightarrow$ Let $\tilde{x} = \tilde{x}_0$: $P_{\tilde{X}}(x_0,\theta) = P(\tilde{X}=\tilde{x}_0;\theta) = P(\tilde{X}=x_0, T(\tilde{X})=T(\tilde{x}_0);\theta)$ (this event is contained in that one!) $= \underbrace{P(\tilde{X}=\tilde{x}_0;\theta \mid T(\tilde{X})=T(\tilde{x}_0))}_{h(\tilde{x}_0)\ (\text{indep of }\theta \text{ as } T \text{ sufficient})} \cdot \underbrace{P(T(\tilde{X})=T(\tilde{x}_0);\theta)}_{g(T(\tilde{x}_0),\theta)}$

$\Leftarrow$ By assumption: $P(\tilde{X}=\tilde{x}_0;\theta) = g(T(\tilde{x}_0);\theta)\,h(\tilde{x}_0)$

So: $P(\tilde{X}=\tilde{x}_0;\theta \mid T(\tilde{X})=T(\tilde{x}_0)) = \dfrac{P(\tilde{X}=x_0,\theta)}{P(T(\tilde{X})=T(\tilde{x}_0))} = \dfrac{g(T(\tilde{x}_0);\theta)\,h(\tilde{x}_0)}{\sum_{\tilde{x}:T(\tilde{x})=T(\tilde{x}_0)} P(\tilde{X}=\tilde{x},\theta)} = \dfrac{g(T(\tilde{x}_0);\theta)\,h(\tilde{x}_0)}{\sum_{\tilde{x}:T(\tilde{x})=T(\tilde{x}_0)} g(T(\tilde{x});\theta)\,h(\tilde{x})} = \dfrac{g(T(\tilde{x}_0);\theta)\,h(\tilde{x}_0)}{\sum_{\tilde{x}:T(\tilde{x})=T(\tilde{x}_0)} g(T(\tilde{x});\theta)\sum_{\tilde{x}:T(\tilde{x})=T(\tilde{x}_0)} h(\tilde{x})}$

$\Rightarrow P(\tilde{X}=\tilde{x}_0;\theta \mid T(\tilde{X})=T(\tilde{x}_0)) = \dfrac{h(\tilde{x}_0)}{\sum_{\tilde{x}:T(\tilde{x})=T(\tilde{x}_0)} h(\tilde{x})}$ indep of $\theta$ $\forall \tilde{x}$ $\Rightarrow$ $T$ sufficient for $\theta$!

**Examples:**

(1) $X_1,\dots,X_n$ IID Bernoulli($p$) r.v., $p$ unknown: $\rightsquigarrow T(\tilde{X}) = \sum_{i=1}^n X_i$ Sufficient for $p$.

(pf) $P(X_1=x_1,\dots,X_n=x_n;p) = \underbrace{p^{\sum x_i}(1-p)^{n-\sum x_i}}_{=g(\sum x_i;p)} \cdot \underbrace{\prod_{i=1}^n I[x_i\in\{0,1\}]}_{=h(\tilde{x})}$ $\overset{\text{NF Thm}}{\Rightarrow}$ $T(\tilde{X}) = \sum_{i=1}^n X_i$ Sufficient for $p$. (NF Thm: True $\forall p$)

(2) $X_1,\dots,X_n$ IID $U(0,\theta)$ r.v., $\theta$ unknown: $\rightsquigarrow T(\tilde{X}) = \max\{X_1,\dots,X_n\}$ is Sufficient for $\theta$.

(pf) $f_{\tilde{X}}(\tilde{x};\theta) = \frac{1}{\theta^n}\prod_{i=1}^n I[0<x_i<\theta] = \frac{1}{\theta^n}\prod_{i=1}^n I[x_i>0]\prod_{i=1}^n I[x_i<\theta] = \underbrace{\frac{1}{\theta^n}\cdot I[\max_{i\in\{1,\dots,n\}} x_i<\theta]}_{g(T(\tilde{x});\theta)} \cdot \underbrace{I[\min_{i\in\{1,\dots,n\}} x_i>0]}_{h(\tilde{x})}$ $\forall\tilde{x},\forall\theta$

(3) $X_1,\dots,X_n$ IID Gamma($\alpha,\beta$) r.v.:
- $\alpha=?$, $\beta$ known: $\rightsquigarrow T(\tilde{X}) = \prod_{i=1}^n X_i$ Sufficient for $\alpha$
- $\alpha$ known, $\beta=?$: $\rightsquigarrow T(\tilde{X}) = \sum_{i=1}^n X_i$ Sufficient for $\beta$
- $\alpha,\beta$ unknown: $\rightsquigarrow \tilde{T}(\tilde{X}) = (\prod_{i=1}^n X_i, \sum_{i=1}^n X_i)$ Sufficient for $(\alpha,\beta)$.

(pf) $f_{\tilde{X}}(\tilde{x};\alpha,\beta) = \frac{1}{(\Gamma(\alpha))^n}\cdot\frac{1}{\beta^{n\alpha}}\cdot\prod_{i=1}^n x_i^{\alpha-1}\cdot e^{-\sum x_i/\beta}\cdot\prod_{i=1}^n I[0<x_i<\infty]$ $\Rightarrow$ arrange terms for each case.

Ex: Case 1: $f_{\tilde{X}}(\tilde{x};\alpha,\beta) = \underbrace{\left(\prod_{i=1}^n x_i\right)^{\alpha-1}\cdot\frac{1}{(\Gamma(\alpha))^n}}_{g(\prod x_i,\alpha)} \cdot \underbrace{\frac{1}{\beta^{n\alpha}}e^{-\sum x_i/\beta}\cdot\prod_{i=1}^n I[0<x_i<\infty]}_{h(\tilde{x})}$ for $\beta$ known

**Note:** even if $T$ is sufficient, you still want to keep your data to check your model (IID Gamma, Binom, Bernoulli, Normal, ...)

**Indicator function:** $I[A] = \begin{cases} 1, & x\in A \\ 0, & x\notin A \end{cases}$

# III) Minimal Sufficiency : → Best Summarization of the data $X_1, \dots, X_n$

**minimal Sufficient statistics** = a statistic $T$ is minimal sufficient if

(1) $T$ is sufficient for $\Theta$

(2) $T$ is a function of any other sufficient statistic $T'$ : if $T'$ is sufficient, then $\exists g \; s.t. \; T = g(T')$

**Lemma** = $\tilde{x}, \tilde{y}$ two observations from $f_{\tilde{X}}(\cdot;\theta)$, $T$ a sufficient statistic :

$T(\tilde{x}) = T(\tilde{y}) \implies \dfrac{f_{\tilde{X}}(\tilde{x};\theta)}{f_{\tilde{Y}}(\tilde{y};\theta)}$ is independent of $\Theta$.

**Proof** : NF thm $\Rightarrow \dfrac{f_{\tilde{X}}(\tilde{x};\theta)}{f_{\tilde{Y}}(\tilde{y};\theta)} = \dfrac{g(T(\tilde{x});\theta)\,h(\tilde{x})}{g(T(\tilde{y});\theta)\,h(\tilde{y})} \overset{T(\tilde{x})=T(\tilde{y})}{=} \dfrac{h(\tilde{x})}{h(\tilde{y})}$

**Theorem** = $T$ a statistic : if " $\dfrac{f_{\tilde{X}}(\tilde{x};\theta)}{f_{\tilde{Y}}(\tilde{y};\theta)}$ is independent of $\Theta \iff T(\tilde{x}) = T(\tilde{y})$ " holds, then $T$ is minimal Sufficient for $\Theta$.

**Proof** = Let $\tilde{X}$ be discrete ( $\tilde{X}$ continuous also works but technical! )

- (sufficiency) : $P(\tilde{X} = x_0 \,|\, T(\tilde{X}) = T(\tilde{x}_0)) = \dfrac{P(\tilde{X} = \tilde{x}_0;\theta)}{\sum_{\tilde{x}:T(\tilde{x})=T(\tilde{x}_0)} P(\tilde{X} = \tilde{x};\theta)} = \dfrac{1}{\sum_{\tilde{x}:T(\tilde{x})=T(\tilde{x}_0)} \frac{P(\tilde{X}=\tilde{x};\theta)}{P(\tilde{X}=x_0;\theta)}}$ = indep of $\Theta$.

  ↳ so indep of $\Theta$ (by "$\Leftarrow$" hypothesis)

- (minimality) : We know that $T$ is sufficient; Let $S$ be any other sufficient statistic :

  $S$ is sufficient: $\forall \tilde{w}, \tilde{z} : s(\tilde{w}) = s(\tilde{z}) \xrightarrow{\text{Lemma}} \dfrac{P_{\tilde{X}}(\tilde{w};\theta)}{P_{\tilde{X}}(\tilde{z};\theta)}$ indep of $\Theta \implies T(\tilde{w}) = T(\tilde{z})$ ("$\Leftarrow$" hypoth)

  So $\forall \tilde{w}, \tilde{z} : s(\tilde{w}) = s(\tilde{z}) \Rightarrow T(\tilde{w}) = T(\tilde{z})$, hence $T$ is a function of $S$, so $T$ minimal sufficient.

**Examples** =

① $X_1, \dots, X_n$ IID Poisson($\lambda$) r.v., $\lambda$ unknown : $T(\tilde{X}) = \sum_{i=1}^n X_i$ is minimal sufficient for $\lambda$

(sol) $\dfrac{P(\tilde{X}=\tilde{x};\lambda)}{P(\tilde{X}=\tilde{y};\lambda)} = \dfrac{\lambda^{\sum_{i=1}^n x_i} e^{-\lambda n} / \prod_{i=1}^n x_i!}{\lambda^{\sum_{i=1}^n y_i} e^{-\lambda n} / \prod_{i=1}^n y_i!} = \left(\prod_{i=1}^n \dfrac{x_i!}{y_i!}\right)\left(\lambda^{\sum_{i=1}^n (x_i - y_i)}\right)$ is indep of $\lambda \iff \sum_{i=1}^n x_i = \sum_{i=1}^n y_i$ ✓ So $T = \sum_{i=1}^n X_i$ is min. Suff.

② $X_1, \dots, X_n$ IID $N(\mu, \sigma^2)$, $\mu$ to known, $\sigma^2$ unknown : $T(\tilde{X}) = \sum_{i=1}^n (X_i - \mu)^2$ is min. Suff. for $\sigma^2$

(sol) $\dfrac{f_{\tilde{X}}(\tilde{x};\mu,\sigma^2)}{f_{\tilde{Y}}(\tilde{y};\mu,\sigma^2)} = \dfrac{\left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i-\mu)^2}}{\left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n (y_i-\mu)^2}} = e^{-\frac{1}{2\sigma^2}\left[\sum_{i=1}^n (x_i-\mu)^2 - \sum_{i=1}^n (y_i-\mu)^2\right]}$

is indep of $\sigma^2 \iff \sum_{i=1}^n (x_i-\mu)^2 = \sum_{i=1}^n (y_i-\mu)^2$

**Note** : if $\mu = 0$, $T(\tilde{X}) = \sum_{i=1}^n X_i^2$ is min. Suff for $\sigma^2$.

# Completeness :

→ How to find UMVUEs

## ① Definitions :

**Completeness of a family** = A family of distribution $\{f_{\tilde{X}}(\tilde{x};\theta);\theta\in\Theta\}$ is complete

$\Leftrightarrow$ "$E_\theta[g(\tilde{X})] \overset{\theta}{=} 0 \Rightarrow g(\tilde{X})=0$ with prob 1" holds $\forall$ measurable fonct° $g$.
($\forall\theta\in\Theta$)

**Completeness of a statistic** = A statistic $T$ is complete if the family of distributions induced by $T$ is complete.

i.e. "$E_\theta[g(T(\tilde{X}))] \overset{\theta}{=} 0 \Rightarrow g(T(\tilde{X}))=0$ with prob 1" holds $\forall$ meas. fonct° $g$.

**Examples** =

① $X \sim$ Poisson $(\lambda)$ is a complete family :

(☆) $E_\lambda[g(X)] \overset{\lambda}{=} 0 \Leftrightarrow \sum_{x=1}^{n} g(x)\frac{\lambda^x e^{-\lambda}}{x!} \overset{\lambda}{=} 0 \Leftrightarrow \sum_{x=1}^{n}\frac{g(x)}{x!}\lambda^x \overset{\lambda}{=} 0$ ($\Leftrightarrow$ power series) $\Leftrightarrow g(x)=0\ \forall x=1,2,\dots,n$

but in Poisson : $P(X\notin\mathbb{N})=0$, so $g(x)=0\ \forall x\in\mathbb{R}$ with probability 1 ✓

② $X_1,\dots,X_n$ IID Poisson($\lambda$) r.v. : $T=\sum_{i=1}^{n}X_i$ is complete.

(☆☆) $\sum_{i=1}^{n}X_i \sim$ Poisson$(n\lambda)$, and Poisson family is complete $\Rightarrow$ T is complete !

③ $X_1,\dots,X_n$ IID $U(0,\theta)$ : $T=Y_n$ (max order statistic) is complete.

(☆) $f_{Y_n}(y)=\frac{n y^{n-1}}{\theta^n}\mathbb{I}[0<y<\theta]$, so $E_\theta[g(Y_n)]=\int_0^\theta g(y)\frac{n y^{n-1}}{\theta^n}dy \overset{\theta>0}{=} 0 \Leftrightarrow \int_0^\theta g(y)y^{n-1}dy=0\ \forall\theta>0$

$\Rightarrow \frac{d}{d\theta}(\dots)\overset{\theta>0}{=}0 \Rightarrow g(\theta)\theta^{n-1}\overset{\theta>0}{=}0 \rightsquigarrow g(x)=0\ \forall 0<x<\theta$

$\Rightarrow g(\tilde{X})=0$ with prob 1

**Rao-Blackwell improvement thm** =

Let $\hat{\theta}$ an unbiased estimator for $\theta$, $T$ a sufficient statistic for $\theta$, define $T' := E(\hat{\theta}|T)$

Then : $T'$ is unbiased for $\theta$ ($E_\theta(T')\overset{\theta}{=}\theta$) and $\mathrm{Var}_\theta(T') \le \mathrm{Var}_\theta(\hat{\theta})$

**Proof** = Recall : $E[\hat{\theta}|T] := E[\hat{\theta}|T=t]\big|_{t=T}$ is a r.v., as it varies randomly with T.

- $T'$ is an estimator for $\theta$ (☆) $T$ is sufficient for $\theta$, so $T'$ does not depend on $\theta$.
- $E[T'] = E_T[E_{\hat{\theta}|T}[\hat{\theta}|T]] = E_\theta(\hat{\theta}) = \theta$ (☆) $\hat{\theta}$ unbiased by assumption.
- $\mathrm{Var}(\hat{\theta}) = \underbrace{E[\mathrm{Var}(\hat{\theta}|T)]}_{\ge 0} + \mathrm{Var}[E(\hat{\theta}|T)] \ge \mathrm{Var}[E(\hat{\theta}|T)] = \mathrm{Var}(T')$

**Note** = Even if $\hat{\theta}$ is not unbiased for $\theta$, R-B thm will improve the Mean Square Error of $\hat{\theta}$.

i.e. $MSE(\hat{\theta}) \ge MSE(T')$

(☆) $\hat{\theta}$ given, $T$ sufficient, $T' := E(\hat{\theta}|T)$ :

$$MSE(\hat{\theta}) = E((\hat{\theta}-\theta)^2) = \mathrm{Var}(\hat{\theta}) + (E[\hat{\theta}-\theta])^2 \ge \mathrm{Var}(T') + (E(\hat{\theta})-\theta)^2 = \mathrm{Var}(T') + \left(E_T[E_{\hat{\theta}|T}[\hat{\theta}|T]]-\theta\right)^2$$

$$= \mathrm{Var}(T') + (E[T']-\theta)^2$$

$$= \mathrm{Var}(T') + (E[T'-\theta])^2$$

$$= MSE(T')$$

# IV Lehmann-Scheffé Theorem: → Link between Sufficiency, Completeness & UMVUEs

**L-S thm** = $\hat\theta$ an unbiased estimator for $\theta$, $T$ a complete & sufficient statistic:

If $\exists g$ ST $\hat\theta = g(T)$, then $\hat\theta$ is the unique UMVUE for $\theta$.

**Proof** = • Uniqueness: Suppose $\exists$ another unbiased $g'(T)$: claim: $g(T) = g'(T)$ with prob 1.

(wlog) $E(g(T)) = E(\hat\theta) \overset{\theta}{=} \theta$, $E(g'(T)) = E(\hat\theta) \overset{\theta}{=} \theta$ $\Rightarrow E(g(T)) - E(g'(T)) \overset{\theta}{=} 0 \Rightarrow E[g(T) - g'(T)] \overset{\theta}{=} 0$

$\Rightarrow E[(g-g')(T)] \overset{\theta}{=} 0 \overset{T \text{ is complete}}{\Longrightarrow} (g-g')(T) = 0$ with prob 1 $\Rightarrow g(T) = g'(T)$ with prob 1

• UMVUE: Suppose $\hat\theta'$ is another unbiased estimator for $\theta$ with $\mathrm{Var}(\hat\theta') < \mathrm{Var}(\hat\theta)$

Then improve $\hat\theta'$ by Rao-Blackwellizing it ($T$ sufficient): $g'(T) = E(\hat\theta' | T)$ unbiased

So $\mathrm{Var}(g'(T)) \le \mathrm{Var}(\hat\theta') < \mathrm{Var}(\hat\theta) = \mathrm{Var}(g(T))$

So $\mathrm{Var}(g'(T)) < \mathrm{Var}(g(T))$ ⇒⇐ (wlog) $g(T), g'(T)$ are unbiased estimators of $\theta$, & functions of $T$

So by "claim" above: $g(T) = g'(T)$ with prob 1.

## How to apply the L-S Thm:

① Hope to find an unbiased estimator $\hat\theta$ of $\theta$, ST $\hat\theta = g(T)$ where $T$ is a complete sufficient statistic.

OR

② If you have: • an unbiased estimator $\hat\theta$ of $\theta$
• a complete sufficient statistic $T$

} Rao-Blackwellize $\hat\theta$ by conditioning on $T$: $T' := E[\hat\theta | T]$ is UMVUE!

Tip: take $T$ as simple as you can!

## Examples:

① $X_1, \dots, X_n$ IID Poisson($\lambda$) r.v., $\lambda$ unknown:

$\hat\lambda = \bar X$ is UMVUE for $\lambda$

$\hat\theta = (1-\frac{1}{n})^{\sum_1^n X_i}$ is UMVUE for $P(X=0) = e^{-\lambda}$

(wlog) • $\hat\lambda := \bar X$ is unbiased for $\lambda$, and $T = \sum_{i=1}^n X_i$ is complete + sufficient for $\lambda \overset{\text{L-S thm}}{\Rightarrow} \hat\lambda = g(T) = \frac{1}{n}T$ is UMVUE for $\lambda$.

• Simple unbiased estimator $\hat\theta$ of $e^{-\lambda}$: $\hat\theta := I[X=0]$ (wlog) $E[I[X=0]] = P(X=0) = e^{-\lambda}$.

$T = \sum_{i=1}^n X_i$ is complete + sufficient $\Rightarrow T' = E\left[I[X_1=0] \,\middle|\, \sum_{i=1}^n X_i = t\right]\Big|_{t=T}$ is UMVUE for $e^{-\lambda}$.

$$T' = P\left[X_1=0 \,\middle|\, \sum_{i=1}^n X_i = t\right]\Big|_{t=T} \overset{\text{Bayes's thm}}{=} \frac{P(X_1=0)\cdot P(\sum_{i=1}^n X_i = t \,|\, X_1=0)}{P(\sum_{i=1}^n X_i = t)} \overset{X_1 \perp \sum_2^n X_i}{=} \frac{P(X_1=0)\cdot P(\sum_2^n X_i = t)}{P(\sum_1^n X_i = t)}$$

$$\Rightarrow T' = \frac{e^{-\lambda}\frac{[(n-1)\lambda]^t e^{-(n-1)\lambda}}{t!}}{(n\lambda)^t e^{-n\lambda}/t!}\Big|_{t=T} = \left(\frac{n-1}{n}\right)^t\Big|_{t=T} \Rightarrow T' = (1-\tfrac{1}{n})^T = (1-\tfrac{1}{n})^{\sum_1^n X_i}$$

Note: $(1-\frac{1}{n})^{\sum_1^n X_i} = (1-\frac{1}{n})^{n\bar X} \xrightarrow[n\to\infty]{} (e^{-1})^\lambda$ as $\bar X \xrightarrow[n\to\infty]{} \lambda$ and $(1-\frac{1}{n})^n \xrightarrow[n\to\infty]{} e^{-1}$

② $X_1, \dots, X_n$ IID $U(0,\theta)$ r.v., $\theta$ unknown: $\hat\theta_n := Y_n$ is UMVUE for $\theta$

(wlog) $\hat\theta_n = \frac{n+1}{n}Y_n$ unbiased for $\theta$, $T = Y_n$ is complete & sufficient. $\overset{\text{L-S thm}}{\Longrightarrow} \hat\theta = g(Y_n) = \frac{n+1}{n}Y_n$ is UMVUE for $\theta$.

③ $N(\mu, \sigma^2)$, $\mu$ unknown, $\sigma^2$ known forms a complete family

(wlog) $E_\theta[g(X)] \overset{\mu}{=} 0 \Leftrightarrow \int_{-\infty}^{\infty} g(x)\cdot\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \overset{\mu}{=} 0 \Leftrightarrow \int_{-\infty}^{\infty} g(x) e^{-\frac{x^2 - 2\mu x}{2\sigma^2}} dx \overset{\mu}{=} 0 \Leftrightarrow \int_{-\infty}^{\infty} \underbrace{\left(g(x) e^{-\frac{x^2}{2\sigma^2}}\right)}_{:=h(x)} e^{-\frac{\mu}{\sigma^2}x} dx \overset{\mu}{=} 0$

$\Leftrightarrow \int_{-\infty}^{\infty} h(x) e^{-\frac{\mu}{\sigma^2}x} dx \overset{\mu}{=} 0 \Rightarrow h(x) = 0 \ \forall x \in \mathbb{R}$ (↑ Laplace Transform) $\Rightarrow g(x) = 0 \ \forall x \in \mathbb{R}$

$\Rightarrow g(X) = 0$ with prob 1.

**Aside** = German Tanks during WW II:

How many tanks are produced? → 1, 2, 3, …, $N$: estimate $N$!

$Y_n$ is sufficient for $N$ → $\boxed{\hat N = \frac{n+1}{n}Y_n - 1}$ UMVUE for $N$ → discrete case!

# Terminology of "completeness":

**Def**: We say that a statistic $T$ is ancillary for $\theta$ if the distribution of $T$ does not depend on $\theta$.

**Def**: We say that a statistic $T$ is 1st order ancillary for $\theta$ if $E(T)$ does not depend on $\theta$.

It might seem that an ancillary statistic is useless when making inference about $\theta$.
However, the reason for the term ancillary (servant, helper, assistant) is conveyed in the following example:

**Ex**: Toss a fair coin once: if the outcome is Heads (with prob $\frac{1}{2}$) then draw a sample of size $n$ from a $N(\mu, 1)$ distribution.
If the outcome is Tails: draw a sample of size $n$ from a $N(\mu, 10)$ distrib.

$X :=$ r.v. defining the outcome of the coin toss:

So $P(X=1) = \frac{1}{2} = P(X=0)$

If we are interested in $\mu$, then the statistic $T = X$ is ancillary for $\mu$.
(esp) $P(X=x)$ does NOT depend on $\mu$ $\quad x \in \{0,1\}$

However, surely the information that we get from $X$, does enhance our inference about $\mu$.

By conditioning on the observed values of $X$, we gain info about the distribution from which our 2nd sample is drawn

Thus $\mathrm{Var}(\bar{Y}) = \hat{\rho}$ will change according to whether $X=1$ or $X=0$

A statistic $T$ is 1st order ancillary $\Rightarrow E(T) \overset{\theta}{\equiv} c$ (cst indep of $\theta$) $\Rightarrow E(T-c) \overset{\theta}{\equiv} 0$

➡ completeness says that a family is complete if the only ancillary statistics are the Trivial ones!

Method of estimation:

# Method of Moments:

→ Given a parameter $\theta$ to estimate: How do we derive a suitable estimator?

## I The Method:

**Setup:**
- $X_1, \dots, X_n$ IID r.v. whose distribution depends on $\theta_1, \theta_2, \dots, \theta_r$.
- ALL $k$-moments exist for $1 \le k \le r$: i.e. $E[|X^r|] < \infty$.
- Let $\mu_k := E(X^k)$, $1 \le k \le r$, be the $k^{th}$ population moment.
- Let $m_k := \frac{1}{n}\sum_{i=1}^n X_i^k$, $1 \le k \le r$, be the $k^{th}$ sample moment.

**WLLN:** $\underbrace{\frac{1}{n}\sum_{i=1}^n X_i^k}_{m_k} \xrightarrow[n\to\infty]{} \underbrace{E[X^k]}_{\mu_k} \Rightarrow m_k \xrightarrow[n\to\infty]{} \mu_k$

**Method:** Set $m_k \simeq \mu_k$, and solve: $\begin{cases} m_1 = h_1(\theta_1, \dots, \theta_r) \\ \vdots \\ m_r = h_r(\theta_1, \dots, \theta_r) \end{cases}$ for $\hat\theta_1, \dots, \hat\theta_r$.
(with $m_k = \frac{1}{n}\sum_{i=1}^n X_i^k$)

**Result:** You obtain: $\begin{cases} \hat\theta_1 = g_1(m_1, \dots, m_r) \\ \vdots \\ \hat\theta_r = g_r(m_1, \dots, m_r) \end{cases}$ } useful when $g_k$'s are continuous

**Note:**
- The MOM is easy to apply (MOM estimators are often used as starting values in iterative methods) Ex: MLE's
- MOM's are not always unbiased.
- MOM's are consistent estimators! $\hat\theta_k \xrightarrow[n\to\infty]{P} \theta_k$, $\forall 1 \le k \le r$

(oo) When $g_k$ are continuous:

$m_k = \frac{1}{n}\sum_{i=1}^n X_i^k \xrightarrow[n\to\infty]{P} E[X^k] = \mu_k$ $\overset{g \text{ continuous}}{\Longrightarrow}$ $\underbrace{g_k(m_1, \dots, m_r)}_{:=\hat\theta_k} \xrightarrow[n\to\infty]{P} g_k(\mu_1, \dots, \mu_r) = \theta_k$ $\forall 1 \le k \le r$

## II Examples:

**Ex 1:** $X_1, \dots, X_n$ IID $\mathcal{N}(\mu, \sigma^2)$, $\mu$ & $\sigma^2$ unknown

**Solution:** $\hat\mu = \frac{1}{n}\sum_{i=1}^n X_i$, $\hat\sigma^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \bar X)^2$ → BIASED!

(oo) $\theta = (\theta_1, \theta_2) = (\mu, \sigma^2)$ unknown: $\begin{cases} m_1 \overset{!}{=} \mu_1 \\ m_2 \overset{!}{=} \mu_2 \end{cases} \Rightarrow \begin{cases} \frac{1}{n}\sum_{i=1}^n X_i = \mu_1 = \mu \\ \frac{1}{n}\sum_{i=1}^n X_i^2 = \mu_2 = \sigma^2 + \mu^2 \end{cases}$

So: $\begin{cases} \mu = \frac{1}{n}\sum_{i=1}^n X_i = \bar X \\ \sigma^2 = \left(\frac{1}{n}\sum_{i=1}^n X_i^2\right) - \bar X^2 = \frac{1}{n}\left[\left(\sum_{i=1}^n X_i^2\right) - n\bar X^2\right] = \frac{1}{n}\sum_{i=1}^n (X_i^2 - \bar X^2) = \frac{1}{n}\sum_{i=1}^n (X_i - \bar X)^2 \end{cases}$

**Ex 2:** $X_1, \dots, X_n$ IID Bern(p), p unknown:

**Solution:** $\hat p = \frac{1}{\bar X}$, where $\bar X$ = trial # of the 1st success

(oo) $X \sim \text{Geom}(p) \Rightarrow m_1 \overset{!}{=} \mu_1 = E(X) = \frac{1}{p} \Rightarrow \hat p = \frac{1}{m_1} = \frac{1}{\bar X}$

OR: $\hat p = \frac{1}{n}\sum_{i=1}^n X_i$

(oo) $P[X_i = 0] = 1-p$, $P[X_i = 1] = p$

So $\begin{cases} m_1 = \frac{1}{n}\sum_{i=1}^n X_i = \bar X \\ \mu_1 = E[X_i] = p \end{cases} \Rightarrow \begin{cases} \mu_1 \overset{!}{=} \bar X \\ \Rightarrow \hat p = \bar X \end{cases}$

**Note:** ∄ a reasonable unbiased estimator for p:

Suppose $T(X)$ is an estimator for p:

Unbiasedness $\Rightarrow E[T(X)] = \sum_{x=1}^{\infty} T(x) \underbrace{p(1-p)^{x-1}}_{\text{polynomial in } p} \overset{!}{=} p$ $\forall p$ $\Rightarrow \begin{cases} T(1) = 1 \\ T(x) = 0 \ \forall x \ge 2 \end{cases}$ ⇒ crazy estimator!

# Method of Maximum Likelihood :

## (I) The Method :

**Likelihood :**
- Discrete : $\ell(\theta ; \tilde{x}) := P_{\tilde{X}}(\tilde{x} ; \theta)$
- Continuous : $\ell(\theta ; \tilde{x}) := f_{\tilde{X}}(\tilde{x} ; \theta)$

} likelihood of $\theta$ at the observation $\tilde{x}$
$\Rightarrow \ell$ = a function of $\theta$
(but $\ell$ is a different function at each $\tilde{x}$)

**MLE :** $\hat{\Theta} = \hat{\Theta}(\tilde{X})$ is a MLE of $\theta$ if :

$$\hat{\Theta}(\tilde{x}) = \arg \sup_{\theta \in \Theta} \ell(\theta ; \tilde{x}) \quad , \forall \tilde{x} \text{ ST } f_{\tilde{X}}(\tilde{x} ; \theta) > 0$$

**Prop :** $\hat{\Theta} = \hat{\Theta}(\tilde{X})$ is a MLE $\iff \sup_{\theta \in \Theta} \ell(\theta ; \tilde{x}) = \ell(\hat{\Theta}(\tilde{x}), \tilde{x}) \quad \forall \tilde{x}$ ST $f_{\tilde{X}}(\tilde{x} ; \theta) > 0$

**Note :**
- The MLE may not exist
- The MLE is not always unique (but it often is !)
- $\hat{\Theta}(\tilde{X})$ is the maximum likelihood estimator ; $\hat{\Theta}(\tilde{x})$ is the maximum likelihood estimate

**Trick :** To find $\hat{\theta}$, maximize $\text{Log}(\ell(\theta ; \tilde{x}))$ WRT $\theta$.

**Method :**
- Continuous case = If $\ell$ is differentiable, use calculus :
  - → Solve the "likelihood equations" : $\frac{\partial}{\partial \theta_i} \text{Log}(\ell(\tilde{\theta} ; \tilde{x})) \overset{!}{=} 0$, for $\tilde{\theta} = (\theta_1, \dots, \theta_r)$
  - → "Score vector" : $\left( \frac{\partial}{\partial \theta_1} \text{Log}(\ell(\tilde{\theta} ; \tilde{x})), \dots, \frac{\partial}{\partial \theta_r} \text{Log}(\ell(\tilde{\theta} ; \tilde{x})) \right)$
- discrete case = harder !

**Ex 1 :** $X_1, \dots, X_n$ IID $\mathcal{N}(\mu, \sigma^2)$ r.v. with $\mu, \sigma^2$ unknown $\Rightarrow$ $\begin{cases} \hat{\mu} = \overline{X} \\ \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \overline{X})^2 \end{cases}$

$$(\ell) \quad \text{Log}(\ell(\mu, \sigma^2 ; \tilde{x})) = \text{Log}\left( \frac{1}{(2\pi)^{n/2}} \cdot \frac{1}{(\sigma^2)^{n/2}} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2} \right) = (\text{CST}) - \frac{n}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2$$

- $\frac{\partial}{\partial \mu} \text{Log}(\ell) \overset{!}{=} 0 \Rightarrow \sum_{i=1}^n (x_i - \mu) = 0 \rightarrow \mu = \frac{1}{n} \sum_{i=1}^n x_i \quad \forall x_i \Rightarrow \hat{\mu} = \frac{1}{n} \sum_{i=1}^n X_i = \overline{X}$
- $\frac{\partial}{\partial \sigma^2} \text{Log}(\ell) \overset{!}{=} 0 \Rightarrow -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2} \sum_{i=1}^n (x_i - \mu)^2 = 0 \Rightarrow \sigma^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu) \Rightarrow \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \underbrace{\hat{\mu}}_{= \overline{X}})^2$

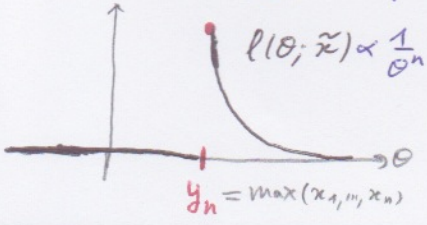
**Ex 2 :** $X_1, \dots, X_n$ IID Poisson($\lambda$) r.v. with $\lambda$ unknown : $\hat{\lambda} = \overline{X}$

$$(\ell) \quad \text{Log}(\ell(\lambda ; \tilde{x})) = \text{Log}\left( \frac{\lambda^{\sum_{i=1}^n x_i} e^{-n\lambda}}{\prod_{i=1}^n x_i !} \right) = \left( \sum_{i=1}^n x_i \right) \log \lambda - n\lambda - (\text{CST})$$

$$\Rightarrow \frac{\partial}{\partial \lambda} \text{Log}(\ell) \overset{!}{=} 0 \Rightarrow \frac{1}{\lambda} \sum_{i=1}^n x_i - n = 0 \Rightarrow \lambda = \frac{1}{n} \sum_{i=1}^n x_i \quad \forall x_i \rightarrow \hat{\lambda} = \frac{1}{n} \sum_{i=1}^n X_i = \overline{X}$$

**Ex 3 :** ⚠ $X_1, \dots, X_n$ IID $U(0, \theta)$ r.v. with $\theta$ unknown : $\hat{\theta} = Y_n$

→ Cannot set up the likelihood equations : $\ell$ is NOT differentiable at the region of interest)




$$\ell(\theta ; \tilde{x}) = \frac{1}{\theta^n} \cdot \mathbb{I}[0 < y_n < \theta] \rightarrow \text{maximized at } \theta = y_n$$

## II Invariance Property of MLEs:

**Thm:** Let $\theta$ an unknown parameter with MLE $\hat{\theta}$:
The MLE of $\eta = \tau(\theta)$ (↳ some f° of $\theta$) is: $\hat{\eta} = \tau(\hat{\theta})$

$\Rightarrow \boxed{\widehat{\tau(\theta)} = \tau(\hat{\theta})}$

**Ex:**
- $X_1, \dots, X_n \sim Exp(\beta)$: let $\theta = P[X > x] = e^{-x/\beta}$
  $\hat{\beta} = \bar{X} \Rightarrow \hat{\theta} = e^{-x/\hat{\beta}} \Rightarrow \widehat{P[X>x]} = e^{-x/\bar{X}}$
- $X_1, \dots, X_n \sim \mathcal{N}(\mu, \sigma^2)$: let $\theta = \frac{\mu^2 + \sin\mu}{\log(\sigma^{-2})} \Rightarrow \hat{\theta} = \frac{\hat{\mu}^2 + \sin(\hat{\mu})}{\log(\hat{\sigma}^{-2})}$

**Proof:** **Case 1:** $\tau(\theta)$ is 1-1 $\rightsquigarrow \tau^{-1}$ exists!
Let $\ell^*(\eta; \tilde{x}) :=$ likelihood of $\eta$ at $\tilde{x}$. and $\tau$ is 1-1 $\Rightarrow \tau(\theta) = \eta \Leftrightarrow \tau^{-1}(\eta) = \theta$
So $\ell^*(\eta; \tilde{x}) = \ell^*(\tau(\theta); \tilde{x}) = \ell(\theta; \tilde{x}) = \ell(\tau^{-1}(\eta); \tilde{x})$
$\Rightarrow \ell^*(\hat{\eta}; \tilde{x}) := \sup_{\eta}[\ell^*(\eta; \tilde{x})] = \sup_{\theta}[\ell(\theta; \tilde{x})] = \ell(\hat{\theta}; \tilde{x}) = \ell(\tau^{-1}(\tau(\hat{\theta})); \tilde{x}) = \ell^*(\tau(\hat{\theta}); \tilde{x})$
$\Rightarrow \hat{\eta} = \tau(\hat{\theta}) \Rightarrow \widehat{\tau(\theta)} = \tau(\hat{\theta})$

**Case 2:** $\tau(\theta)$ is not 1-1:
Induced likelihood: $\ell^*(\eta; \tilde{x}) := \sup_{\theta: \tau(\theta)=\eta} \ell(\theta; \tilde{x})$
So $\ell^*(\hat{\eta}; \tilde{x}) := \sup_{\eta} \ell^*(\eta; \tilde{x}) = \sup_{\eta} \sup_{\theta:\tau(\theta)=\eta} \ell(\theta; \tilde{x}) = \sup_{\theta} \ell(\theta; \tilde{x}) = \ell(\hat{\theta}; \tilde{x}) = \sup_{\theta: \tau(\theta)=\tau(\hat{\theta})} \ell(\theta; \tilde{x}) = \ell^*(\tau(\hat{\theta}); \tilde{x})$ (iterated Supremum)
$\Rightarrow \ell^*(\hat{\eta}; \tilde{x}) = \ell^*(\tau(\hat{\theta}); \tilde{x}) \Rightarrow \widehat{\tau(\theta)} = \tau(\hat{\theta})$.

## III Efficiency of MLEs:

**Asymptotic Unbiasedness** = $\hat{\theta}$ asympt. unbiased for $\theta \Leftrightarrow \exists K_n \in \mathbb{R}$ s.t. "$K_n \cdot (\hat{\theta}_n - \theta) \xrightarrow[n\to\infty]{D} W \Rightarrow E(W) = 0$"

⚠ this says something about the limiting distrib° of $\hat{\theta}_n$ ($\neq$ unbiased as $n \to \infty$ : $E(\hat{\theta}_n) \xrightarrow[n\to\infty]{} 0$)

**Thm:** $X_1, \dots, X_n$ IID r.v. with unknown scalar parameter $\theta$
$\hat{\theta}_n$ = MLE of $\theta_0$ for each $n$:
**Then:** (need regularity cond°) $\boxed{\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow[n\to\infty]{D} W \sim \mathcal{N}\left(0, \frac{1}{I(\theta_0)}\right)}$, $I(\theta_0) = Var\left[\frac{\partial}{\partial\theta} \log(\theta, X)\right]\Big|_{\theta=\theta_0}$
↳ CR Lower bound ↳ generic $X_i$

**Corr:** $\hat{\theta}_n$ is asymptotically unbiased

**Note** = MLEs are asymptotically efficient: $Var(\hat{\theta}_n) \simeq \frac{1}{n \cdot I(\theta_0)}$ → CR Lower bound! MLEs have the lowest large sample variance than any other estimator!

**Proof** = $\begin{cases} L(\theta; \tilde{X}) := \log(\ell(\theta, \tilde{X})) \\ L'(\theta; \tilde{X}) := \frac{\partial}{\partial\theta} \log(\dots) \\ L''(\theta; \tilde{X}) := \frac{\partial^2}{\partial\theta^2} \log(\dots) \end{cases}$

$\hat{\theta}_n \xrightarrow[n\to\infty]{P} \theta \Rightarrow \underbrace{L'(\hat{\theta}; \tilde{X})}_{=0 \text{ as } \hat{\theta} \text{ is MLE}} \simeq L'(\theta; \tilde{X}) + (\hat{\theta} - \theta) \cdot L''(\theta; \tilde{X})$

So $L'(\theta, \tilde{X}) = -(\hat{\theta} - \theta) \cdot L''(\theta; \tilde{X}) \Rightarrow \sum_{i=1}^n L'(\theta; X_i) = -(\hat{\theta} - \theta) \cdot \sum_{i=1}^n L''(\theta; \tilde{X})$

($\because \tilde{X} \sim IID \Rightarrow \ell \to \prod \ell_i \to L \to \sum L_i$)

And: • $E_\theta(L'(\theta; X_i)) = \frac{\partial}{\partial\theta} \underbrace{\int \dots dx}_{=1} = 0 \quad \forall i$
• $Var_\theta(L'(\theta; X_i)) = Var_\theta\left[\frac{\partial}{\partial\theta} \log(\theta; X_i)\right] \quad \forall i$

$\Rightarrow \frac{\sum_{i=1}^n L'(\theta; X_i) - E(L'(\theta; X_i))}{\sqrt{Var_\theta\left(\sum_{i=1}^n L'(\theta; X_i)\right)}} = \frac{-(\hat{\theta} - \theta) \cdot \sum_{i=1}^n L''(\theta; \tilde{X}) - 0}{\sqrt{n \cdot Var_\theta[L'(\theta; X_i)]}} \xrightarrow{D} \mathcal{N}(0, 1)$

Also: $-\frac{1}{n} \sum_{i=1}^n \underbrace{L''(\theta; X_i)}_{IID} \xrightarrow[n\to\infty]{P} -E_\theta[L''(\theta; X_i)] = +E_\theta[L'(\theta; X_i)^2] = Var_\theta[L'(\theta; X_i)]$

$\Rightarrow \sqrt{n}(\hat{\theta}_n - \theta) \cdot \sqrt{Var_\theta[L'(\theta; X_i)]} \xrightarrow[n\to\infty]{D} \mathcal{N}(0, 1) \Rightarrow \sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow[n\to\infty]{D} \mathcal{N}\left(0, \frac{1}{Var[L'(\theta; X_i)]}\right)$

(multiply top & bottom by $\frac{1}{n}$)

# IV) Consistency of MLEs:

**Consistency:** $X_1, \dots, X_n$ r.v. with parameter $\theta$:
An Estimator $\hat{\theta}_n$ is consistent for $\theta$ if: $\hat{\theta}_n \xrightarrow[n\to\infty]{P} \theta \quad \forall \theta \in \Theta$

**Ex:** $X_1, \dots, X_n \sim \mathcal{N}(\mu, \sigma^2)$ with $(\mu, \sigma^2)$ unknown: $\begin{cases} \hat{\mu}_n = \frac{1}{n}\sum_{i=1}^n X_i \\ \hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2 \end{cases}$ are consistent

(@) WLLN: $\hat{\mu}_n \xrightarrow[n\to\infty]{P} \mu$ and $\hat{\sigma}_n^2 \xrightarrow[n\to\infty]{P} \sigma^2$

**Assumptions** - To prove that $\hat{\theta}_n$ is consistent when $X_1, \dots, X_n$ are IID and $\Theta$ is finite, need:

- A1: The family of distributions $\{P_\theta : \theta \in \Theta\}$ is identifiable: $\theta \neq \theta' \Rightarrow P_\theta \neq P_{\theta'}$
- A2: The support of $P_\theta$ does not depend on $\theta$.
- A3: $X_i$'s are IID
- A4: $\Theta$ contains an open interval ST $\theta_0$ is an interior point of this interval.

**Lemma** = $X_1, \dots, X_n$ IID r.v. with likelihood $\ell(\theta; \tilde{x})$; Let $\theta_0$ be the "true" value of $\theta$.
Then: $P_{\theta_0}\left(\ell(\theta_0; \tilde{X}) > \ell(\theta; \tilde{X})\right) \xrightarrow[n\to\infty]{} 1 \quad \forall \theta \in \Theta$

**Note** = Jensen's inequality: $\begin{cases} E[g(X)] \geq g(E(X)), & \text{if } g \text{ is convex} \\ E[g(X)] \leq g(E(X)), & \text{if } g \text{ is concave} \end{cases}$ $\langle x^2 \rangle \geq \langle x \rangle^2$
↳ strict if strictly convex/concave

**Proof** = $P_{\theta_0}(\ell(\theta_0; \tilde{X}) > \ell(\theta; \tilde{X})) \overset{\text{indep}}{=} P_{\theta_0}\left(\prod_{i=1}^n \ell(\theta_0; X_i) > \prod_{i=1}^n \ell(\theta; X_i)\right) = P_{\theta_0}\left(\prod_{i=1}^n \frac{\ell(\theta; X_i)}{\ell(\theta_0; X_i)} < 1\right)$

So $P_{\theta_0}(\ell(\theta_0; \tilde{X}) > \ell(\theta; \tilde{X})) = P_{\theta_0}\left(\frac{1}{n}\sum_{i=1}^n \log\left(\frac{\ell(\theta; X_i)}{\ell(\theta_0; X_i)}\right) < 0\right)$

- Let $Y_n := \frac{1}{n}\sum_{i=1}^n \underbrace{\log\left(\frac{\ell(\theta; X_i)}{\ell(\theta_0; X_i)}\right)}_{\text{IID}} \Rightarrow \text{WLLN: } Y_n \xrightarrow[n\to\infty]{P} E_{\theta_0}\left[\log\left(\frac{\ell(\theta; X_i)}{\ell(\theta_0; X_i)}\right)\right]$ (*)
- Log is strictly concave $\Rightarrow$ Jensen: $E_{\theta_0}\left[\log\left(\frac{\ell(\theta; X_i)}{\ell(\theta_0; X_i)}\right)\right] < \log\left[E_{\theta_0}\left(\frac{\ell(\theta; X_i)}{\ell(\theta_0; X_i)}\right)\right] = \log\left[\int \frac{\ell(\theta; x)}{\ell(\theta_0; x)} \ell(x; \theta_0)\, dx\right]$
$= \log\left[\int \ell(\theta; x)\, dx\right] = \log[1] = 0$

So: $E_{\theta_0}\left[\log\left(\frac{\ell(\theta; X_i)}{\ell(\theta_0; X_i)}\right)\right] < 0, \quad \forall i \leq n$

- So $\forall \varepsilon > 0$: $P_{\theta_0}\left[|Y_n - E_{\theta_0}[\cdots]| \leq \varepsilon\right] \xrightarrow[n\to\infty]{} 1$ by (*)
$\Rightarrow P_{\theta_0}\left[Y_n - E_{\theta_0}[\cdots] \leq \varepsilon\right] \xrightarrow[n\to\infty]{} 1$
$\Rightarrow P_{\theta_0}\left[Y_n \leq \varepsilon + \underbrace{E_{\theta_0}[\cdots]}_{<0}\right] \xrightarrow[n\to\infty]{} 1 \quad \forall \varepsilon > 0$

Pick $\varepsilon$ small enough ST $\varepsilon + E_{\theta_0}[\cdots] < 0$
then: $P_{\theta_0}[Y_n < 0] \xrightarrow[n\to\infty]{} 1$

So $P_{\theta_0}\left[\ell(\theta_0; \tilde{X}) > \ell(\theta; \tilde{X})\right] \xrightarrow[n\to\infty]{} 1 \quad \forall \theta \in \Theta$

**Thm:** Let $\Theta = \{\theta_1, \dots, \theta_n\}$ be a finite discrete parameter space.

Then under $A_1, \dots, A_4$: (1) The MLE $\hat{\theta}_n$ for $\theta$ is consistent $\iff P_\theta[\hat{\theta}_n = \theta] \xrightarrow[n\to\infty]{} 1, \ \forall \theta \in \Theta$

(2) The MLE $\hat{\theta}_n$ is consistent for $\theta$: $\hat{\theta}_n \xrightarrow[n\to\infty]{P} \theta, \ \forall \theta \in \Theta$

**Proof:** (1) $P_\theta(|\hat{\theta}_n - \theta| < \varepsilon) \xrightarrow[n\to\infty]{} 1 \ \forall \varepsilon > 0 \iff P_\theta(\hat{\theta}_n = \theta) \xrightarrow[n\to\infty]{} 1$

$\Rightarrow$ $\Theta = \{\theta_1, \dots, \theta_n\}$ : WLOG $\theta_1 < \theta_2 < \dots < \theta_n$. Since $\theta \in \Theta$, $\exists i$ ST $\theta = \theta_i$ $\rightarrow$ Pick $\varepsilon < \theta_{i+1} - \theta_{i-1}$

So $P_\theta(\underbrace{\theta - \varepsilon < \hat{\theta}_n < \theta + \varepsilon}) = P_\theta(|\hat{\theta}_n - \theta| < \varepsilon) \xrightarrow[n\to\infty]{} 1$

$= P_\theta(\hat{\theta}_n = \theta)$

So $\forall \varepsilon > 0$, $\exists \varepsilon_n > 0$ ST $P_\theta(\hat{\theta}_n = \theta) = P_\theta(|\hat{\theta}_n - \theta| < \varepsilon_n) \Rightarrow P_\theta(\hat{\theta}_n = \theta) \xrightarrow[n\to\infty]{} 1$

$\Leftarrow$ $\{\hat{\theta}_n : |\hat{\theta}_n - \theta| < \varepsilon\} \supset \{\hat{\theta}_n : \hat{\theta}_n = \theta\} \Rightarrow 1 \geq P_\theta(|\hat{\theta}_n - \theta| < \varepsilon) \geq P_\theta(\hat{\theta}_n = \theta) \xrightarrow[n\to\infty]{} 1$

$\rightarrow P_\theta(|\hat{\theta}_n - \theta| < \varepsilon) \xrightarrow[n\to\infty]{} 1$ by Squeeze Thm.

(2) Particular case: $\Theta = \{\theta_1, \theta_2\}$

Enough to show that $P_{\theta_1}[\hat{\theta}_n = \theta_1] \xrightarrow[n\to\infty]{} 1$ and $P[\hat{\theta}_n = \theta_2] \xrightarrow[n\to\infty]{} 1$.

$$P_{\theta_1}[\hat{\theta}_n = \theta_1] = P_{\theta_1}[\hat{\theta}_n = \theta_1 \mid \ell(\theta_1; \tilde{X}) > \ell(\theta_2; \tilde{X})] \cdot P[\ell(\theta_1; \tilde{X}) > \ell(\theta_2 > \tilde{X})]$$
$$+ P_{\theta_1}[\hat{\theta}_n = \theta_1 \mid \ell(\theta_1; \tilde{X}) < \ell(\theta_2; \tilde{X})] \cdot P[\ell(\theta_1; \tilde{X}) < \ell(\theta_2 > \tilde{X})]$$

But:

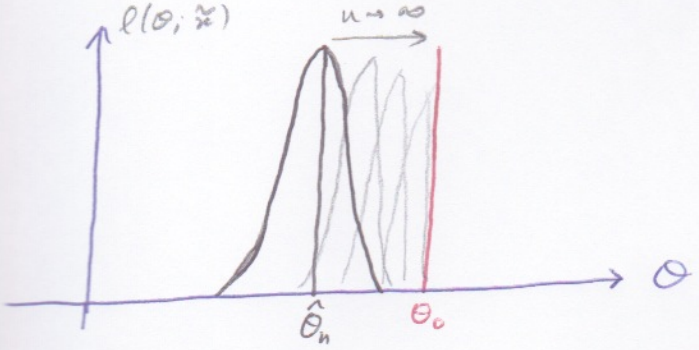
- $P_{\theta_1}[\hat{\theta}_n = \theta_1 \mid \ell(\theta_1; \tilde{X}) > \ell(\theta_2; \tilde{X})] = P_{\theta_1}[\hat{\theta}_n = \theta_1 \mid \hat{\theta}_n = \theta_1] = 1$
  ↳ $\hat{\theta}_n = \theta_1$ as MLE
- $P_{\theta_1}[\hat{\theta}_n = \theta_1 \mid \ell(\theta_1; \tilde{X}) < \ell(\theta_2; \tilde{X})] = P_{\theta_1}[\hat{\theta}_n = \theta_1 \mid \hat{\theta}_n = \theta_2] = 0$ as $\theta_1 \neq \theta_2$.
  ↳ $\hat{\theta}_n = \theta_2$ as MLE
- $P_{\theta_1}[\ell(\theta_1; \tilde{X}) > \ell(\theta_2; \tilde{X})] \xrightarrow[n\to\infty]{} 1$ by the previous lemma.

So $P_{\theta_1}[\hat{\theta}_n = \theta_1] = 1 \cdot P_{\theta_1}[\ell(\theta_1; \tilde{X}) > \ell(\theta_2; \tilde{X})] + 0 \xrightarrow[n\to\infty]{} 1$

(Similar for $P_{\theta_2}[\hat{\theta}_n = \theta_2]$) ∎

**Thm:** If $\ell(\theta; \tilde{X})$ is differentiable, then:

with probability 1, $\exists$ a consistent solution to the likelihood equations $\frac{\partial \text{Log}(\ell(\theta))}{\partial \theta} = 0$



# Interval Estimation: → Confidence Intervals!

## ① General Picture:

$\tilde{X} = (X_1, \dots, X_n)$ a random vector, $\theta$ an unknown parameter:

- $L = L(\tilde{X})$ and $R = R(\tilde{X})$: left & Right random endpoints of the interval $(L(\tilde{X}), R(\tilde{X}))$
- $(L(\tilde{X}), R(\tilde{X}))$: random interval.

Goal = Find L & R, ST: given $\alpha$, $\boxed{P_\theta(L < \theta < R) = 1 - \alpha}$ $\forall \theta \in \Theta$

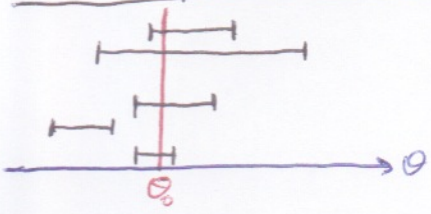
Note: • $(L(\tilde{X}), R(\tilde{X}))$ is really a probability interval: $\theta \in (L(\tilde{X}), R(\tilde{X}))$ with probability $(1-\alpha)$!

• After the experiment, $\tilde{x} = (x_1, \dots, x_n)$ have fixed values, so $(L(\tilde{x}), R(\tilde{x}))$ is a fixed interval.
↳ $\theta \in (L(\tilde{x}), R(\tilde{x}))$ = 0 or 1!

Coverage probability = $(1-\alpha)$ → usually: $(1-\alpha) \sim 0{,}95$
(confidence level / coeff)

$100(1-\alpha)\%$ Confidence interval for $\theta$ = $(L(\tilde{x}), R(\tilde{x}))$ (FIXED!) as well as $(L(\tilde{X}), R(\tilde{X}))$ (RANDOM!)

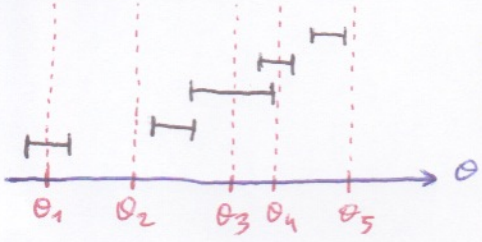
**Weak interpretation of C.I.:** → Repeat the experiment for the Same $\theta$




If we repeat the experiment many times with the Same unknown $\theta$, and compute the interval $(L(\tilde{x}), R(\tilde{x}))$ each time, then:

"$100(1-\alpha)\%$ of the times, roughly, we expect $\theta \in (L(\tilde{x}), R(\tilde{x}))$".

**Strong interpretation of C.I.:** → Repeat the experiment for possibly different $\theta$'s.




If we use the random interval in a large nb of experiments, with possibly different $\theta$'s, then:

"$100(1-\alpha)\%$ of the times, roughly, we expect $(L(\tilde{x}), R(\tilde{x}))$ to capture its respective $\theta$".

Reason: Let $Y_j = \begin{cases} 1, & \text{if } \theta_j \in I_j \\ 0, & \text{if } \theta_j \notin I_j \end{cases}$ where $I_j = j^{th}$ confidence interval

- The $j$ trials are independent $\Rightarrow$ $Y_j$ are IID Bernoulli(p) r.v.
- $p = P_{\theta_j}[\theta_j \in I_j] = 1 - \alpha$ $\forall \theta_j$
- So $E[Y_j] = p = 1 - \alpha$ $\Rightarrow$ WLLN: $\underbrace{\frac{1}{n}\sum_{j=1}^{n} Y_j}_{\text{Proportion of Intervals that capture their } \theta_j\text{'s.}} \xrightarrow[n\to\infty]{P} E[Y_j] = 1 - \alpha$

⚠ L & R should NOT depend on unknown parameters!

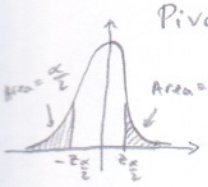
# (IV) The method of pivots:

**Pivot** = a random variable of the form $g(\tilde{X};\theta)$ whose distribution does **not** depend on $\theta$.
For each $\tilde{x}$, $g(\tilde{x},\theta)$ is monotone in $\theta$.

**Setting 1:** $X_1,\dots,X_n$ IID $\mathcal{N}(\mu,\sigma^2)$ r.v., $\sigma^2$ **known** $\Rightarrow$ $\begin{cases} L = \bar{X} - z_{\frac{\alpha}{2}}\,\sigma/\sqrt{n} \\ R = \bar{X} + z_{\frac{\alpha}{2}}\,\sigma/\sqrt{n} \end{cases}$ for $\mu$

**Solution:** Goal: find $L \& R$ s.t. $P_\mu(L<\mu<R) = 1-\alpha \ \ \forall \mu$

Pivot: $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1)$ as $X_i \sim \mathcal{N}(\mu,\sigma^2)$, so: $P(-z_{\frac{\alpha}{2}} < Z < z_{\frac{\alpha}{2}}) = 1-\alpha$ (where $P(Z>z_{\frac{\alpha}{2}}) = \frac{\alpha}{2}$), where $Z \sim \mathcal{N}(0,1)$




So: $P_\mu\left(\frac{\bar{X}-R}{\sigma/\sqrt{n}} < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < \frac{\bar{X}-L}{\sigma/\sqrt{n}}\right) = 1-\alpha \Rightarrow \begin{cases} \frac{\bar{X}-R}{\sigma/\sqrt{n}} = -z_{\alpha/2} \\ \frac{\bar{X}-L}{\sigma/\sqrt{n}} = z_{\alpha/2} \end{cases} \Rightarrow \begin{cases} R = \bar{X} + z_{\frac{\alpha}{2}}\cdot\frac{\sigma}{\sqrt{n}} \\ L = \bar{X} - z_{\frac{\alpha}{2}}\cdot\frac{\sigma}{\sqrt{n}} \end{cases}$

**Setting 2:** $X_1,\dots,X_n$ IID $\mathcal{N}(\mu,\sigma^2)$ r.v., $\sigma^2$ **unknown** $\Rightarrow$ $\begin{cases} L = \bar{X} - t_{n-1}(\frac{\alpha}{2})\cdot S/\sqrt{n} \\ R = \bar{X} + t_{n-1}(\frac{\alpha}{2})\cdot S/\sqrt{n} \end{cases}$ for $\mu$

**Solution:** Goal: find $L \& R$ s.t. $P_\mu(L<\mu<R) = 1-\alpha \ \ \forall \mu$

Pivot: $\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{n-1} \Rightarrow P(-t_{n-1}(\frac{\alpha}{2}) < W < t_{n-1}(\frac{\alpha}{2})) = 1-\alpha$, $W \sim t_{n-1}$

So $P_\mu\left(\frac{\bar{X}-R}{S/\sqrt{n}} < \frac{\bar{X}-\mu}{S/\sqrt{n}} < \frac{\bar{X}-L}{S/\sqrt{n}}\right) = 1-\alpha \Rightarrow \begin{cases} \frac{\bar{X}-R}{S/\sqrt{n}} = -t_{n-1}(\alpha/2) \\ \frac{\bar{X}-L}{S/\sqrt{n}} = t_{n-1}(\alpha/2) \end{cases}$

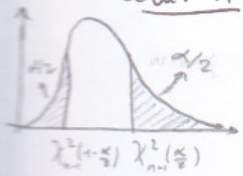
**Setting 3:** $X_1,\dots,X_n$ IID, $\mu \& \sigma^2$ **unknown**, $n \geq 30$ $\Rightarrow$ $\begin{cases} L \approx \bar{X} - z_{\frac{\alpha}{2}}\cdot S/\sqrt{n} \\ R \approx \bar{X} + z_{\frac{\alpha}{2}}\cdot S/\sqrt{n} \end{cases}$ for $\mu$

**Solution:** Pivot: $\frac{\bar{X}-\mu}{S/\sqrt{n}} \xrightarrow[n\to\infty]{CLT} \mathcal{N}(0,1)$

So $P_\mu\left(\frac{\bar{X}-R}{S/\sqrt{n}} < \frac{\bar{X}-\mu}{S/\sqrt{n}} < \frac{\bar{X}-L}{S/\sqrt{n}}\right) = 1-\alpha \Rightarrow \begin{cases} \frac{\bar{X}-R}{S/\sqrt{n}} \approx -z_{\frac{\alpha}{2}} \\ \frac{\bar{X}-L}{S/\sqrt{n}} \approx +z_{\frac{\alpha}{2}} \end{cases}$

**Setting 4:** $X_1,\dots,X_n$ IID $\mathcal{N}(\mu,\sigma^2)$ r.v., $\mu \& \sigma^2$ unknown $\Rightarrow$ $\begin{cases} L = S^2\cdot\frac{n-1}{\chi^2_{n-1}(\alpha/2)} \\ R = S^2\cdot\frac{n-1}{\chi^2_{n-1}(1-\frac{\alpha}{2})} \end{cases}$ for $\sigma^2$

**Solution:** Pivot: $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$, Goal: $P_{\sigma^2}(L<\sigma^2<R) = 1-\alpha$




So $P_{\sigma^2}\left(\frac{(n-1)S^2}{R} < \frac{(n-1)S^2}{\sigma^2} < \frac{(n-1)S^2}{L}\right) = 1-\alpha \Rightarrow \begin{cases} \frac{(n-1)S^2}{R} = \chi^2_{n-1}(1-\frac{\alpha}{2}) \\ \frac{(n-1)S^2}{L} = \chi^2_{n-1}(\frac{\alpha}{2}) \end{cases}$

**Setting 5:** $X_1,\dots,X_n$ IID Bernoulli$(p)$ r.v. $\Rightarrow$ $\begin{cases} L \simeq \hat{p} - z_{\frac{\alpha}{2}}\sqrt{\hat{p}(1-\hat{p})/n} \\ R \simeq \hat{p} + z_{\frac{\alpha}{2}}\sqrt{\hat{p}(1-\hat{p})/n} \end{cases}$ for $p$, where $\hat{p} := \frac{1}{n}\sum_{i=1}^{n} X_i$

**Solution:** Pivot: $\frac{\hat{p}-p}{\sqrt{\hat{p}(1-\hat{p})/n}} \xrightarrow[n\to\infty]{CLT} \mathcal{N}(0,1)$ : Goal: $P_p(L<p<R) = 1-\alpha$

So: $P_p\left(\frac{\hat{p}-R}{\sqrt{\cdots}} < \frac{\hat{p}-p}{\sqrt{\cdots}} < \frac{\hat{p}-L}{\sqrt{\cdots}}\right) = 1-\alpha \Rightarrow \begin{cases} \frac{\hat{p}-R}{\sqrt{\cdots}} \approx -z_{\frac{\alpha}{2}} \\ \frac{\hat{p}-L}{\sqrt{\cdots}} \approx +z_{\frac{\alpha}{2}} \end{cases}$

**Setting 6:** $X_1,\dots,X_n$ IID r.v., $\mu$ **known** $\rightarrow$ $\begin{cases} L \approx \\ R \approx \end{cases}$

**Solution:** Pivot: $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \xrightarrow[n\to\infty]{CLT} \mathcal{N}(0,1)$ : Goal: $P_{\sigma^2}(L<\sigma^2<R) = 1-\alpha$

So $P_{\sigma^2}\left(\frac{\bar{X}-\mu}{\sqrt{R/n}} < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < \frac{\bar{X}-\mu}{\sqrt{L/n}}\right) = 1-\alpha \Rightarrow \begin{cases} \frac{\bar{X}-\mu}{\sqrt{R/n}} \simeq -z_{\alpha/2} \\ \frac{\bar{X}-\mu}{\sqrt{L/n}} \simeq z_{\alpha/2} \end{cases} \Rightarrow \begin{cases} R \approx \frac{(\bar{X}-\mu)^2 n}{(z_{\alpha/2})^2} \\ L \approx \frac{(\bar{X}-\mu)^2 n}{(z_{\alpha/2})^2} \end{cases}$ ... ?

↑ ↑
Require $\bar{X} > \mu$

# III) 2-Samples Confidence Intervals :

**Idea :** • up to now : "single-sample problems" → draw a sample from a single population & make inference about it

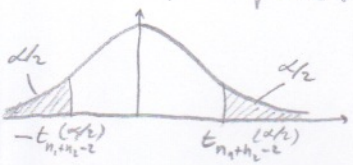
• "2-samples problems" → compare 2 populations : make inference about $(\mu_1-\mu_2)$, $\frac{\sigma_1^2}{\sigma_2^2}$, $(p_1-p_2)$, $\frac{p_1}{p_2}$ ...


**Setting 1 :** $X_1, \dots, X_{n_1}$ IID $\mathcal{N}(\mu_1, \sigma^2)$
$Y_1, \dots, Y_{n_2}$ IID $\mathcal{N}(\mu_2, \sigma^2)$
$\tilde{X} \perp\!\!\!\perp \tilde{Y}$, $(\mu_1, \mu_2, \sigma^2)$ unknown
$\Rightarrow P_{\mu_1,\mu_2}(L < \mu_1-\mu_2 < R) = 1-\alpha$ , $\forall \mu_1 \in \mathbb{R}$, $\forall \mu_2 \in \mathbb{R}$

• Pivot : $U = \dfrac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$ where $S_p^2 = \dfrac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$ → "Pooled Sample Variance"

(unbiased : $E[S_p^2] = \frac{(n_1-1)\sigma^2+(n_2-1)\sigma^2}{n_1+n_2-2} = \sigma^2$ ✓)

If we prove that $U \sim t_{n_1+n_2-2}$ : $P_{\mu_1,\mu_2}\left(\frac{(\bar{X}-\bar{Y})-R}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} < \frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} < \frac{(\bar{X}-\bar{Y})-L}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\right) = 1-\alpha$




$\Rightarrow \begin{cases} \frac{(\bar{X}-\bar{Y})-R}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = -t_{n_1+n_2-2}^{(\alpha/2)} \\ \frac{(\bar{X}-\bar{Y})-L}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = t_{n_1+n_2-2}^{(\alpha/2)} \end{cases} \Rightarrow \boxed{\begin{cases} L = (\bar{X}-\bar{Y}) - t_{n_1+n_2-2}^{(\frac{\alpha}{2})} \cdot S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} \\ R = (\bar{X}-\bar{Y}) + t_{n_1+n_2-2}^{(\frac{\alpha}{2})} \cdot S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} \end{cases}}$

• **Claim :** $\dfrac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t_{n_1+n_2-2}$ Goal : write $U \sim \dfrac{\mathcal{N}(0,1)}{\sqrt{\chi^2_{n_1+n_2-2}/(n_1+n_2-2)}}$ ↳ independent RV.

**Step 1 :** $U = \underbrace{\dfrac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}}_{V} \Big/ \sqrt{\underbrace{\dfrac{(n_1+n_2-2)S_p^2}{\sigma^2}}_{W}\Big/(n_1+n_2-2)} = \dfrac{V}{\sqrt{W}}$

**Step 2 :** $V \sim \mathcal{N}(0,1)$ (*) $\tilde{X} \perp\!\!\!\perp \tilde{Y} \Rightarrow \bar{X} \perp\!\!\!\perp \bar{Y} \Rightarrow (\bar{X}-\bar{Y})$ is Normal as $\bar{X}$ & $\bar{Y}$ are normal.

$E[\bar{X}-\bar{Y}] = E[\bar{X}]-E[\bar{Y}] = \mu_1-\mu_2$

$\mathrm{Var}[\bar{X}-\bar{Y}] = \mathrm{Var}[\bar{X}]+\mathrm{Var}[\bar{Y}] = \frac{\sigma^2}{n_1}+\frac{\sigma^2}{n_2} = \sigma^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)$ } $V \sim \mathcal{N}(0,1)$

**Step 3 :** $W \sim \chi^2_{n_1+n_2-2}/(n_1+n_2-2)$

(**) $\dfrac{(n_1+n_2-2)S_p^2}{\sigma^2} = \underbrace{\frac{(n_1-1)}{\sigma^2}S_1^2}_{\chi^2_{n_1-1}} + \underbrace{\frac{(n_2-1)}{\sigma^2}S_2^2}_{\chi^2_{n_2-1}} \sim \chi^2_{n_1+n_2-2}$ because $\tilde{X} \perp\!\!\!\perp \tilde{Y} \Rightarrow S_1^2 \perp\!\!\!\perp S_2^2$

So both $\chi^2$ are indep.

**Step 4 :** $V \perp\!\!\!\perp W$, i.e. $(\bar{X}-\bar{Y}) \perp\!\!\!\perp S_p^2$ → Goal : $f_{\bar{X},\bar{Y},S_1^2,S_2^2}^{(r,s,u,v)} = f_{\bar{X},\bar{Y}}^{(r,s)} \cdot f_{S_1^2,S_2^2}^{(u,v)}$ $\forall r,s,u,v$

$f_{\bar{X},\bar{Y},S_1^2,S_2^2}^{(r,s,u,v)} = f_{(\bar{X},S_1^2),(\bar{Y},S_2^2)}^{(r,u,s,v)} = f_{\bar{X},S_1^2}^{(r,u)} \cdot f_{\bar{Y},S_2^2}^{(s,v)} = f_{\bar{X}}(r)\, f_{S_1^2}(u)\, f_{\bar{Y}}(s)\, f_{S_2^2}(v) = f_{\bar{X},\bar{Y}}^{(r,s)} \cdot f_{S_1^2,S_2^2}^{(u,v)}$ $\forall r,s,u,v$

↑ $\tilde{X} \perp\!\!\!\perp \tilde{Y}$ ; ↑ $\bar{X} \perp\!\!\!\perp S_1^2$, $\bar{Y} \perp\!\!\!\perp S_2^2$ } under Normality ; ↑ $\bar{X} \perp\!\!\!\perp \bar{Y}$

So $(\bar{X},\bar{Y}) \perp\!\!\!\perp (S_1^2, S_2^2)$

$\Rightarrow (\bar{X}-\bar{Y}) \perp\!\!\!\perp S_p^2$ as funct° of independent vars are indep.

**Setting 2:** $X_1, \dots, X_{n_1}$ IID r.v. with $(\mu_1, \sigma_1^2)$
$Y_1, \dots, Y_{n_2}$ IID r.v. with $(\mu_2, \sigma_2^2)$
$\tilde{X} \perp\!\!\!\perp \tilde{Y}$ and $(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2)$ unknown

$n_1, n_2 \geq 30 \Rightarrow$

$$\begin{cases} L \approx (\bar{X} - \bar{Y}) - z_{\frac{\alpha}{2}} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \\ R \approx (\bar{X} - \bar{Y}) + z_{\frac{\alpha}{2}} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \end{cases}$$

for $P_{\mu_1, \mu_2}(L < \mu_1 - \mu_2 < R) = 1 - \alpha$

- Pivot: $\dfrac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} = Z$

If we prove that $Z \xrightarrow[n \to \infty]{D} \mathcal{N}(0,1)$, then $P_{\mu_1, \mu_2}\left( \underbrace{\frac{(\bar{X} - \bar{Y}) - R}{\sqrt{\dots}}}_{= -z_{\frac{\alpha}{2}}} < \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\dots}} < \underbrace{\frac{(\bar{X} - \bar{Y}) - L}{\sqrt{\dots}}}_{= +z_{\frac{\alpha}{2}}} \right) = 1 - \alpha$

- **claim:** $\dfrac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \xrightarrow[n_1, n_2 \to \infty]{D} \mathcal{N}(0,1)$

(∞,∞) $\dfrac{\bar{X} - \mu_1}{\sqrt{S_1^2/n_1}} \xrightarrow[n_1 \to \infty]{D} \mathcal{N}(0,1)$ and $\dfrac{\bar{Y} - \mu_2}{\sqrt{S_2^2/n_2}} \xrightarrow[n_2 \to \infty]{D} \mathcal{N}(0,1)$ by CLT

So $\dfrac{\bar{X} - \mu_1}{\sqrt{S_1^2/n_1}} \cdot \underbrace{\dfrac{\sqrt{S_1^2/n_1}}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}}_{\sqrt{1 + \frac{S_2^2}{S_1^2} \cdot \frac{n_1}{n_2}} \xrightarrow[n_1 \to \infty]{} 1} + \dfrac{\bar{Y} - \mu_2}{\sqrt{S_2^2/n_2}} \cdot \underbrace{\dfrac{\sqrt{S_2^2/n_2}}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}}_{\xrightarrow[n_2 \to \infty]{} 1} \xrightarrow[n_1, n_2 \to \infty]{D} \mathcal{N}(0,1)$ as $\tilde{X} \perp\!\!\!\perp \tilde{Y}$ so indep normals

**Setting 3:** $X_1, \dots, X_n$ IID Bernoulli$(p_1)$
$Y_1, \dots, Y_n$ IID Bernoulli$(p_2)$
$\tilde{X} \perp\!\!\!\perp \tilde{Y}$ and $(p_1, p_2)$ unknown

$n_1, n_2 \geq 30 \Rightarrow$
$\hat{p}_1 = \frac{1}{n_1} \sum_{i=1}^{n_1} X_i$
$\hat{p}_2 = \frac{1}{n_2} \sum_{i=1}^{n_2} Y_i$

$$\begin{cases} L \approx (\hat{p}_1 - \hat{p}_2) - z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} \\ R \approx (\hat{p}_1 - \hat{p}_2) + z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} \end{cases}$$

for $P_{p_1, p_2}(L < p_1 - p_2 < R) = 1 - \alpha$

- Pivot: $Z = \dfrac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}} \xrightarrow[n_1, n_2 \to \infty]{D} \mathcal{N}(0,1)$ by CLT

(∞) $E[\hat{p}_1 - \hat{p}_2] = E\left[\frac{1}{n_1} \sum^{n_1} X_i\right] - E\left[\frac{1}{n_2} \sum^{n_2} Y_i\right]$
$= p_1 - p_2$

$\hat{p}_1 \perp\!\!\!\perp \hat{p}_2$
$\mathrm{Var}[\hat{p}_1 - \hat{p}_2] = \mathrm{Var}[\hat{p}_1] + \mathrm{Var}[\hat{p}_2]$
$= \frac{1}{n_1} \hat{p}_1(1-\hat{p}_1) + \frac{1}{n_2} \hat{p}_2(1-\hat{p}_2)$

- So: $P_{p_1, p_2}\left( \underbrace{\frac{(\hat{p}_1 - \hat{p}_2) - R}{\sqrt{\dots}}}_{\approx -z_{\alpha/2}} < \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\dots}} < \underbrace{\frac{(\hat{p}_1 - \hat{p}_2) - L}{\sqrt{\dots}}}_{\approx z_{\alpha/2}} \right) = 1 - \alpha$

Setting 4:

**Example 7.** ***Confidence interval for the ratio $\sigma_1^2/\sigma_2^2$ of variances of two normal distributions $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$.***

We shall assume we have independent random samples of sizes $n_1$ and $n_2$ from these distributions, and that $\mu_1$ and $\mu_2$ are unknown. From chapter 1, the quantities $Y_1 = \frac{(n_1-1)s_1^2}{\sigma_1^2}$ and $Y_2 = \frac{(n_2-1)s_2^2}{\sigma_2^2}$ have chi-square distributions with $n_1 - 1$ and $n_2 - 1$ d.f. respectively, and so

$$\frac{Y_2/(n_2-1)}{Y_1/(n_1-1)} = \frac{s_2^2/\sigma_2^2}{s_1^2/\sigma_1^2}$$

has the $F$-distribution with $n_2 - 1, n_1 - 1$ degrees of freedom. Hence our pivot will be

$$\frac{s_2^2/\sigma_2^2}{s_1^2/\sigma_1^2} \sim F_{n_2-1,n_1-1}.$$

which has the $F$-distribution with $n_2 - 1, n_1 - 1$ degrees of freedom. We have

$$\begin{aligned} 1-\alpha &= P\left[F_{1-\alpha/2,n_2-1,n_1-1} < \frac{s_2^2/\sigma_2^2}{s_1^2/\sigma_1^2} < F_{\alpha/2,n_2-1,n_1-1}\right] \\ &= P\left[\frac{s_1^2}{s_2^2} \cdot F_{1-\alpha/2,n_2-1,n_1-1} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{s_1^2}{s_2^2} \cdot F_{\alpha/2,n_2-1,n_1-1}\right] \end{aligned}$$

and so we find the confidence interval to be

$$\boxed{\frac{s_1^2}{s_2^2} \cdot F_{1-\alpha/2,n_2-1,n_1-1} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{s_1^2}{s_2^2} \cdot F_{\alpha/2,n_2-1,n_1-1}.}$$

Sometimes the fact that

$$F_{1-\alpha/2,n_2-1,n_1-1} = \frac{1}{F_{\alpha/2,n_1-1,n_2-1}} \quad \text{is used in this c.i.}$$

| | Population(s) | To test | calculate | and reject $H_0$ at level $\alpha$ if | $(1-\alpha)\times 100\%$ Confidence Interval |
|---|---|---|---|---|---|
| 1. | $N(\mu,\sigma^2)$ (or *any* population if $n$ is large ($\geq 25$)) with $\sigma^2$ known. | $H_0: \mu = \mu_0$<br>$H_1: \mu \neq \mu_0$<br>$(\mu > \mu_0)$<br>$(\mu < \mu_0)$ | $z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$<br>If $n$ is large, then $\sigma$ may be replaced by $s$. | $\lvert z\rvert > z_{\alpha/2}$<br>$(z > z_\alpha)$<br>$(z < -z_\alpha)$ | $\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$<br>If $n$ is large, then $\sigma$ may be replaced by $s$. |
| 2. | $N(\mu,\sigma^2)$ with both $\mu$ and $\sigma^2$ unknown | same as above | $t = \frac{\bar{x}-\mu_0}{s/\sqrt{n}}$ | $\lvert t\rvert > t_{\alpha/2,n-1}$<br>$(t > t_{\alpha,n-1})$<br>$(t < -t_{\alpha,n-1})$ | $\bar{x} - t_{\alpha/2,n-1}\frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}$ |
| 3. | Bin($p$) (large sample case($n \geq 25$)).<br>$\hat{p} = x/n, \hat{q} = 1-\hat{p}$ | $H_0: p = p_0$<br>$H_1: p \neq p_0$<br>$(p > p_0)$<br>$(p < p_0)$ | $z = \frac{\hat{p}-p_0}{\sqrt{\frac{p_0 q_0}{n}}}$ | $\lvert z\rvert > z_{\alpha/2}$<br>$(z > z_\alpha)$<br>$(z < -z_\alpha)$ | $\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} < p < \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$ |
| 4. | $N(\mu,\sigma^2)$ with both $\mu$ and $\sigma^2$ unknown | $H_0: \sigma^2 = \sigma_0^2$<br>$H_1: \sigma^2 \neq \sigma_0^2$<br>$(\sigma^2 > \sigma_0^2)$<br>$(\sigma^2 < \sigma_0^2)$ | $\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$ | $\chi^2 > \chi^2_{\alpha/2,n-1}$ or $< \chi^2_{1-\alpha/2,n-1}$<br>$(\chi^2 > \chi^2_{\alpha,n-1})$<br>$(\chi^2 < \chi^2_{1-\alpha,n-1})$ | $\frac{(n-1)s^2}{\chi^2_{\alpha/2,n-1}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{1-\alpha/2,n-1}}$ |
| 5. | $N(\mu_1,\sigma_1^2)$ and $N(\mu_2,\sigma_2^2)$ (or *any* populations provided $n_1$ and $n_2$ are large ($\geq 25$)), where $\sigma_1^2$ and $\sigma_2^2$ are known. | $H_0: \mu_1 - \mu_2 = D_0$<br>$H_1: \mu_1 - \mu_2 \neq D_0$<br>$(\mu_1 - \mu_2 > D_0)$<br>$(\mu_1 - \mu_2 < D_0)$ | $z = \frac{\bar{x}_1-\bar{x}_2-D_0}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$<br>If $n_1$ and $n_2$ are large, $\sigma_1^2$ and $\sigma_2^2$ may be replaced by $s_1^2$ and $s_2^2$ | $\lvert z\rvert > z_{\alpha/2}$<br>$(z > z_\alpha)$<br>$(z < -z_\alpha)$ | $\bar{x}_1 - \bar{x}_2 - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}} < \mu_1 - \mu_2 < \bar{x}_1 - \bar{x}_2 + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$<br>If $n_1$ and $n_2$ are large, $\sigma_1^2$ and $\sigma_2^2$ may be replaced by $s_1^2$ and $s_2^2$ |
| 6. | $N(\mu_1,\sigma_1^2)$ and $N(\mu_2,\sigma_2^2)$, where $\sigma_1^2$ and $\sigma_2^2$ are unknown, but we assume $\sigma_1^2 = \sigma_2^2$. | same as above | $t = \frac{\bar{x}_1-\bar{x}_2-D_0}{s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$<br>where $s_p^2 = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}$ | $\lvert t\rvert > t_{\alpha/2,n_1+n_2-2}$<br>$(t > t_\alpha, n_1+n_2-2)$<br>$(t < -t_\alpha, n_1+n_2-2)$ | $\bar{x}_1 - \bar{x}_2 - t_{\alpha/2,n_1+n_2-2}s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} < \mu_1 - \mu_2 < \bar{x}_1 - \bar{x}_2 + t_{\alpha/2,n_1+n_2-2}s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$ |
| 7. | Bin($p_1$) and Bin($p_2$) (large sample case-both $n_1, n_2 \geq 25$)<br>$\hat{p}_1 = x_1/n_1, \hat{q}_1 = 1-\hat{p}_1$<br>$\hat{p}_2 = x_2/n_2, \hat{q}_2 = 1-\hat{p}_2$ | $H_0: p_1 - p_2 = D_0$<br>$H_1: p_1 - p_2 \neq D_0$<br>$(p_1 - p_2 > D_0)$<br>$(p_1 - p_2 < D_0)$ | $z = \begin{cases} \frac{\hat{p}_1-\hat{p}_2-D_0}{\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1}+\frac{\hat{p}_2\hat{q}_2}{n_2}}} & \text{if } D_0 \neq 0, \\ \frac{\hat{p}_1-\hat{p}_2}{\sqrt{\frac{\hat{p}\hat{q}}{n_1}+\frac{\hat{p}\hat{q}}{n_2}}} & \text{if } D_0 = 0, \end{cases}$<br>where $\hat{p} = \frac{x_1+x_2}{n_1+n_2}$ | $\lvert z\rvert > z_{\alpha/2}$<br>$(z > z_\alpha)$<br>$(z < -z_\alpha)$ | $\hat{p}_1 - \hat{p}_2 - z_{\alpha/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1}+\frac{\hat{p}_2\hat{q}_2}{n_2}} < p_1 - p_2 < \hat{p}_1 - \hat{p}_2 + z_{\alpha/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1}+\frac{\hat{p}_2\hat{q}_2}{n_2}}$ |
| 8. | $N(\mu_1,\sigma_1^2)$ and $N(\mu_2,\sigma_2^2)$, where $\mu_1$ and $\mu_2$ are unknown. | $H_0: \sigma_1^2 = \sigma_2^2$<br>$H_1: \sigma_1^2 \neq \sigma_2^2$<br>$(\sigma_1^2 > \sigma_2^2)$<br>$(\sigma_1^2 < \sigma_2^2)$ | $F = \frac{s_1^2}{s_2^2}$<br>Take pop'n 1 to be that with the larger sample variance. | $F \geq F_{\alpha/2,n_1-1,n_2-1}$<br>$(F > F_{\alpha,n_1-1,n_2-1})$<br>do not reject | $\frac{s_1^2}{s_2^2}\frac{1}{F_{\alpha/2,n_1-1,n_2-1}} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{s_1^2}{s_2^2}F_{\alpha/2,n_2-1,n_1-1}$ |

ll samples are assumed randomly drawn. In the two-population scenarios 5,6,7,and 8, the two samples are assumed independent of each other.

### 2.1.2 Method of Moments.

This method is based on the strong law of large numbers: if $X_1, X_2, X_3, \ldots$ is a sequence of i.i.d. random variables with finite $k$th moment $\mu_k' = E(X^k)$, then

$$\frac{1}{n}\sum_{i=1}^{n} X_i^k \longrightarrow \mu_k' \quad \text{as } n \to \infty.$$

Thus if $X_1, \cdots, X_n$ is a random sample and $n$ is large, and if we put

$$m_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k,$$

then we have

$$m_k \approx \mu_k', \quad k \geq 1.$$

The method of moments consists of solving as many of the equations

$$m_k = \mu_k', \quad k \geq 1,$$

starting with the case $k = 1$, as are necessary to identify the unknown parameters.

***Example.*** Suppose $X_1, \cdots, X_n$ is a random sample from an exponential distribution with mean $\theta$. Then $\mu_1' = \theta$ and $m_1 = \bar{X}$, so the moment estimator is $\hat{\theta} = \bar{X}$.

***Example.*** Find the moment estimator of $\theta$ given a random sample $X_1, \ldots, X_n$ from the Bernoulli distribution

$$P[X = 0] = 1 - \theta, \quad P[X = 1] = \theta.$$

***Solution.*** we have $m_1 = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}$ and $\mu_1' = EX = \theta$, so the moment estimator of $\theta$ is $\bar{X}$.

***Example.*** Suppose $X_1, \cdots, X_n$ is a random sample from a gamma distribution with parameters $\alpha$ and $\beta$. Find moment estimators of $\alpha$ and $\beta$.

***Solution.*** We have $\mu_1' = \alpha\beta$ and $\mu_2' = E(X^2) = \mathrm{Var}(X) + (EX)^2 = \alpha\beta^2 + \alpha^2\beta^2$. Hence $m_1 = \alpha\beta$ and $m_2 = \alpha\beta^2(1 + \alpha)$. Solving these two equations for $\alpha$ and $\beta$, we find

$$\hat{\alpha} = \frac{m_1^2}{m_2 - m_1^2}, \qquad \hat{\beta} = \frac{m_2 - m_1^2}{m_1}.$$