6th March:

Quiz-3 next tutorial, 11:45 AM – 12:15PM, LH102, Syllabus sufficiency and minimal sufficiency

Def: A sufficient statistic is called minimal, when for any other T', T is a function of T'
i.e $T'(\underset{\sim}{x}) = T'(\underset{\sim}{y})$
$\Rightarrow T(\underset{\sim}{x}) = T(\underset{\sim}{y})$

Theorem: $X_1, X_2, \dots X_n \overset{iid}{\sim} f(x|\theta)$, any statistic T

① any two $\underset{\sim}{x}, \underset{\sim}{y} \in \mathcal{X}$

$$\{\theta \in \Theta \mid L(\underset{\sim}{x}|\theta) > 0\} = \{\theta \in \Theta \mid L(\underset{\sim}{y}|\theta) > 0\}$$

② $\underset{\sim}{x}, \underset{\sim}{y} \in \mathcal{X}$ $\dfrac{L(\underset{\sim}{x}|\theta)}{L(\underset{\sim}{y}|\theta)}$ is constant as function of θ iff $T(\underset{\sim}{x}) = T(\underset{\sim}{y})$

then T is minimal sufficient statistic

Note: Above we are starting with any statistic, we don't need it to be sufficient

Note: first condition always true if range does not depend on parameter

Eg: $X_1, X_2, \dots X_n \overset{iid}{\sim} N(\mu, \sigma^2)$ both μ, σ are unknown parameters

$\mu \in \mathbb{R}$, $\sigma \in (0, \infty)$

we want to find minimal sufficient statistic for (μ, σ)

for any $\underset{\sim}{x} \in \mathcal{X} = \mathbb{R}^n$
as $\mathcal{X} = \mathbb{R}^n$, first condition already satisfied

now, likelihood function is

$$L(\underset{\sim}{x}|\mu, \sigma) = (2\pi)^{-n/2} \sigma^{-n} \exp\left\{-\frac{(n-1)s_x^2}{2\sigma^2}\right\} \exp\left\{-\frac{n(\bar{x}-\mu)^2}{2\sigma^2}\right\}$$

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

now, taking $\underset{\sim}{x}, \underset{\sim}{y} \in \mathbb{R}^n$ we get

$$\frac{L(\underset{\sim}{x}|\mu,\sigma)}{L(\underset{\sim}{y}|\mu,\sigma)} = \exp\left\{-\frac{(n-1)(s_x^2 - s_y^2)}{2\sigma^2}\right\} \exp\left\{-\frac{n}{2\sigma^2}\left((\bar{x}-\mu)^2 - (\bar{y}-\mu)^2\right)\right\}$$

$$= \exp\left\{-\frac{(n-1)}{2\sigma^2}(s_x^2 - s_y^2)\right\} \exp\left\{-\frac{n}{2\sigma^2}(\bar{x}^2 - \bar{y}^2)\right\} \exp\left\{\frac{\mu n}{\sigma^2}(\bar{x} - \bar{y})\right\}$$

if $\bar{x} = \bar{y}$ and $s_x^2 = s_y^2$

then $\dfrac{L(\underset{\sim}{x}|\mu,\sigma)}{L(\underset{\sim}{y}|\mu,\sigma)} = 1$ i.e constant as function of (μ, σ)

conversely suppose that $\dfrac{L(\underset{\sim}{x}|\mu\sigma)}{L(\underset{\sim}{y}|\mu,\sigma)}$ is a function of (μ, σ), in particular

this means that, the likelihood ratio is a constant function of μ when

$\sigma=1$, only when $\bar{x}=\bar{y}$

So, this implies $\bar{x}=\bar{y}$, then

$$\frac{L(\underset{\sim}{x}|\mu,\sigma)}{L(\underset{\sim}{y}|\mu,\sigma)} = \exp\left\{-\frac{(n+1)}{2\sigma^2}\underbrace{(s_x^2-s_y^2)}_{s_x^2=s_y^2}\right\}$$

then as fution of σ^2 constant

$\Rightarrow s_x^2=s_y^2$ then $\frac{L(\underset{\sim}{x}|\mu,\sigma)}{L(\underset{\sim}{y}|\mu,\sigma)}$ is constant fution of σ^2

So we have shown that $\forall\ \underset{\sim}{x},\underset{\sim}{y}\in\mathcal{X}$

$$\frac{L(\underset{\sim}{x}|\mu,\sigma)}{L(\underset{\sim}{y}|\mu,\sigma)} = \text{constant as fution of } \mu,\sigma \iff \bar{x}=\bar{y} \text{ and } s_x^2=s_y^2$$

therefore by the theorm $T(\underset{\sim}{x})=(\bar{x},s_x^2)$ is a minimal sufficient statistic

Note: If we go to uniform, we have to see the first part also

Eg: $X_1,X_2,\dots X_n \overset{iid}{\sim} \text{unif}(0,\theta)$, where $\theta\in(0,\infty)$ is unknown

$$L(\underset{\sim}{x}|\theta)=\begin{cases}\frac{1}{\theta^n}; & x_{(1)}>0,\ x_{(n)}<\theta\\ 0; & \text{otherwise}\end{cases}$$

So, for two sample points $\underset{\sim}{x},\underset{\sim}{y}$

$L(\underset{\sim}{x}|\theta)$ and $L(\underset{\sim}{y}|\theta)$ will be positive for same values of θ iff $x_{(n)}=y_{(n)}$

in this case likelyhood ratio $\frac{L(\underset{\sim}{x}|\theta)}{L(\underset{\sim}{y}|\theta)}=1$ which is a constant fution of θ

So, $T(\underset{\sim}{X})=X_{(n)}$ is minimal sufficient for θ

Note: It is not explicitly written in above example but it's hidden in second that it denotes constant when $T(\underset{\sim}{X})=X_{(n)}$

Exe: If $X_1,X_2,\dots X_n \overset{iid}{\sim} \text{unif}(\theta,\theta+1)$, then show that $T(\underset{\sim}{X})=(X_{(1)},X_{(n)})$ the minimal sufficient statistic for θ → done down

Exe: Show the following:

(a) any one-to-one fution of a sufficient statistic is sufficient → done down

(b) minimal sufficient statistic one-to-one fution is minimal sufficient

Note: minimal sufficient statistic, if exist, is not unique

Applications of not uniqueness:

If $X_1,X_2,\dots X_n \overset{iid}{\sim} N(\mu,\sigma^2)$ then

$T(\underset{\sim}{x})=(\bar{x},S^2)$ is minimal and

$T'(\underset{\sim}{X}) = \left(\sum_{i=1}^{n} X_i, \sum X_i^2\right)$ is a minimal sufficient statistic for (μ, σ)

also, if $X_1, X_2, \ldots X_n \overset{iid}{\sim} \text{unif}(\theta, \theta+1)$ then $T'(\underset{\sim}{X}) = (X_{(n)} - X_{(1)}, X_{(n)} + X_{(1)})$ is minimal sufficient statistic for θ

Exe: Suppose $X_1, X_2, \ldots X_n \overset{iid}{\sim} N(\mu, \mu^2)$ show that $\left(\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i^2\right)$ is minimal sufficient statistic for μ → done down

Exe: Suppose $X_1, X_2, \ldots X_n \overset{iid}{\sim} N(\mu, \mu)$ show the following:

(a) $\left(\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i^2\right)$ is sufficient for μ but not minimal sufficient

(b) $\sum_{i=1}^{n} X_i^2$ is minimal sufficient for μ → done down

Ancillary Statistics:

$X_1, X_2, \ldots X_n \overset{iid}{\sim} f(x|\theta)$ pmf/pdf

Defn: A statistic $S = S(\underset{\sim}{X})$ is called an ancillary statistic for θ if the distribution of S does not depend on θ

Note: An ancillary statistic alone contains no information about θ

Eg: $X_1, X_2, \ldots X_n \overset{iid}{\sim} \text{Unif}(\theta, \theta+1)$ then $R = X_{(n)} - X_{(1)}$ (sample range) is an ancillary statistic for θ

Exe: Find the pdf of R and show that it is indeed an ancillary statistic for θ → done down

one more way to show it is $Y_i = X_i - \theta \quad i = 1, 2, \ldots n$

$Y_1, Y_2, \ldots Y_n \overset{iid}{\sim} \text{Unif}(0,1)$

$X_i = \theta + Y_i$

$(X_{(1)}, X_{(2)}, \ldots X_{(n)}) = (Y_{(1)} + \theta, Y_{(2)} + \theta, \ldots Y_{(n)} + \theta)$

$\Rightarrow R = X_{(n)} - X_{(1)} = Y_{(n)} - Y_{(1)}$

→ sample range of Unif(0,1) whose distribution will be independent of θ

$\Rightarrow R = X_{(n)} - X_{(1)}$ is an ancillary statistic for θ

Exe: Suppose $X_1, X_2, \ldots X_n \overset{iid}{\sim} F(x-\theta)$ (cdf) where $\theta \in \mathbb{R} = \Theta$ is the location parameter

($\{F(x-\theta): \theta \in \mathbb{R}\}$ is called a location family of distributions or location parameter family)

show that in this case, $R = X_{(n)} - X_{(1)}$ is an ancillary statistic for θ → done down

So, if $X_1, X_2, \ldots X_n \overset{iid}{\sim} F(x-\theta)$ type was R as ancillary

Eg: Suppose $X_1, X_2, \dots X_n \overset{iid}{\sim} Exp(\lambda)$ then $S = \frac{X_1}{\sum_{i=1}^{n} X_i}$ is ancillary statistic for λ $(\lambda > 0)$

$Y_1 = X_1 \sim Exp(\lambda) \equiv gamma(1, \lambda)$

$Y_2 = \sum_{i=2}^{n} X_i \sim gamma(n-1, \lambda)$

$S = \frac{X_1}{\sum_{i=1}^{n} X_i} = \frac{Y_1}{Y_1 + Y_2}$ ← $gamma(\alpha_1, \lambda)$, → $gamma(\alpha_2, \lambda)$

then $S \sim Beta(\alpha_1, \alpha_2)$

$= Beta(1, n-1)$

So, $S \sim Beta(1, n-1)$ does not depend on λ

Ex: If $X_1, X_2, \dots X_n \overset{iid}{\sim} unif(\theta, \theta+1)$, then show that $T(\underset{\sim}{X}) = (X_{(1)}, X_{(n)})$ is minimal sufficient statistic for θ

Ans:

$$L(\underset{\sim}{x}|\theta) = \mathbb{I}_{x_{(1)} > \theta} \, \mathbb{I}_{x_{(n)} < 1+\theta}$$

$$\{\theta \in \Theta \mid L(\underset{\sim}{x}|\theta) > 0\} = \{\theta \in \Theta \mid L(\underset{\sim}{y}|\theta) > 0\}$$

$$\Rightarrow \{\theta \in \Theta \mid \mathbb{I}_{x_{(1)} > \theta} \, \mathbb{I}_{x_{(n)} < 1+\theta}\} = \{\theta \in \Theta \mid \mathbb{I}_{y_{(1)} > \theta} \, \mathbb{I}_{y_{(n)} < 1+\theta}\}$$

only when $x_{(1)} = y_{(1)}$

$x_{(n)} = y_{(n)}$

then $\frac{L(\underset{\sim}{x}|\theta)}{L(\underset{\sim}{y}|\theta)} = 1$ when $x_{(1)} = y_{(1)}$, $x_{(n)} = y_{(n)}$

$\Rightarrow T(\underset{\sim}{x}) = (x_{(1)}, x_{(n)})$ is minimal sufficient statistic

Ex: Show the following:

(a) any one-to-one function of a sufficient statistic is sufficient

(b) minimal sufficient statistic one-to-one function is minimal sufficient

Ans:

T is a sufficient statistic

let $\varphi: \{T(\underset{\sim}{x}) \mid \underset{\sim}{x} \in \mathcal{X}\} \to \{\varphi(T(\underset{\sim}{x})) \mid \underset{\sim}{x} \in \mathcal{X}\}$

φ is one-one and $T'(\underset{\sim}{x}) = \varphi(T(\underset{\sim}{x}))$ ↓ we can make it this function

So, φ is also onto as we have taken same domain

now, by factorisation criterion,

$\Rightarrow L(\underset{\sim}{x}|\theta) = h(\underset{\sim}{x})\, g(T(\underset{\sim}{x})|\theta)$

$\Rightarrow L(\underset{\sim}{x}|\theta) = h(\underset{\sim}{x})\, g(\varphi^{-1}(T'(\underset{\sim}{x}))|\theta)$

$= h(\underset{\sim}{x})\, \tilde{g}(T'(\underset{\sim}{x})|\theta)$

$\Rightarrow T'$ is sufficient ($\because$ factorisation criterion)

now, if T is minimal sufficient, for any T'' sufficient

$\exists h$ s.t. $h(T''(\underset{\sim}{x})) = T(\underset{\sim}{x})$

h: Range$(T'') \longrightarrow$ Range(T)

$\varphi \circ h$: Range$(T'') \longrightarrow$ Range(T')

($\because \varphi$ is one-one)

so, $\exists\ \varphi \circ h$ s.t. $\varphi \circ h(T''(\underset{\sim}{x})) = T'(\underset{\sim}{x})$

so, T' is also minimal sufficient ($\because$ sufficient) by defn

Ex: Suppose $X_1, X_2, \dots X_n \overset{iid}{\sim} N(\mu, \mu^2)$ show that $\left(\sum_{i=1}^n x_i, \sum_{i=1}^n x_i^2\right)$ is minimal sufficient statistic for μ

Ans:

$X_1, X_2, \dots X_n \overset{iid}{\sim} N(\mu, \mu^2)$

$\{\theta \in \Theta \mid L(\underset{\sim}{x}|\theta) > 0\} = \{\theta \in \Theta \mid L(\underset{\sim}{y}|\theta) > 0\}$

also, $\dfrac{L(\underset{\sim}{x}|\theta)}{L(\underset{\sim}{y}|\theta)} = \dfrac{\exp\left\{-\sum_{i=1}^n \frac{(x_i-\mu)^2}{2\mu^2}\right\}}{\exp\left\{-\sum_{i=1}^n \frac{(y_i-\mu)^2}{2\mu^2}\right\}}$

$= \dfrac{\exp\left\{-\frac{1}{2\mu^2}\left[\sum_{i=1}^n x_i^2 + n\mu^2 - 2\mu\sum_{i=1}^n x_i\right]\right\}}{\exp\left\{\frac{-1}{2\mu^2}\left[\sum_{i=1}^n y_i^2 + n\mu^2 - 2\mu\sum_{i=1}^n y_i\right]\right\}}$

$= \exp\left\{-\frac{1}{2\mu^2}\left[\sum_{i=1}^n x_i^2 - \sum_{i=1}^n y_i^2\right] + \frac{1}{\mu}\left[\sum_{i=1}^n x_i - \sum_{i=1}^n y_i\right]\right\}$

now this will be constant wrt $\mu \Leftrightarrow \sum_{i=1}^n x_i^2 = \sum_{i=1}^n y_i^2$, $\sum_{i=1}^n x_i = \sum_{i=1}^n y_i$

so, $T(\underset{\sim}{x}) = \left(\sum_{i=1}^n x_i^2, \sum_{i=1}^n x_i\right)$ is a minimal sufficient statistic

Ex: Suppose $X_1, X_2, \dots X_n \overset{iid}{\sim} N(\mu, \mu)$ show the following:

(a) $\left(\sum_{i=1}^{n} x_i, \sum_{i=1}^{n} x_i^2\right)$ is sufficient for μ but not minimal sufficient

(b) $\sum_{i=1}^{n} x_i^2$ is minimal sufficient for μ

Ans: $L(\underset{\sim}{x}|\mu) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\mu}} e^{-(x_i-\mu)^2}$

$= (2\pi\mu)^{-n/2} \exp\left\{-\sum_{i=1}^{n} \frac{(x_i-\mu)^2}{2\mu}\right\}$

$= (2\pi\mu)^{-n/2} \exp\left\{-\frac{\left(\sum_{i=1}^{n} x_i^2\right) - n\mu^2 + 2\mu\sum_{i=1}^{n} x_i}{2\mu}\right\}$

$= (2\pi\mu)^{-n/2} \exp\left\{-\frac{\left(\sum_{i=1}^{n} x_i^2\right)}{2\mu} - \frac{n\mu}{2} + \sum_{i=1}^{n} x_i\right\}$

then trivial to see

$= \underbrace{h(\underset{\sim}{x})}_{1}\, g(T(\underset{\sim}{x})|\mu)$

$g(t_1, t_2|\mu) = (2\pi\mu)^{-n/2} \exp\left\{\frac{-t_1}{2\mu} - \frac{n\mu}{2} + t_2\right\}$

So, $T(\underset{\sim}{x}) = \left(\sum_{i=1}^{n} x_i^2, \sum_{i=1}^{n} x_i\right)$ is sufficient

also, $\left\{\theta \in \Theta \,\middle|\, L(\underset{\sim}{x}|\theta) > 0\right\} = \left\{\theta \in \Theta \,\middle|\, L(\underset{\sim}{y}|\theta) > 0\right\}$

$$\frac{L(\underset{\sim}{x}|\mu)}{L(\underset{\sim}{y}|\mu)} = \exp\left\{-\frac{\left[\sum_{i=1}^{n} x_i^2 - \sum y_i^2\right]}{2\mu} + \left[\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} y_i\right]\right\}$$

const wrt $\mu \Leftrightarrow \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} y_i^2$

So, $T(\underset{\sim}{x}) = \left(\sum_{i=1}^{n} x_i^2\right)$ is minimal sufficient statistic

Exe: Find the pdf of R and show that it is indeed an auxiliary statistic for θ

Ans:

$X_1, X_2, \ldots X_n \overset{iid}{\sim} \text{unif}(\theta, 1+\theta)$

$f_R(r) = \int_{\mathbb{R}} (n)(n-1)[F(u+r) - F(u)]^{n-2} f(u+r) f(u)\, du$

$\theta < u+r < \theta+1$

$\theta < u < \theta+1 \quad 1 > r > 0$

$\theta - r < u < \theta+1-r \qquad \theta < u < \theta+1 \Rightarrow \theta < u < \theta+1-r$

$$= \int_{\theta}^{\theta+1-\gamma} (n)(n-1)[\gamma]^{n-2} du$$

$$f_R(\gamma) = (n)(n-1)[\gamma]^{n-2}[1-\gamma]$$

$$\gamma \in (0,1)$$

ex: suppose $X_1, X_2, \ldots X_n \overset{iid}{\sim} F(x-\theta)$ (cdf) where $\theta \in \mathbb{R} = \Theta$ is the location parameter
($\{F(x-\theta) : \theta \in \mathbb{R}\}$ is called a location family of distributions or location parameter family)
show that in this case, $R = X_{(n)} - X_{(1)}$ is an ancillary statistic for θ

Ans:

let $Y_i = X_i + \theta$ $\quad X_i \sim F(x-\theta)$ (cdf)

$\Rightarrow Y_i \sim F(x+\theta-\theta) = F(x)$

$(X_1, \ldots X_n) = (Y_1 - \theta, \ldots Y_n - \theta)$

$\Rightarrow (X_{(1)}, X_{(2)} \ldots X_{(n)}) = (Y_{(1)} - \theta, \ldots Y_{(n)} - \theta)$

$\Rightarrow R = X_{(n)} - X_{(1)} = \underbrace{Y_{(n)} - Y_{(1)}}$

does not depend on θ

10th March :

Eg: Suppose F is a CDF

$$X_1, X_2, \dots X_n \overset{iid}{\sim} P[X_i \le x] = F\left(\frac{x}{\sigma}\right)$$

$\sigma \in (0,\infty) = \Theta$ is unknown (but fixed) parameter
(the family, $\{F(\frac{x}{\sigma}) ; \sigma \in (0,\infty)\}$ is called a scale parameter or just scale family → done done)

then show that $S = \frac{X_1}{\sum_{i=1}^{n} X_i}$ is an ancillary statistic for σ

minimal sufficient statistic is not completely unrelated to ancillary statistic

Eg: $X_1, X_2, \dots X_n \overset{iid}{\sim} \text{Unif}(\theta, \theta+1)$ then

$T(\underset{\sim}{X}) = (X_{(n)} - X_{(1)},\ X_{(n)} + X_{(1)})$ is a minimal sufficient statistic for θ

but T is not independent of R, where

$R :=$ Ancillary statistic

So, even though T is minimally sufficient, it can have ancillary component

However, T is minimal sufficient, but not complete. A complete statistic avoids maximum data reduction, without any ancillary statistic

Defn: Suppose $X_1, X_2, \dots X_n \overset{iid}{\sim} f(x|\theta)$ pmf/pdf, $\theta \in \Theta$ = parameter space
A statistic $T := T(\underset{\sim}{X})$ is called complete if for any function

$g : \text{Range}(T) \to \mathbb{R}$, g does not depend on θ

$$\mathbb{E}_\theta[g(T)] = 0 \ \forall \theta \in \Theta \Rightarrow P_\theta[g(T) = 0] = 1 \ \forall \theta \in \Theta$$

So, if on average $g(T) = 0 \ \forall \theta \in \Theta$ then $g(T) = 0$ with probability 0

Eg: If $X_1, X_2, \dots X_n \overset{iid}{\sim} \text{Unif}(\theta, \theta+1)$ $\theta \in \mathbb{R} = \Theta$

then $T(\underset{\sim}{X}) = (X_{(n)} - X_{(1)},\ X_{(n)} + X_{(1)})$ is not complete

as, $\mathbb{E}_\theta[\underbrace{X_{(n)} - X_{(1)}}_{R \text{ (ancillary)}}] = c \in (0,1)$ s.t c does not depend on θ

Let $g : \text{Range } T \to \mathbb{R}$, $g(u,v) = u - c$

then $\mathbb{E}_\theta[g(T)] = \mathbb{E}_\theta[X_{(n)} - X_{(1)} - c]$

$= c - c$

$= 0$

$\Rightarrow \mathbb{E}_\theta[g(T)] = 0 \ \forall \theta \in \Theta$

but $P_\theta[g(T) = 0] = P_\theta[R = c] = 0 \ \forall \theta \in \Theta$ ($\because R$ is continuous random variable)

so, T is not complete

Eg: $X_1, X_2, \dots X_n \overset{iid}{\sim} Ber(p)$ $p \in \Theta = (0,1)$

then $T(\underset{\sim}{X}) = \sum_{i=1}^{n} X_i$ is a complete statistic

$T(\underset{\sim}{X}) \sim Bin(n, p)$

$Range(T) = \{0, 1, 2, \dots, n\}$

$g : \{0, 1, 2, \dots n\} \longrightarrow \mathbb{R}$

$\mathbb{E}_p[g(T)] = 0 \quad \forall p \in (0,1)$

if $P_p[g(T) = 0] = 1 \quad \forall p \in (0,1)$ then we are done

now, $\mathbb{E}_p[g(T)] = 0$

$\Leftrightarrow \sum_{k=0}^{n} g(k) \binom{n}{k} p^k (1-p)^{n-k} = 0 \quad \forall p \in (0,1)$

$\Rightarrow \sum_{k=0}^{n} g(k) \binom{n}{k} \left(\frac{p}{1-p}\right)^k = 0 \quad (\because (1-p)^n \neq 0 \text{ common})$

$r = \frac{p}{1-p}$ $\quad p \in (0,1) \Rightarrow r \in (0, \infty)$

then $\sum_{k=0}^{n} g(k) \binom{n}{k} r^k = 0$

polynomial in r with degree almost n but was uncountable many roots

$\Rightarrow$ polynomial is zero polynomial

$\Rightarrow g(k) \binom{n}{k} = 0 \quad \forall k \in \{0, 1, \dots n\}$

$\Rightarrow g(k) = 0 \quad \left(\because \binom{n}{k} \neq 0, \forall k \in \{0, 1, \dots n\}\right)$

so, $P_p[g(T) = 0] = 1 \quad \forall p \in (0,1)$

$\Rightarrow T(\underset{\sim}{X})$ is complete sufficient statistic for p

Theorem: (Basu) If $T = T(\underset{\sim}{X})$ is complete and sufficient statistic for θ and $S = S(\underset{\sim}{X})$ is an ancillary statistic of θ, then $T \perp\!\!\!\perp S$

Theorem: Suppose $X_1, X_2, \dots X_n \overset{iid}{\sim} f(x|\theta)$ pmf/pdf where $\mathcal{X}$ does not depend on θ and

$$f(x|\theta) = \begin{cases} h(x) \exp\left\{\sum_{i=1}^{k} \eta_i(\theta) T_i(x) - \psi(\theta)\right\}; & \text{if } x \in \mathcal{X} \\ 0 & \text{; if } x \in \mathcal{X}^c \end{cases}$$

is in exponential family with $d = k$, so, $\underset{\sim}{\theta} = (\theta_1, \theta_2, \dots \theta_k) \in \Theta \subseteq \mathbb{R}^k$

suppose parameter space Θ contains an open set in $\mathbb{R}^k$, then

$T(\underset{\sim}{X}) := \left(\sum_{j=1} T_1(X_j), \sum_{j=1} T_2(X_j), \dots \sum_{j=1}^{n} T_k(X_j)\right)$

is a complete sufficient statistic for $\underset{\sim}{\theta}$, it is also minimal sufficient

Eg: $X_1, X_2, \dots X_n \overset{iid}{\sim} Exp(\lambda)$, parameter space $\Theta = (0, \infty)$, which is an open subset of $\mathbb{R}$

$$f(x|\lambda) = \lambda e^{-\lambda x} = \exp\{\underbrace{\lambda}_{\eta_1(\lambda)}\underbrace{(-x)}_{T_1(x)} + \underbrace{\log\lambda}_{-\psi(\lambda)}\}$$

by above theorm,

$T(\underset{\sim}{X}) = -\sum_{j=1}^{n} X_j$ is complete sufficient for λ

$\Rightarrow \sum_{j=1}^{n} X_j$ is complete sufficient for λ

also, $\frac{X_1}{\sum_{i=1}^{n} X_i} \sim Beta(1, n-1)$ is ancillary, so by Basu's theorm

$\sum_{i=1}^{n} X_i \perp\!\!\!\perp \frac{X_1}{\sum_{i=1}^{n} X_i}$, in particular $n=2$: $X_1, X_2 \overset{iid}{\sim} Exp(\lambda)$

$\Rightarrow X_1 + X_2 \sim gamma(2, \lambda)$

$\frac{X_1}{X_1+X_2} \sim Beta(1,1) = Unif(0,1)$

and $X_1 + X_2 \perp\!\!\!\perp \frac{X_1}{X_1+X_2}$

by Basu's theorm

ind
- $X_1 + X_2 \sim gamma(2, \lambda)$
- $\frac{X_1}{X_1+X_2} \sim Unif(0,1)$

Exe: Using above theorm, show that $\bar{X}$ is complete and sufficient for μ where $X_1, X_2, \dots X_n \overset{iid}{\sim} N(\mu, \sigma_0^2)$ σ_0^2 is known → done down

also, $S^2 = \frac{1}{(n-1)} \sum_{i=1}^{n} (X_i - \bar{X})^2$ is an ancillary statistic

Eg: $X_1, X_2, \dots X_n \overset{iid}{\sim} N(\mu, \sigma_0^2)$ where σ_0^2 is known and $\mu \in \mathbb{R}$ is unknown

then from above exercise we get

$\bar{X} \perp\!\!\!\perp S^2$ by Basu's theorm as $\frac{\bar{X}}{S^2}$ is complete sufficient is ancillary

Theorm: If a minimal sufficient statistic exist, then any complete sufficient statistic is also a minimal sufficient statistic

Theorm: (Basu) If $T = T(\underset{\sim}{X})$ is complete and sufficient statistic for θ and $S = S(\underset{\sim}{X})$ is an ancillary statistic of θ, then $T \perp\!\!\!\perp S$

proof: In discrete case

$\forall s \in Range(S) = \{$

$P_\theta[S = s] = P[S = s]$ does not depend on θ as S is an ancillary statistic

Also, for each $s \in \mathcal{S}$ and $\forall\, t \in \text{Range}(T) = \mathcal{T}$, the conditional probability

$$P_\theta[S(\underset{\sim}{X}) = s \mid T = t] = \sum_{\underset{\sim}{x}:\, S(\underset{\sim}{x}) = s} P_\theta[\underset{\sim}{X} = \underset{\sim}{x} \mid T = t]$$

$P_\theta[\underset{\sim}{X} = \underset{\sim}{x} \mid T = t]$ is free of θ as $T := T(\underset{\sim}{X})$ is sufficient for θ

$$\Rightarrow P_\theta[S(\underset{\sim}{X}) = s \mid T = t] = \sum_{\underset{\sim}{x}:\, S(\underset{\sim}{x}) = s} P[\underset{\sim}{X} = \underset{\sim}{x} \mid T = t]$$

$$= P[S(\underset{\sim}{X}) = s \mid T = t]$$

to show that $T \perp\!\!\!\perp S$, it is sufficient to show that $\forall\, t \in \mathcal{T}, \forall\, s \in \mathcal{S}$

$$P[S = s \mid T = t] = P[S = s]$$

fix $s \in \mathcal{S} = \text{Range}(S)$, note that $\forall\, \theta \in \Theta$

$$P[S = s] = \sum_{t \in \mathcal{T}} P[S = s \mid T = t]\, P_\theta[T = t]$$

also note that,

$$P[S = s] = P[S = s] \sum_{t \in \mathcal{T}} P_\theta[T = t]$$

$$= \sum_{t \in \mathcal{T}} P[S = s]\, P_\theta[T = t]$$

$$\Rightarrow \sum_{t \in \mathcal{T}} P[S = s]\, P_\theta[T = t] = \sum_{t \in \mathcal{T}} P[S = s \mid T = t]\, P_\theta[T = t]$$

$$\Rightarrow \sum_{t \in \mathcal{T}} \underbrace{\left(P[S = s \mid T = t] - P[S = s]\right)}_{\text{free of } \theta} P_\theta[T = t] = 0$$

$$g(t) = P[S = s \mid T = t] - P[S = s] : \mathcal{T} \longrightarrow \mathbb{R}$$

↳ completely free of θ

$$\Rightarrow \sum_{t \in \mathcal{T}} g(t)\, P_\theta[T = t] = 0 \quad \forall\, \theta \in \Theta$$

$$\Rightarrow \mathbb{E}_\theta[g(T)] = 0 \quad \forall\, \theta \in \Theta$$

$$\Rightarrow P_\theta[g(T) = 0] = 1 \quad \forall\, \theta \in \Theta \; (\because \text{completeness of } T)$$

in discrete setup T is a discrete statistic,

in Range of T, $g(t) = 0\ \forall\, t \in \mathcal{T}$

$$\Rightarrow P[S = s \mid T = t] = P[S = s] \quad \forall\, t \in \mathcal{T}, \forall\, s \in \mathcal{S}$$

$$\Rightarrow T \perp\!\!\!\perp S$$

Ex: Suppose F is a CDF

$$X_1, X_2, \ldots X_n \overset{iid}{\sim} P[X_1 \le x] = F\left(\frac{x}{\sigma}\right)$$

$\sigma \in (0,\infty) = \Theta$ is unknown (but fixed) parameter
(the family, $\{F(\frac{x}{\sigma}); \sigma \in (0,\infty)\}$ is called a scale parameter or just scale family)

then show that $S = \frac{X_1}{\sum_{i=1}^{n} X_i}$ is an ancillary statistic for σ

Ans: $(X_1, X_2, \dots X_n) = (\sigma Y_1, \sigma Y_2, \dots \sigma Y_n)$

then $Y_1, Y_2, \dots Y_n \overset{iid}{\sim} F(x)$

also $\frac{X_1}{\sum_{i=1}^{n} X_i} = \frac{\sigma X_1}{\sum_{i=1}^{n} \sigma X_i} = \frac{Y_1}{\sum_{i=1}^{n} Y_i}$ ← does not depend on σ

Ex: Using above theorem, show that $\bar{X}$ is complete and sufficient for μ where $X_1, X_2, \dots X_n \overset{iid}{\sim} N(\mu, \sigma_0^2)$ σ_0^2 is known

also, $s^2 = \frac{1}{(n-1)} \sum_{i=1}^{n} (X_i - \bar{X})^2$ is an ancillary statistic

Ans:

$$f(x|\mu) = \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{\frac{-(x-\mu)^2}{2\sigma_0^2}}$$

$$= \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{1}{2\sigma_0^2}(x^2+\mu^2-2x\mu)}$$

$$= \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{x^2}{2\sigma_0^2}} e^{-\frac{\mu^2}{2\sigma_0^2}} e^{\frac{x\mu}{2\sigma_0^2}}$$

$k=1=d$ $T_1(x) = \frac{x}{n}$ $\eta_1(\mu) = \frac{\mu \times n}{2\sigma_0^2}$ $\psi(\mu) = \frac{\mu^2}{2\sigma_0^2}$

then trivially $\Theta = \mathbb{R}$ contains open set in $\mathbb{R}$

by above theorem $\sum_{i=1}^{n} T_1(X_i) = \sum \frac{X_i}{n}$ is complete sufficient

now $\frac{(n-1)s^2}{\sigma_0^2} \sim \chi^2_{n-1}$ ← always as does not depend on μ

$\Rightarrow s^2 \sim \frac{\sigma_0^2}{n-1} \chi^2_{n-1}$ always

so, by Basu's theorem

$s^2 \perp\!\!\!\perp \bar{X}$

13th March:

Quiz-4 : Next Tuesday 3:45 pm – 4:15 pm LA002 (complete, Ancillary, metod of moments) for point estimation

Estimation of parameters:

setup is $X_1, X_2, \dots X_n \overset{iid}{\sim} f(x|\theta)$ pmf/pdf

θ (possibly a vector) is unknown but fixed

goal is to estimate θ based on the random sample $X_1, X_2, \dots X_n$
more specifically we would like to estimate θ using "suitable" statistic

$$T(X) = T(X_1, X_2, \dots X_n)$$

in this case, we will write

$$\hat{\theta} = \hat{\theta}_n = T(X_1, X_2, \dots X_n)$$

↑ point estimator of θ based on a random sample of size

Estimation of parameter

- point estimator
  - methods of finding point estimator
    - metod of moments
    - metod of maximum likelyhood
    - Bayasian metod
  - evaluating point estimators
    - unbiasness
    - UMVUE
    - MSE
    - consistency
    - Efficiency
- interval estimator

Estimator and Estimate:

Estimator is $T(X_1, X_2, \dots X_n)$, A function of $\underset{\sim}{X} = (X_1, X_2, \dots X_n)$

Estimate is value of an estimator $T(x_1, x_2, \dots x_n)$

Method of moments:

suppose $X_1, X_2, \dots X_n \overset{iid}{\sim} f(x|\theta_1, \theta_2, \dots \theta_k)$ (PMF/pdf)

let $\mu_j^\circ = \mu_j^\circ(\theta_1, \theta_2, \dots \theta_k) = E[X_1^j]$

$m_j^\circ = \frac{1}{n} \sum_{j=1}^{n} X_j^j$ ← j^{th} sample moment

j^{th} population moment

metod of moment estimator or moment estimator

$$\mu_1(\theta_1, \theta_2, \ldots \theta_k) = m_1$$
$$\mu_2(\theta_1, \theta_2, \ldots \theta_k) = m_2$$
$$\vdots$$
$$\mu_k(\theta_1, \ldots \theta_k) = m_k$$

solving the above system of equations we get $(\theta_1, \theta_2, \ldots \theta_k)$

Eg: $X_1, X_2, \ldots X_n \overset{iid}{\sim} N(\mu, \sigma^2)$ μ, σ unknown

$$m_1 = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}$$

$$m_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$$

$$\mu_1(\mu, \sigma) = \mathbb{E}[X_1] = \mu$$
$$\mu_2(\mu, \sigma) = \sigma^2 + \mu^2$$

to find moment estimator of μ and σ

$$\mu = \bar{X}$$
$$\sigma^2 + \mu = \frac{1}{n}\sum_{i=1}^{n} X_i^2$$

$$\Rightarrow \hat{\mu} = \bar{X}$$

& $$\hat{\sigma}^2 = m_2 - m_1^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - (\bar{X})^2$$
$$= \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

Exe: $X_1, X_2, \ldots X_n \overset{iid}{\sim} \text{Gamma}(\alpha, \lambda)$ $(\alpha, \lambda) \in (0, \infty)^2$ unknown
show that $m_2 - m_1^2 \neq 0$ with prob 1 and → done down

$$\hat{\alpha} = \frac{m_1^2}{m_2 - m_1^2} \qquad \hat{\lambda} = \frac{m_1}{m_2 - m_1^2}$$ are moment estimators

Eg: $X_1, X_2, \ldots X_n \overset{iid}{\sim} \text{Unif}(-\theta, \theta)$ $\theta \in (0, \infty)$ unknown

$$\mu_1 = \mathbb{E}[X_1] = 0$$

$\mu_1 = m_1$ does not help us to get an estimator of θ

$$\mu_2 = \mathbb{E}[X_1^2] = \text{var}[X_1] = \frac{(\theta - (-\theta))^2}{12} = \frac{4\theta^2}{12} = \frac{\theta^2}{3}$$

$$\frac{\theta^2}{3} = m_2$$

So, moment estimator of θ can be found by solving

$$\hat{\theta} = \sqrt{3m_2}$$

i.e $\hat{\theta} = \sqrt{\frac{3}{n}\sum_{i=1}^{n} X_i^2}$

Exe: $X_1, X_2, \dots X_n \overset{iid}{\sim} Bin(k, p)$ $k \in \mathbb{N}$, $p \in (0,1)$ both unknown

then show that $\hat{k} = \frac{\bar{X}^2}{\bar{X} - \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2}$ → done down

$\hat{p} = \frac{\bar{X}}{\hat{k}}$

Note: Above, both $\hat{k}$, $\hat{p}$ can be negative and so outside the parameter space, $\hat{k}$ will rarely be an integer. So, moment estimators can therefore give us very bad estimates.

Note: moment estimators are useful as the initial values of an iterative procedure like the Newton-Raphson. So, they can be used for finding another type of estimator like MLE

Exe: $X_1, X_2, \dots X_n \overset{iid}{\sim} gama(\alpha, \lambda)$ $(\alpha, \lambda) \in (0, \infty)^2$ unkown

show that $m_2 - m_1^2 \neq 0$ with prob 1 and

$\hat{\alpha} = \frac{m_1^2}{m_2 - m_1^2}$ $\hat{\lambda} = \frac{m_1}{m_2 - m_1^2}$ are moment estimators

Ans:

$X_1, X_2, \dots X_n \sim gama(\alpha, \lambda)$ $(\alpha, \lambda) \in (0, \infty)^2$

$\mathbb{E}[X_1^1] = \frac{\alpha}{\lambda} = m_1$

$\mathbb{E}[X_1^2] = var + (\mathbb{E}[X_1^1])^2$

$= \frac{\alpha}{\lambda^2} + \frac{\alpha^2}{\lambda^2} = m_2$

$\alpha = m_1 \lambda \Rightarrow \alpha^2 = m_1^2 \lambda^2$

$\alpha + \alpha^2 = m_2 \lambda^2$

$\alpha^2 = m_1^2 \lambda^2$

$\frac{\alpha + \alpha^2}{m_2} = \frac{\alpha^2}{m_1^2}$

$\frac{1}{\alpha} + 1 = \frac{m_2}{m_1^2}$

$\frac{1}{\alpha} = \frac{m_2 - m_1^2}{m_1^2}$

$\hat{\alpha} = \frac{m_1^2}{m_2 - m_1^2}$

now $\alpha = m_1 \lambda \Rightarrow \frac{m_1^2}{m_2 - m_1^2} = m_1 \hat{\lambda} \Rightarrow \hat{\lambda} = \frac{m_1}{m_2 - m_1^2}$

Exe: $X_1, X_2, \dots X_n \overset{iid}{\sim} Bin(k, p)$ $k \in \mathbb{N}$, $p \in (0,1)$ both unknown

then show that $\hat{k} = \dfrac{\bar{X}^2}{\bar{X} - \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2}$

$$\hat{p} = \frac{\bar{X}}{\hat{k}}$$

Ans: $X_1, X_2, \dots X_n \overset{iid}{\sim} Bin(k, p)$ $k \in \mathbb{N}$ $p \in (0,1)$

Now, $\mathbb{E}[X_1] = kp = m_1$

$\mathbb{E}[X_1^2] - (\mathbb{E}[X_1])^2 = kp(1-p)$

$\mathbb{E}[X_1^2] = kp(1-p) + k^2p^2 = m_2$

$\hat{p} = \dfrac{\bar{X}}{\hat{k}}$ and $p = \dfrac{m_1}{k}$ $\quad k\left(\dfrac{m_1}{k}\right)\left(1 - \dfrac{m_1}{k}\right) + k^2\dfrac{m_1^2}{k^2} = m_2$

$m_1 - \dfrac{m_1^2}{k} + m_1^2 = m_2$

$m_1 + m_1^2 - m_2 = \dfrac{m_1^2}{k}$

$k = \dfrac{m_1^2}{m_1 + m_1^2 - m_2}$

$m_2 - m_1^2 = \frac{1}{n}\Sigma(X_i^2) - (\bar{X})^2$

now $\dfrac{1}{n}\sum_{i=1}^{n}(X_i^2 - \bar{X})^2$

$= \dfrac{1}{n}\sum_{i=1}^{n}\left[X_i^2 + (\bar{X})^2 - 2X_i^2\bar{X}\right]$

$= \dfrac{1}{n}\sum_{i=1}^{n}(X_i)^2 + (\bar{X})^2 - 2(\bar{X})^2$

$= m_2 - m_1^2$

so, $\hat{k} = \dfrac{(\bar{X})^2}{\bar{X} - \frac{1}{n}\left(\sum_{i=1}^{n}(X_i - \bar{X})^2\right)}$

20th March:

Method of maximum likelihood:

$X_1, X_2, \dots X_n \overset{iid}{\sim} f(x|\theta)$ (pmf/pdf)

$\theta = (\theta_1, \theta_2, \dots \theta_k) \in \Theta \subseteq \mathbb{R}^k$

parameter space

$L(\underset{\sim}{x}|\theta) = \prod_{i=1}^{n} f(x_i|\theta)$ is the likelihood function

Defn: for each sample point $\underset{\sim}{x} \in \mathcal{X}$ let $\hat{\theta}(\underset{\sim}{x})$ be a parameter value at which $L(\underset{\sim}{x}|\theta)$ attains its maximum as a function of $\theta \in \Theta$ with $\underset{\sim}{x}$ held fixed

So, $\forall \underset{\sim}{x} \in \mathcal{X}$ let $\max_{\theta \in \Theta} L(\underset{\sim}{x}|\theta) = \tilde{L}(\underset{\sim}{x}|\hat{\theta}(\underset{\sim}{x}))$

then $\hat{\theta} = \hat{\theta}(\underset{\sim}{X})$ is called maximal likelihood estimator (MLE) of θ based on $\underset{\sim}{X}$

Note: ① most popular method of point estimator
② MLE always belongs to parameter space
③ MLE may not exist/may not be unique if exist
④ MLE has a few optimality properties

if $k=1$, then MLE is for $\theta = \theta_1$
for $\underset{\sim}{x} \in \mathcal{X}$

we will try to maximise log likelihood

$\tilde{L}(\underset{\sim}{x}|\theta) = \log L(\underset{\sim}{x}|\theta) = \sum_{i=1}^{n} \log f(x_i|\theta)$

as in many cases, the likelihood would be a smooth and strictly log-concave function of θ i.e $\tilde{L}(\underset{\sim}{x}|\theta)$ will be smooth and strictly concave function of $\theta \in \Theta$

$\frac{d^2}{d\theta^2} \tilde{L}(\underset{\sim}{x}|\theta) < 0 \quad \forall \theta \in \Theta$

So, if $\frac{d}{d\theta} \tilde{L}(\underset{\sim}{x}|\theta) = 0$ has a solution, then $\hat{\theta}(\underset{\sim}{x})$ is unique maximiser of $L(\underset{\sim}{x}|\theta)$

So, MLE of θ is $\hat{\theta} = \hat{\theta}(\underset{\sim}{x})$

Note: If $\tilde{L}(\underset{\sim}{x}|\theta)$ is not smooth, then we cannot use this method

Eg: $X_1, X_2, \dots X_n \overset{iid}{\sim} Poi(\lambda) \quad \lambda \in (0, \infty) = \Theta$

$f(x|\lambda) = e^{-\lambda} \frac{\lambda^x}{x!} \quad x \in \mathbb{N} \cup \{0\}$

$\log f(x|\lambda) = -\lambda + x \log \lambda - \log(x!)$

$\underset{\sim}{x} \in \mathcal{X}$

then

$\tilde{L}(\underset{\sim}{x}|\lambda) = \sum_{i=1}^{n} -\lambda + x_i \log \lambda - \log(x_i!)$

$$= -n\lambda + \log\lambda\left(\sum_{i=1}^{n} x_i^0\right) - \sum_{i=1}^{n} \log(x_i^0!)$$

$$\frac{d}{d\lambda}\mathcal{L}(\underset{\sim}{x}|\lambda) = -n + \sum_{i=1}^{n}\frac{x_i}{\lambda}$$

$$\frac{d^2}{d\lambda^2}\mathcal{L}(\underset{\sim}{x}|\lambda) = -\sum_{i=1}^{n}\frac{x_i}{\lambda}$$

now if $\sum_{i=1}^{n} x_i^0 > 0$ then $\hat{\lambda}(\underset{\sim}{x}) = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}$

if $\sum_{i=1}^{n} x_i^0 = 0$ then $x_i^0 = 0 \ \forall i$

$$L(\underset{\sim}{x}|\lambda) = e^{-n\lambda}$$

does not have a maxima for $\lambda \in (0,\infty)$

so in this case MLE does not exist

Ex: $x_1, x_2 \ldots x_n \overset{iid}{\sim} Exp(\lambda)$ with $E[X_i] = \frac{1}{\lambda}$ find MLE of λ

Ans: $f(x|\lambda) = \lambda e^{-\lambda x}$ now $\lambda \in (0,\infty)$

$$\log f(x|\lambda) = \log\lambda - \lambda x$$

$$\sum_{i=1}^{n}\log f(x_i|\lambda) = n\log\lambda - \lambda\sum_{i=1}^{n} x_i$$

$$\frac{d}{d\lambda}\mathcal{L}(\underset{\sim}{x}|\lambda) = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i$$

$$\frac{d^2}{d\lambda^2}\mathcal{L}(\underset{\sim}{x}|\lambda) = -\frac{n}{\lambda^2} < 0$$

so, $\hat{\lambda} = \frac{1}{\bar{x}}$

Eg: $x_1, x_2 \ldots x_n \overset{iid}{\sim} Unif[0,\theta]$

$$f(x|\theta) = \begin{cases} \frac{1}{\theta} & ; x \in [0,\theta] \\ 0 & ; \text{otherwise} \end{cases}$$

$$L(\underset{\sim}{x}|\theta) = \prod_{i=1}^{n} f(x_i|\theta) = \frac{1}{\theta^n}\, \mathbb{I}_{(x_{(1)} \geq 0)}\, \mathbb{I}_{(x_{(n)} \leq \theta)}$$

now $x_{(n)} \leq \theta$ enough to maximise $\frac{1}{\theta^n}$

$L(\underset{\sim}{x}|\theta) = \frac{1}{\theta^n}$ maximised when $\theta = x_{(n)}$

so, $\hat{\theta}(\underset{\sim}{x}) = x_{(n)}$ for fixed $\underset{\sim}{x} \in \mathcal{X} \Rightarrow \hat{\theta} = \hat{\theta}(\underset{\sim}{X}) = X_{(n)}$

Note: If $x_1, x_2 \ldots x_n \overset{iid}{\sim} Unif(0,\theta)$ with $f(x|\theta) = \begin{cases} 1/\theta & ; 0 < x < \theta \\ 0 & ; \text{otherwise} \end{cases}$

then MLE does not exist as $\theta > x_{(n)}$ does not maxima for $\frac{1}{\theta^n}$

so MLE can be very sensitive to choice of pdf

Exe: $X_1, X_2, \dots X_n \overset{iid}{\sim} N(\mu, 1)$ $\mu \in \Theta = \mathbb{R}$

$$f(x|\mu) = \frac{1}{\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2}} \quad x \in \mathbb{R}$$

$\underset{\sim}{x} \in \mathcal{X} = \mathbb{R}^n$

$$L(\underset{\sim}{x}|\theta) = \sum_{i=1}^{n} \log(x_i|\theta)$$

$$= n \log\left(\frac{1}{\sqrt{2\pi}}\right) - \sum_{i=1}^{n} \frac{(x_i-\mu)}{2}$$ smooth function wrt μ

$$\Rightarrow \frac{d}{d\mu} L(\underset{\sim}{x}|\mu) = \sum_{i=1}^{n} (x_i - \mu) = \sum_{i=1}^{n} x_i - n\mu$$

$$\frac{d^2}{d\mu^2} L(\underset{\sim}{x}|\mu) = -n < 0 \quad \forall \mu \in \mathbb{R} = \Theta$$

So, $\frac{d}{d\mu} L(\underset{\sim}{x}|\mu) = 0 \Rightarrow \sum_{i=1}^{n} x_i = n\mu$

$$\Rightarrow \hat{\mu}(\underset{\sim}{x}) = \frac{\sum_{i=1}^{n} x_i}{n}$$

$\Rightarrow \hat{\mu} = \bar{x}$ unique MLE

Now let $X_1, X_2, \dots X_n \overset{iid}{\sim} f(x|\theta)$ (pdf)

$f(x|\theta) = \frac{1}{2} e^{-|x-\theta|}$ $x \in \mathbb{R}$ double exponential density

where $\theta \in \mathbb{R} = \Theta$ fix $\underset{\sim}{x} \in \mathbb{R}^n = \mathcal{X}$ likelyhood, then

$$L(\underset{\sim}{x}|\theta) = \frac{1}{2} e^{-\sum_{i=1}^{n} |x_i - \theta|} \quad \theta \in \mathbb{R}$$

maximising this is same as minimising $\sum_{i=1}^{n} |x_i - \theta|$

Exe: (a) If n is odd then show that θ has unique MLE

$$\hat{\theta} = X_{(\frac{n+1}{2})}$$

(b) If n is even then show that any value b/w $X_{(\frac{n}{2})}$, $X_{(\frac{n}{2}+1)}$ (both inclusive) is MLE of θ

So, MLE not unique

Ans: If n is odd then $\sum_{i=1}^{n} |x_i - \theta| = \sum_{i=1}^{n} |x_{(i)} - \theta|$

then $\log L(\underset{\sim}{x}|\theta) = -\frac{1}{2} \sum_{i=1}^{n} |x_i - \theta|$

$$\frac{dL}{d\theta} \log L(\underset{\sim}{x}|\theta) = \frac{1}{2} \sum_{i=1}^{n} (-1) \frac{(x_i - \mu)}{|x_i - \mu|} = 0 \text{ or undefined}$$

$$\Rightarrow \sum_{i=1}^{n} \frac{(x_i - \mu)}{|x_i - \mu|} = 0 \text{ or undefined}$$

only when $x_{(\frac{n+1}{2})} = \mu$ trivial
when n is odd
otherwise $1, -1$ will come other
case for n even $x_{(\frac{n}{2})} < \mu < x_{(\frac{n}{2}+1)}$
we get 0
$x_{(\frac{n}{2})} = \mu$ or $x_{(\frac{n}{2}+1)} = \mu$
we get undefined

in both cases we will get minima (trivial)

as for $\mu \leftarrow$ we get $\frac{dL}{d\mu} > 0$, $\mu \rightarrow$ we get $\frac{dL}{d\mu} < 0$

K>1 :

Here, assuming $\log L(\underset{\sim}{x}|\theta)$ is differtiable, we have to solve the following:

$$\frac{\partial}{\partial \theta_1} L(\underset{\sim}{x}|\theta_1 \dots \theta_K) = 0$$

$$\frac{\partial}{\partial \theta_2} L(\underset{\sim}{x}|\theta_1, \dots \theta_K) = 0$$

$$\vdots$$

$$\frac{\partial}{\partial \theta_K} L(\underset{\sim}{x}|\theta_1 \dots \theta_K) = 0$$

and then we have to show μ is global maxima of

$$L(\underset{\sim}{x}|\theta_1 \dots \theta_K)$$

Eg: $X_1, X_2, \dots X_n \overset{iid}{\sim} N(\mu, \sigma^2)$ both unknown $\mu \in \mathbb{R}, \sigma \in (0, \infty)$

$$L(\underset{\sim}{x}|\mu, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2\right\} \quad \underset{\sim}{x} \in \mathcal{X} = \mathbb{R}^n$$

fix $\underset{\sim}{x} \in \mathbb{R}^n = \mathcal{X}$ $\tau = \sigma^2$

$$\tilde{L}(\underset{\sim}{x}|\mu, \sigma^2) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\tau) - \frac{1}{2\tau}\sum_{i=1}^{n}(x_i - \mu)^2$$

$$\frac{\partial \tilde{L}}{\partial \mu} = -\frac{1}{\tau}\sum_{i=1}^{n}(x_i - \mu) = 0 \Rightarrow \hat{\mu} = \bar{x}$$

$$\frac{\partial \tilde{L}}{\partial \tau} = -\frac{n}{2} + \frac{1}{2\tau^2}\sum_{i=1}^{n}(x_i - \mu)^2 = 0 \Rightarrow \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

now, we shall verify that they are global maximas

$$\sum_{i=1}^{n}(x_i-\mu)^2 = \sum_{i=1}^{n}(x_i-\bar{x})^2 + n(\bar{x}-\mu)^2$$
$$> \sum_{i=1}^{n}(x_i-\bar{x})^2 \quad \mu \neq \bar{x}$$

so, for any value of σ^2

$$L(\underset{\sim}{x}\,|\,\bar{x},\sigma^2) \geq L(\underset{\sim}{x}\,|\,\mu,\sigma^2)$$

equality for $\bar{x}=\mu$

Exe: show that $L(\underset{\sim}{x}\,|\,\bar{x},\sigma^2) \leq L(\underset{\sim}{x}\,|\,\bar{x}, \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2)$ $\forall \sigma^2$, and equality for $\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2$

Ans: if $\sigma^2 < \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2 \neq 0$ $(\because > \sigma^2)$

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

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

$$\frac{\partial \tilde{L}}{\partial \sigma} = -\frac{n}{2} + \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\bar{x})^2 > -\frac{n}{2}+\frac{n}{2} = 0$$

$\Rightarrow \tilde{L}$ is increasing $\Rightarrow \tilde{L}(\mu,\sigma^2) < \tilde{L}(\mu, \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2)$

so, $L(\underset{\sim}{x}\,|\,\hat{\mu},\hat{\sigma}^2) = L(\underset{\sim}{x}\,|\,\bar{x},\hat{\sigma}^2)$

$\geq L(\underset{\sim}{x}\,|\,\bar{x},\sigma^2)$ $(\because$ above$)$

$\geq L(\underset{\sim}{x}\,|\,\mu,\sigma^2)$ $\forall \mu,\sigma^2$

so, $\forall \mu,\sigma$, $\hat{\mu},\hat{\sigma}^2$ are global maxima of $L(\underset{\sim}{x}\,|\,\mu,\sigma^2)$, so MLE is

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

24th March :

Exe: $X_1, X_2, \dots X_n \overset{iid}{\sim} f(x|\beta,\lambda)$ where

$$f(x|\beta,\lambda) = \begin{cases} \lambda e^{-\lambda(x-\beta)} ; & x \geq \beta \\ 0 ; & x < \beta \end{cases}$$ → done down

Here $\lambda \in (0,\infty)$, $\beta \in \mathbb{R}$ are both unknown, what is MLE of (β,λ)

Exe: Let $X_1, X_2, \dots X_n$ be a random sample from a shifted pareto distribution with probability density function

$$f(x|\alpha,\beta) = \begin{cases} \alpha \beta^{\alpha} x^{-(\alpha+1)} ; & \text{if } x \geq \beta \\ 0 ; & \text{if } x < \beta \end{cases}$$ → done down

Here $\alpha \in (0,\infty)$, and $\beta \in (0,\infty)$ are both unknown parameters, find the MLE of α and β

Exe: Now that MLE, if exist was to be a function of any sufficient statistic using factorization criterion → done down

Exe: Let $X_1, X_2, \dots X_n \overset{iid}{\sim}$ Unif $[-\theta,\theta]$ with pdf

$$f(x|\theta) = \begin{cases} \frac{1}{2\theta} ; & x \in [-\theta,\theta] \\ 0 ; & \text{o.w} \end{cases}$$ → done down

find the MLE of θ

Bayesian method of point estimation :

Here the underlying parameter θ will no longer be fixed we will have some prior knowledge on the variation of the parameter and this knowledge will be quantified with the help of a continious distribution called prior distribution

let $\pi(\theta)$ be the pdf of the prior distribution

$\pi(\theta)$ is supported on parameter space Θ

as before $f(x|\theta)$ denotes the pmf/pdf of the observations in the sample

$X_1, X_2, \dots X_n$ were the underlying parameter take value θ

in other words,

$$X_1, X_2, \dots X_n \overset{iid}{\sim} f(x|\theta) \text{ pmf/pdf}$$

conditional on value of θ

Here, the conditional pmf/pdf of

$\underset{\sim}{X} = (X_1, X_2, \dots X_n)$ given θ is

$$f_{\underset{\sim}{X}|\theta}(\underset{\sim}{x}|\theta) = \prod_{i=1}^{n} f(x_i|\theta) \quad \underset{\sim}{x} = (x_1, x_2, \dots x_n) \in \mathbb{R}^n$$

now marginal pmf / pdf of $\underset{\sim}{X}$:

$$m(\underset{\sim}{x}) = \int_{\theta} f_{\underset{\sim}{X}|\theta}(\underset{\sim}{x}|\theta)\pi(\theta)d\theta$$

$$= \int_{\Theta} \left(\prod_{i=1}^{n} f(x_i|\theta)\right)\pi(\theta)d\theta$$

Def: (posterior distribution) It is the conditional distribution of θ given $\underset{\sim}{X}$

in this case, the pdf of posterior distribution is

$$\pi(\theta|\underset{\sim}{x}) = \frac{f_{\underset{\sim}{X}|\theta}(\underset{\sim}{x}|\theta)\pi(\theta)}{m(\underset{\sim}{x})}$$

$$\Rightarrow \pi(\theta|\underset{\sim}{x}) = \frac{\left(\prod_{i=1}^{n} f(x_i|\theta)\right)\pi(\theta)}{\int_{\Theta}\left(\prod_{i=1}^{n} f(x_i|\theta)\right)\pi(\theta)d\theta}$$

the posterior distribution is used to carry out statistical interfence on θ using the bayesian approach

Def: Bayes estimate of θ is the mean of the posterior distribution, i.e

$$\hat{\theta}_{Bayes} = \mathbb{E}[\theta|\underset{\sim}{X}]$$

suppose T is a sufficient statistic for θ then by factorization criterion, $\exists$ fuction $h(\underset{\sim}{x})$ and $g(t|\theta)$ s.t the product

$$\prod_{i=1}^{n} f(x_i|\theta) = h(\underset{\sim}{x})g(T(\underset{\sim}{x})|\theta)$$

then $\pi(\theta|\underset{\sim}{x}) = \frac{\left(\prod_{i=1}^{n} f(x_i|\theta)\right)\pi(\theta)}{\int_{\theta}\left(\prod_{i=1}^{n} f(x_i|\theta)\right)\pi(\theta)d\theta}$

$$= \frac{h(\underset{\sim}{x})g(T(\underset{\sim}{x})|\theta)\pi(\theta)}{h(\underset{\sim}{x})\int_{\Theta} g(T(\underset{\sim}{x})|\theta)\pi(\theta)d\theta}$$

$$\Rightarrow \pi(\theta|\underset{\sim}{x}) = \frac{g(T(\underset{\sim}{x})|\theta)\pi(\theta)}{\int_{\Theta} g(T(\underset{\sim}{x})|\theta)\pi(\theta)d\theta}$$

Note: $\pi(\theta|x)$ depends through x only through $T(x)$, so the bayes estimate of θ is a function of T

so, it is enough to compute Bayes estimate using T

so, $\hat{\theta}_{Bayes} = E[\theta|T]$

Eg: let $X_1, X_2, \ldots X_n$ random sample from Ber(p) dist^n

we assume that prior distribution on p is Beta(α, β) with α, β known

now, we want to find Bayes statistic

$$T(\underset{\sim}{X}) = \sum_{i=1}^{n} X_i \quad (\because \text{sufficient from previous})$$

is sufficient for Ber(p)

let $Y = \sum_{i=1}^{n} X_i$ is sufficient statistic for p

we will carry out computation based on $Y \sim Bin(n, p)$ conditionally on p

In this case, the conditional pmf

$$f_{Y|p}(y|p) = \binom{n}{y} p^y (1-p)^{n-y} \qquad y \in \{0, 1, \ldots n\}$$

prior pdf $\pi(p) = \dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1} \qquad p \in (0,1) = \Theta$

the marginal pmf of Y is

$$m(y) = \int_{(0,1)} f_{Y|p}(y|p)\,\pi(p)\,dp$$

$$= \int_0^1 \binom{n}{y} p^y (1-p)^{n-y} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1}\,dp$$

$$= \int_0^1 \binom{n}{y} p^{(y+\alpha-1)} (1-p)^{n-y+\beta-1} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\,dp$$

$$= \binom{n}{y} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\alpha+y)\Gamma(\beta+n-y)}{\Gamma(\alpha+\beta)}$$

$$= \binom{n}{y} \frac{\Gamma(\alpha+y)\Gamma(\beta+n-y)}{\Gamma(\alpha)\Gamma(\beta)} \qquad y \in \{0, 1, \ldots n\}$$

so, the posterior density

$$\pi(p|y) = \frac{f_{Y|p}(y|p)\,\pi(p)}{m(y)}$$

$$= \frac{\binom{n}{y} p^{(y+\alpha-1)} (1-p)^{n-y+\beta-1} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}}{\binom{n}{y} \frac{\Gamma(\alpha+y)\Gamma(\beta+n-y)}{\Gamma(\alpha+\beta)}}$$

for fixed $y \in \{0, \dots n\}$

$$\pi(p|y) = \frac{p^{(y+\alpha-1)} (1-p)^{(n-y+\beta-1)}}{\beta(\alpha+y, \beta+n-y)}$$

→ posterior distribution of $Y=y$ is $\text{Beta}(\alpha+y, n-y+\beta)$

mean of $\pi(p|y) = \mathbb{E}[p|Y=y]$

$$= \frac{\alpha+y}{\alpha+\beta+n}$$

$$\Rightarrow \mathbb{E}[p|Y] = \frac{\alpha+Y}{\alpha+\beta+n}$$

so, $\hat{p}_{\text{Bayes}} = \dfrac{\alpha + \sum_{i=1}^{n} x_i}{\alpha+\beta+n}$

Note: above can be written as

$$\hat{p}_{\text{Bayes}} = \frac{n\bar{X}}{\alpha+\beta+n} + \frac{\alpha+\beta}{\alpha+\beta+n}\left(\frac{\alpha}{\alpha+\beta}\right)$$

$$= \left(\frac{n}{\alpha+\beta+n}\right)\bar{X} + \left(\frac{\alpha+\beta}{\alpha+\beta+n}\right)\left(\frac{\alpha}{\alpha+\beta}\right)$$

where $\bar{X}$ is classical estimator of p

so, $\hat{p}_{\text{Bayes}}$ is a convex combination of the classical estimator $\bar{X}$ of p and the prior mean

as n becomes larger, the posterior distribution varies

Note: In general, the Bayes estimator has to be computed by the following formula:

$$\hat{\theta}_{\text{Bayes}}(\underset{\sim}{x}) = \int_{\Theta} \theta \pi(\theta|\underset{\sim}{x}) d\theta \quad \text{for each } \underset{\sim}{x} \in \mathbb{R}^n$$

Note: Bayes estimator will not exist if the posterior distribution does not have a finite mean

Ex: suppose $X_1, X_2, \dots X_n$ is a random sample from $N(\theta, \sigma_0^2)$ dist^n with σ_0^2 known

Assume that the prior dist^n on θ is $N(\mu, \tau^2)$ with μ, τ^2 known → done down

find bayes estimate of θ

Ex: Suppose $X_1, X_2, \ldots X_n$ is a random sample from poi(λ) distⁿ. Assume that the prior distⁿ on λ is gamma(α, β) distⁿ (with α, β known), find the Bayes estimator of λ → done down

Methods of evaluating point estimators:

$X_1, X_2, \ldots X_n \overset{iid}{\sim} f(x|\theta)$ pmf/pdf of data

$\theta \in \Theta \subseteq \mathbb{R}$ (parameter space)
unknown but fixed

Suppose $T = T(\underset{\sim}{X})$ is a point estimator of θ

Defⁿ: (Mean Square error) The MSE of point estimator T of θ is a fution of θ defined by

$$\mathbb{E}_\theta[(T-\theta)^2]$$

Note: MSE measures average squared difference between the estimator T and the parameter θ, and so, it is a reasonable measure of performance of a point estimator

Ex: $X_1, X_2, \ldots X_n \overset{iid}{\sim} f(x|\beta,\lambda)$ where

$$f(x|\beta,\lambda) = \begin{cases} \lambda e^{-\lambda(x-\beta)} & ; x \geq \beta \\ 0 & ; x < \beta \end{cases}$$

Here $\lambda \in (0,\infty)$, $\beta \in \mathbb{R}$ are both unknowns, what is MLE of (β,λ)

Ans:

$$f(x|\beta,\lambda) = \begin{cases} \lambda e^{-\lambda(x-\beta)} & ; x \geq \beta \\ 0 & ; x < \beta \end{cases}$$

$\lambda \in (0,\infty)$, $\beta \in \mathbb{R}$

$$L(\underset{\sim}{x}|\beta,\lambda) = \lambda^n e^{-\lambda(\sum_{i=1}^n x_i - \beta)} \quad \underset{\sim}{x} \in (\beta,\infty)^n$$

if λ is known, then

$\beta \to e^{\lambda(n\beta - \sum_{i=1}^n x_i)}$

strictly inc wrt β and $x_i \in (\beta,\infty)$

$\Rightarrow x_i > \beta \ \forall i \in \{1,\ldots n\}$

So, likelihood maximised when $\beta = x_{(1)}$

$$L(\underset{\sim}{x}|\hat{\beta},\lambda) = L(\underset{\sim}{x}|x_{(1)},\lambda) \geq L(\underset{\sim}{x}|\beta,\lambda) \quad \text{— (1)}$$

fixing β then putting $\beta = x_{(1)}$

$$l' = \log(L(\underset{\sim}{x}|\beta,\lambda)) = n\log\lambda - \lambda\sum_{i=1}^n (x_i - x_{(1)})$$

$$\Rightarrow \frac{\partial l'}{\partial \lambda} = \frac{n}{\lambda} - \sum_{i=1}^n (x_i - x_{(1)})$$

$$\frac{\partial^2 L'}{\partial \lambda^2} = -\frac{n}{\lambda^2} < 0 \quad \forall \lambda$$

$\Rightarrow$ max when $\frac{n}{\hat{\lambda}} - \sum_{i=1}^{n}(x_i - x_{(1)}) = 0$

$$\Rightarrow \hat{\lambda} = \frac{n}{\sum_{i=1}^{n}(x_i - x_{(1)})}$$

$$\hat{\lambda} = \frac{1}{\bar{x} - x_{(1)}}$$

now, $L(\underset{\sim}{x} | \hat{\beta}, \hat{\lambda}) \geq L(\underset{\sim}{x} | x_{(1)}, \hat{\lambda})$

$\geq L(\underset{\sim}{x} | x_{(1)}, \lambda)$

$\geq L(\underset{\sim}{x} | \beta, \lambda)$ $(\because (1))$

$\Rightarrow$ MLE of β and λ are

$\hat{\beta} = X_{(1)}$

$\hat{\lambda} = \frac{1}{\bar{X} - X_{(1)}}$

Ex: Let $X_1, X_2, \ldots X_n$ be a random sample from a suited pareto distribution with probability density function

$$f(x | \alpha, \beta) = \begin{cases} \alpha \beta^{\alpha} x^{-(\alpha+1)} & ; \text{if } x \geq \beta \\ 0 & ; \text{if } x < \beta \end{cases}$$

Here $\alpha \in (0, \infty)$, and $\beta \in (0, \infty)$ are both unknown parameters, find the MLE of α and β

Ans:

$$L(\underset{\sim}{x} | \alpha, \beta) = \begin{cases} \alpha^n \beta^{n\alpha} (x_1 x_2 \ldots x_n)^{-(\alpha+1)} & ; x_i \geq \beta \ \forall i \in \{1 \ldots n\} \\ 0 & ; \text{otherwise} \end{cases}$$

fixing α: $\beta^{n\alpha} (x_1 \ldots x_n)^{-(\alpha+1)}$ inc

$\Rightarrow$ max when $\beta = x_{(1)}$

so, $L(\underset{\sim}{x} | \alpha, \hat{\beta}) \geq L(\underset{\sim}{x} | \alpha, \beta)$

i.e $L(\underset{\sim}{x} | \alpha, x_{(1)}) \geq L(\underset{\sim}{x} | \alpha, \beta)$

now fixing $\beta = x_{(1)}$ then

$$L = L(\underset{\sim}{x} | \alpha, x_{(1)}) = \alpha^n \beta^{n\alpha} (x_1 \ldots x_n)^{-(\alpha+1)}$$

$$L' = \log L = n \log \alpha + n\alpha \log \beta - (\alpha+1) \log (x_1 \ldots x_n)$$

$$\frac{dL'}{d\alpha} = \frac{n}{\alpha} + n\log(x_{(1)}) - \log(x_1 \dots x_n)$$

$$\frac{d^2L'}{d\alpha^2} = -\frac{n}{\alpha^2} + 0 < 0$$

$$\Rightarrow \frac{n}{\alpha} + n\log(x_{(1)}) - \log(x_1 \dots x_n) = 0$$

$$\Rightarrow \frac{n}{\alpha} = \log(x_1 \dots x_n) - n\log(x_{(1)})$$

$$\Rightarrow \hat{\alpha} = \frac{n}{\log(x_1 \dots x_n) - n\log(x_{(1)})}$$

$$L(\underset{\sim}{x} | \hat{\alpha}, \hat{\beta}) \geq L(\underset{\sim}{x} | \alpha, \hat{\beta}) \geq L(\underset{\sim}{x} | \alpha, \beta)$$

Exe: now that MLE, if exist was to be a function of any sufficient statistic using factorization criterion

Ans: $U[\theta, \theta+1]$ uniform dist

$(X_{(1)}, X_{(n)})$ is minimal sufficient statistic is known but

MLE for $L(\underset{\sim}{x} | \theta) = 1$

every θ maximizes

given $x_{(1)} \geq \theta \quad x_{(n)} - 1 \leq \theta \leq x_{(1)}$

$$\hat{\theta} = \frac{x_{(n)} - 1 + x_{(1)}}{2}$$

then $L(\underset{\sim}{x} | \hat{\theta}(\underset{\sim}{x})) = 1$

$\Rightarrow \hat{\theta}$ is MLE but not sufficient

Exe: let $x_1, x_2, \dots x_n \overset{iid}{\sim}$ Unif $[-\theta, \theta]$ with pdf

$$f(x|\theta) = \begin{cases} \frac{1}{2\theta} ; & x \in [-\theta, \theta] \\ 0 ; & \text{o.w} \end{cases}$$

find the MLE of θ

Ans:

$$L(\underset{\sim}{x} | \theta) = \begin{cases} \left(\frac{1}{2\theta}\right)^n ; & x_{(1)} \geq -\theta, \ x_{(n)} \leq \theta \\ 0 ; & \text{otherwise} \end{cases}$$

$$\log(L(\underset{\sim}{x} | \theta)) = \begin{cases} -n\log(2\theta) ; & x_{(1)} \geq -\theta, x_{(n)} \leq \theta \\ 0 ; & \text{otherwise} \end{cases}$$

$$\frac{d\log(L(\underset{\sim}{x} | \theta))}{d\theta} = -\frac{n}{2\theta} < 0 \rightarrow \text{always decreasing if } \theta \in (0, \infty)$$

$$-x_{(1)} \leq \theta, \ x_{(n)} \leq \theta$$

$\theta > x_{(n)}$ and $-x_{(1)}$

$\Rightarrow \theta > \max\left(x_{(n)}, -x_{(1)}\right)$

$\Rightarrow \hat{\theta} = \max\left(x_{(n)}, -x_{(1)}\right)$

MLE is $\hat{\theta}(\underset{\sim}{x}) = \max\left(x_{(n)}, -x_{(1)}\right)$

Ex: suppose $x_1, x_2, \ldots x_n$ is a random sample from $N(\theta, \sigma_0^2)$ dist^n with σ_0^2 known

Assume that the prior dist^n on θ is $N(\mu, \tau^2)$ with μ, τ^2 known

find bayes estimate of θ

Ans:

$x_1, x_2, \ldots x_n \overset{iid}{\sim} N(\theta, \sigma_0^2)$ (σ_0^2 known)

$$\pi(\theta) = \frac{1}{\sqrt{2\pi\tau^2}} e^{-\frac{(\theta-\mu)^2}{2\tau^2}}$$

now $\bar{x}$ is minimal for θ

$$\bar{x} \sim N\left(\theta, \frac{\sigma_0^2}{n}\right)$$

$$\text{now } \pi(\theta) f(\bar{x}|\theta) \propto e^{-\frac{(\theta-\mu)^2}{2\tau^2}} e^{-\frac{(\bar{x}-\theta)^2}{2\sigma_0^2/n}}$$

$$\pi(\theta|\bar{x}) \propto e^{-\frac{(\theta-\mu)^2}{2\tau^2} - \frac{(\bar{x}-\theta)^2}{2\sigma_0^2/n}}$$

now $\frac{(\theta-\mu)^2}{2\tau^2} + \frac{(\bar{x}-\theta)^2}{\frac{2\sigma_0^2}{n}}$

$$= \frac{1}{2}\left[\frac{\theta^2}{\tau^2} + \frac{\mu^2}{\tau^2} - \frac{2\theta\mu}{\tau^2} + \frac{\bar{x}^2}{\frac{\sigma_0^2}{n}} + \frac{\theta^2}{\frac{\sigma_0^2}{n}} - \frac{2\bar{x}\theta}{\frac{\sigma_0^2}{n}}\right]$$

now $\theta^2\left[\underbrace{\frac{1}{2\tau^2} + \frac{1}{\frac{2\sigma_0^2}{n}}}\right] - 2\theta\left[\frac{\mu}{2\tau^2} + \frac{\bar{x}}{\frac{2\sigma_0^2}{n}}\right] + \frac{1}{2\tau^2}\mu^2 + \frac{\bar{x}^2}{\frac{2\sigma_0^2}{n}}$

$$\frac{1}{2\tau^2} + \frac{n}{2\sigma_0^2}$$

$$= \frac{\sigma_0^2 + n\tau^2}{2\tau^2\sigma_0^2}$$

$$\Rightarrow \underbrace{\frac{1}{\left(\frac{2\tau^2\sigma_0^2}{\sigma_0^2+n\tau^2}\right)}}_{2c^2}\left[\theta^2 - 2\theta\underbrace{\left(\frac{\mu}{2\tau^2} + \frac{\bar{x}}{\frac{2\sigma_0^2}{n}}\right)\left(\frac{2\tau^2\sigma_0^2}{\sigma_0^2+n\tau^2}\right)}_{d} + C\right]$$

so, $\pi(\theta|\bar{x}) \propto e^{-\frac{1}{2c^2}[(\theta-d)^2]} \sim N(d, c^2)$

then $\hat{\theta}_{Bayes}(\bar{x}) = \int_\theta \theta\pi(\theta|\bar{x})\,d\theta = d = \left(\frac{\mu}{2\tau^2} + \frac{\bar{x}}{\frac{2\sigma_0^2}{n}}\right)\left(\frac{2\tau^2\sigma_0^2}{\sigma_0^2+n\tau^2}\right)$

(mean of normal)

$$\Rightarrow \hat{\theta}_{Bayes}(\bar{x}) = \left(\frac{\mu}{\sigma\tau^2} + \frac{n\bar{x}}{\sigma\sigma_0^2}\right)\left(\frac{\sigma\tau^2\sigma_0^2}{\sigma_0^2+n\tau^2}\right)$$

$$= \left(\frac{\mu\sigma_0^2+n\tau^2\bar{x}}{\sigma\tau^2\sigma_0^2}\right)\left(\frac{\sigma\tau^2\sigma_0^2}{\sigma_0^2+n\tau^2}\right)$$

$$\hat{\theta}_{Bayes}(x_1,x_2,..x_n) = \frac{\mu\sigma_0^2+\tau^2\left(\sum_{i=1}^{n}x_i\right)}{\sigma_0^2+n\tau^2}$$

ex: suppose $X_1, X_2, \ldots X_n$ is a random sample from poi(λ) dist^n Assume that the prior dist^n on λ is gamma(α, β) dist^n (with α, β known), find the Bayes estimator of λ

Ans:

$X_1, X_2, \ldots X_n \overset{iid}{\sim} poi(\lambda)$ $\pi(\theta) \sim$ Gamma(α, β)

now $T(\underset{\sim}{X}) = \sum_{i=1}^{n} X_i \sim Poi(n\lambda)$ is sufficient for λ we know

$$\pi(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)}\theta^{\alpha-1}e^{-\beta\theta} \qquad y = \sum_{i=1}^{n}x_i$$

then $\pi(\theta|y) \propto \pi(\theta) \underbrace{f(y|\theta)}_{Poi(n\lambda)}$

$$\propto \frac{\beta^\alpha\theta^{\alpha-1}}{\Gamma(\alpha)}e^{-\beta\theta}\frac{e^{-n\lambda}(n\lambda)^y}{y!}$$

$\theta = \lambda$ and so $\pi(\lambda|y) \propto \frac{\beta^\alpha}{\Gamma(\alpha)}\lambda^{\alpha-1}e^{-\beta\lambda}\frac{e^{-n\lambda}(n\lambda)^y}{y!}$

$$\propto \lambda^{\alpha-1+y}e^{-(\beta+n)\lambda}$$

so, like gamma $(\alpha+y, \beta+n)$ distribution

i.e $\pi(\lambda|y) \sim$ Gamma($\alpha+y, \beta+n$)

now, $\hat{\lambda}_{Bayes}(y) = \frac{\alpha+y}{\beta+n}$

$$\hat{\lambda}_{Bayes}(\underset{\sim}{x}) = \frac{\alpha+\sum_{i=1}^{n}x_i}{\beta+n}$$

27th March :

Quiz-5, 11:45 - 12:15, LH102

Methods of evaluation of Estimators :

$X_1, X_2, \ldots X_n \overset{iid}{\sim} f(x|\theta)$ pmf/pdf

$\theta \in \Theta \subseteq \mathbb{R}$

Defn : $MSE_\theta(T) = \mathbb{E}_\theta[(T-\theta)^2]$

Defn : The bias of a point estimator T of θ is

$$Bias_\theta(T) = \mathbb{E}_\theta[T] - \theta$$

$Bias_\theta(T) < 0 \Rightarrow T$ underestimates θ

$Bias_\theta(T) > 0 \Rightarrow T$ overestimates θ

Defn : An estimator T of θ is called unbiased if $\mathbb{E}_\theta[T] = \theta \ \forall \theta \in \Theta$

Note : T is unbiased for $\theta \Leftrightarrow Bias_\theta(T) = 0 \ \forall \theta \in \Theta \Leftrightarrow$ on the avg T is neither underestimates nor overestimates θ

Now, $MSE_\theta[T] = \mathbb{E}[(T-\theta)^2]$

$= \mathbb{E}[\{(T - \mathbb{E}[T]) + (\mathbb{E}[T] - \theta)\}^2]$

$= \mathbb{E}[(T - \mathbb{E}[T])^2] + \mathbb{E}[(\mathbb{E}[T] - \theta)^2]$

$+ 2\,\mathbb{E}[T - \mathbb{E}[T]][\mathbb{E}[T] - \theta]$

$= \mathbb{E}[(T - \mathbb{E}[T])^2] + (\mathbb{E}[T] - \theta)^2$

$$MSE_\theta(T) = Var_\theta(T) + (Bias_\theta(T))^2$$

($Var_\theta(T)$: measure of precision; $Bias_\theta(T)$: measure of accuracy)

Note : There is always a tradeoff b/w variance and bias

Var ↑, Bias ↓, new Var ↓, Bias ↑

There is a tradeoff b/w the precision and accuracy of the point estimator this is called the bias-variance tradeoff

people give prefer to accuracy, we need to make it unbiased, and among that group we want to reduce precision as much as possible

Eg : $X_1, X_2, \ldots X_n \overset{iid}{\sim} N(\mu, \sigma^2)$ ($n \geq 2$, μ, σ^2 are both unknown)

$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is unbiased for μ as

$\mathbb{E}[\bar{X}] = \mu \ \forall \mu$

$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ is unbiased for σ^2 as $\mathbb{E}[S^2] = \sigma^2$

$$MSE(\bar{X}) = var(\ ) + \underbrace{(Bias(\ ))^2}_{0}$$

$$= var(\bar{X})$$

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

$$MSE(S^2) = var(S^2) + \underbrace{(Bias(S^2))^2}_{0}$$

$$= var(S^2)$$

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1} = gamma\left(\frac{n-1}{2}, \frac{1}{2}\right)$$

$$var\left(\frac{(n-1)S^2}{\sigma^2}\right) = \frac{n-1}{2} = 2(n-1)$$

$$\Rightarrow \frac{(n-1)^2}{\sigma^4} var(S^2) = 2(n-1)\left(\frac{1}{2}\right)^2$$

$$\Rightarrow MSE(S^2) = \frac{2\sigma^4}{n-1}$$

automatically, we can use another estimator of σ^2 as MLE

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

$$E[\hat{\sigma}^2_{MLE}] = \frac{n-1}{n}\sigma^2$$

$$Bias(\hat{\sigma}^2_{MLE}) = \frac{n-1}{n}\sigma^2 - \sigma^2 = -\frac{\sigma^2}{n}$$

$$\Rightarrow \left(Bias(\hat{\sigma}^2_{MLE})\right)^2 = \frac{\sigma^4}{n^2}$$

$$var(\hat{\sigma}^2_{MLE}) = var\left(\frac{n-1}{n}S^2\right)$$

$$= \frac{(n-1)^2}{(n)^2}\frac{2\sigma^4}{n-1} = \frac{(n-1)2\sigma^4}{n^2}$$

$$MSE_\theta(\hat{\sigma}^2_{MLE}) = \frac{(2n-1)\sigma^4}{n^2}$$

now $(2n-1)(n-1) < 2(n^2)$

$$\Rightarrow MSE_\theta(\hat{\sigma}_{MLE}) < MSE_\theta(S^2)$$

So, from above example we see that even with biased estimator, it has lower MSE

Therefore, by compromising a bit on bias, we can gain a lot more

in the variance leading to lower MSE

Ex: suppose $X_1, X_2, \ldots X_n \overset{iid}{\sim} \text{Unif}(0,\theta)$, consider the estimators of θ of the form

$$T_c := c X_{(n)}, \quad c > 0$$

(a) find the values of c for which T_c is an unbiased estimator

(b) consider another estimator W of θ defined by

$W := 2\bar{X}$, show that W is also unbiased for θ → done down

(c) Among the unbiased estimators attained in (a) and (b) above, which one would you prefer, and why?

(d) Find a value of c for which T_c has the minimum MSE, is this estimator unbiased for θ?

uniformly minimum variance unbiased estimator (UMVUE):

This is also called as best unbiased estimator

Def: An estimator T of θ is called UMVUE if

(i) T is unbiased for θ, i.e $E_\theta[T] = \theta \ \forall \theta \in \Theta$ and

(ii) for any unbiased estimator W of θ

$$\text{Var}_\theta(T) \leq \text{Var}_\theta(W) \ \forall \theta \in \Theta$$

Note: UMVUE may not exist

Theorem: UMVUE, if it exist is unique

proof: Suppose W and T are two UMVUE

if $P_\theta[T = W] = 1 \ \forall \theta \in \Theta$, then we are done

consider the estimator $S = \frac{W+T}{2}$

clearly S is unbiased for θ

$$\text{also, } \text{Var}_\theta(S) = \text{Var}_\theta\left(\frac{1}{2}W + \frac{1}{2}T\right)$$

$$= \frac{1}{4}\text{Var}_\theta(W) + \frac{1}{4}\text{Var}_\theta(T) + \frac{1}{2}\text{Cov}_\theta(W,T)$$

$$\leq \frac{1}{4}\text{Var}_\theta(W) + \frac{1}{4}\text{Var}_\theta(T) + \frac{1}{2}\left(\text{Var}_\theta(W)\text{Var}_\theta(T)\right)^{\frac{1}{2}}$$

(∵ CS inequality)

$$\text{Var}_\theta(T) \leq \text{Var}_\theta(W) \ \forall \theta \in \Theta$$
$$\& \ \text{Var}_\theta(T) \geq \text{Var}_\theta(W) \Rightarrow \text{Var}_\theta(T) = \text{Var}_\theta(W)$$

then $\text{Var}_\theta\left(\frac{1}{2}T+\frac{1}{2}W\right) \leq \frac{1}{4}\text{Var}_\theta(W)+\frac{1}{4}\text{Var}_\theta(W)$

$$+\frac{1}{2}\text{Var}_\theta(W)$$

$$\Rightarrow \text{Var}_\theta\left(\frac{1}{2}T+\frac{1}{2}W\right) \leq \text{Var}_\theta(W) = \text{Var}_\theta(V)$$

$$\Rightarrow \text{Var}_\theta\left(\frac{1}{2}T+\frac{1}{2}W\right) = \text{Var}_\theta(W) = \text{Var}_\theta(V)$$

$$\Rightarrow \text{Var}_\theta(S) \leq \text{Var}_\theta(T) = \text{Var}_\theta(W)$$

if above inequality is strict for some $\theta \in \Theta$, then W will no longer be the UMVUE

$\Rightarrow$ the $\leq$ in CS should be equal to not have the contradiction

$\Rightarrow \text{Cov}(W,T) = \sqrt{\text{Var}_\theta(W)\text{Var}_\theta(T)} = \text{Var}_\theta(W)$ — ①

$\Rightarrow \exists\ a(\theta)>0$ and $b(\theta)\in\mathbb{R}$ s.t $a(\theta), b(\theta)$ functions and $\forall\ \theta\in\Theta$, $P_\theta[T = a(\theta)W + b(\theta)] = 1$

$\Rightarrow \text{Cov}_\theta(W,T) = \text{Cov}_\theta(W, a(\theta)W+b(\theta)) = a(\theta)\text{Var}_\theta(W)$

$= \text{Var}_\theta(W)\ \forall\theta\in\Theta\ (\because ①)$

as $\text{Cov}(W,T) = \mathbb{E}[(W-\theta)(T-\theta)]$

$\Rightarrow a(\theta) \equiv 1$ — ②

and $\mathbb{E}_\theta[T] = \mathbb{E}_\theta[W] = \theta\ \forall\theta\in\Theta$

$\Rightarrow a(\theta)\mathbb{E}_\theta[W] + b(\theta) = \mathbb{E}_\theta[W]$ $(\because ②)$

$\Rightarrow b(\theta) \equiv 0\ \forall\ \theta\in\Theta$

so, $P_\theta[T=W] = 1\ \forall\theta\in\Theta$

Finding UMVUE:

Theorem: (Rao-Blackwell) Let W be any unbiased estimator for θ and let T be a sufficient statistic for θ.

Let $\phi(T) = \mathbb{E}[W|T]$ (free of θ), then $\phi(T)$ is another unbiased estimator of θ and it satisfies

$$\text{Var}_\theta(\phi(T)) \leq \text{Var}_\theta(W)\ \forall\ \theta\in\Theta$$

Ex: suppose $X_1, X_2, \ldots X_n \overset{iid}{\sim} \text{Unif}(0,\theta)$, consider the estimators of θ of the form

$$T_c := cX_{(n)},\ c>0$$

(a) find the values of c for which T_c is an unbiased estimator

(b) consider another estimator W of θ defined by

$W := 2\bar{X}$, show that W is also unbiased for θ

(c) Among the unbiased estimators obtained in (a) and (b) above, which one would you prefer, and why?

(d) Find a value of c for which T_c has the minimum MSE, is this estimator unbiased for θ?

Ans:

$X_1, X_2, \dots X_n \overset{iid}{\sim} \text{unif}(0, \theta)$

$T_c = c X_{(n)} \quad c > 0$

(a) now $P(X_{(n)} \le x) = P(x_1 \le x, x_2 \le x, \dots x_n \le x)$

$\uparrow$ max

$= \left(\left(\frac{1}{\theta}\right) x\right)^n$

$F_{X_{(n)}}(x) = \left(\frac{x}{\theta}\right)^n$

$f_{X_{(n)}}(x) = n(x)^{n-1}\left(\frac{1}{\theta}\right)^n \qquad x \in (0, \theta)$

now, $X_{(n)} \sim f_{X_{(n)}}$ as above

$\mathbb{E}[X_{(n)}] = \int_0^\theta x f_{X_{(n)}}(x)\, dx$

$= \int_0^\theta x\, n(x)^{n-1} \left(\frac{1}{\theta}\right)^n dx$

$= n\left(\frac{\theta^{n+1}}{n+1}\right)\left(\frac{1}{\theta}\right)^n$

$\mathbb{E}[X_{(n)}] = \frac{n}{n+1}\theta$

now $\mathbb{E}[T_c] = c\left(\frac{n}{n+1}\theta\right) = \theta \quad (\because \text{unbiased})$

$\Rightarrow c = \frac{n+1}{n}$

(b) now $\bar{X} = \frac{\sum_{i=1}^{n} x_i}{n}$

$\Rightarrow \mathbb{E}[\bar{X}] = \frac{n\,\mathbb{E}[X_1]}{n} = \mathbb{E}[X_1] = \int_0^\theta \left(\frac{1}{\theta}\right) x\, dx = \frac{\theta}{2}$

$\Rightarrow \mathbb{E}[2\bar{X}] = 2\,\mathbb{E}[\bar{X}] = \theta$

so, $W := 2\bar{X}$ is unbiased

(c) now, $MSE_\theta(T_c) = \underset{\theta}{var}(T_c) + \left(\underset{\theta}{Bias}(T_c)\right)^2$

$MSE_\theta(T_c) = var_\theta(T_c) = \mathbb{E}[T_c^2] - \left(\mathbb{E}[T_c]\right)^2$ $\rightsquigarrow 0$

$$var(c\,X_{(n)}) = c^2 var(X_{(n)})$$

$$var(X_{(n)}) = \mathbb{E}[X_{(n)}^2] - \left(\mathbb{E}[X_{(n)}]\right)^2$$

$$= \int_0^\theta x^2 \frac{n x^{n-1}}{\theta^n} dx - \left(\frac{n}{n+1}\theta\right)^2$$

$$= \frac{n}{n+2}\theta^2 - \frac{n^2}{(n+1)^2}\theta^2$$

$$= \left(\frac{n(n+1)^2 - n^2(n+2)}{(n+2)(n+1)^2}\right)\theta^2$$

$$= \left(\frac{n(n^2+2n+1) - n^3 - 2n^2}{(n+2)(n+1)^2}\right)\theta^2$$

$$= \left(\frac{\cancel{n^3} + 2n^2 + n - \cancel{n^3} - 2n^2}{(n+2)(n+1)^2}\right)\theta^2$$

$$var(X_{(n)}) = \frac{n}{(n+2)(n+1)^2}\theta^2$$

so, $Var(T_c) = \left(\frac{\cancel{n+1}}{n}\right)^2 \frac{n}{(n+2)\cancel{(n+1)}^2}\theta^2$

$$MSE_\theta(T_c) = \frac{1}{n(n+2)}\theta^2$$

now, $var(2\bar{X}) = 4\,var(\bar{X})$

$$= \frac{4}{n^2} var\left(\sum_{i=1}^n X_i\right)$$

$$= \frac{4}{n^2} \sum_{i=1}^n var(X_i^0)$$

$$= \frac{4}{n^2} \times n\, var(X_1)$$

$$var(X_1) = \int_0^\theta x^2 \frac{1}{\theta} dx - \left(\int_0^\theta x \frac{1}{\theta} dx\right)^2$$

$$= \frac{1}{3}\theta^2 - \frac{1}{4}\theta^2$$

$$= \frac{1}{12}\theta^2$$

so, $MSE_\theta(W) = \frac{\cancel{4}}{n} \times \frac{1}{3\cancel{12}}\theta^2 = \frac{\theta^2}{3n}$

now as, $n^2 + 2n > 3n \quad \forall n \in \mathbb{N}$

$MSE_\theta(W) \geq MSE_\theta(T_c)$ for $c = \frac{n+1}{n}$

so, T_c is better estimate

(a) $MSE_\theta(T_c) = c^2 Var(X_{(n)}) + \underbrace{(c\,\mathbb{E}[X_{(n)}] - \theta)}_{\text{Bias}}^2$

$$= c^2\left[\frac{n}{(n+2)(n+1)^2}\theta^2\right] + \left[\frac{c\,n}{n+1}\theta - \theta\right]^2$$

then $\frac{d}{dc} MSE_\theta(T_c)$

$$= 2c\left[\frac{n}{(n+2)(n+1)^2}\theta^2\right] + 2\left[\frac{nc}{n+1}\theta - \theta\right]\frac{n}{n+1}\theta$$

$\frac{d^2}{dc^2} MSE_\theta(T_c) = \frac{2n}{(n+2)(n+1)^2}\theta^2 + 2\left[\frac{n}{n+1}\theta\right]^2 > 0$ always

$\Rightarrow \frac{d}{dc} MSE_\theta(T_c) = 0$ the c which does this, minimizes $MSE_\theta(T_c)$

so, $c\left[\frac{n}{(n+2)(n+1)^2}\theta^2\right] + 2\left[\frac{nc}{n+1}\theta - \theta(n+1)\right]\frac{n}{n+1}\theta = 0$

$$\frac{nc\theta^2}{n+2} + (nc\theta - n\theta - \theta)n\theta = 0$$

$$\frac{c}{n+2} + (nc - n - 1) = 0$$

$$\frac{c}{n+2} + nc = n+1$$

$$\frac{(n+1)^2 c}{n+2} = n+1$$

$$c = \frac{n+2}{n+1}$$

this minimises $MSE_\theta(T_c)$

it is biased as $\mathbb{E}[cX_{(n)}] = \frac{n+2}{n+1}\frac{n}{n+1}\theta \neq \theta$

29th March:

Theorm: (Rao-Blackwell) Let W be any unbiased estimator of θ and T be a sufficient statistic for θ

let

$$\phi(T) = E_\theta[W|T] \text{ is free of } \theta$$

and, $\phi(T)$ is another unbiased estimator of θ s.t

$$Var_\theta(\phi(T)) \leq Var_\theta(W) \quad \forall \theta \in \Theta$$

proof: $\forall \theta \in \Theta$ we have

$$E_\theta[\phi(T)] = E_\theta[E_\theta[W|T]]$$
$$= E_\theta[W] = \theta$$

$\Rightarrow \phi(T)$ is unbiased for θ, also $\forall \theta \in \Theta$

$$Var_\theta(W) = Var_\theta(E_\theta[W|T]) + E_\theta[\underbrace{Var(W|T)}_{\geq 0}]$$

$$\geq Var_\theta(E_\theta[W|T])$$
$$= Var_\theta(\phi(T))$$

Note: $\phi(T)$ is UMVUE when T is complete sufficient statistic

Theorm: (Lehmann-Scheffe) Let T be a complete sufficient statistic for θ and $\psi(T)$ be an unbiased estimator of θ based on T, then $\psi(T)$ is the unique UMVUE of θ

Corr: Let W be any unbiased estimator of θ and T be a complete sufficient statistic for θ, then

$$\phi(T) = E_\theta[W|T] \text{ is unique UMVUE of } \theta$$

Note: existance of unbiased estimator of θ and existance of complete sufficient statistic $\Rightarrow$ existance of UMVUE of θ

Theorm: (Lehmann-Scheffe) Let T be a complete sufficient statistic for θ and $\psi(T)$ be an unbiased estimator of θ based on T, then $\psi(T)$ is the unique UMVUE of θ

proof: for some unbiased estimator of θ we want to show that

$$Var_\theta(\psi(T)) \leq Var_\theta(W) \quad \forall \theta \in \Theta$$

as T is sufficient, $\phi(T) = E_\theta[W|T]$ is unbiased estimator of θ and

$$Var_\theta(\phi(T)) \leq Var_\theta(W) \quad \forall \theta \in \Theta \quad (\because \text{previous theorm})$$

moreover, $\phi(T)$ and $\psi(T)$ are both

$$\Rightarrow E_\theta[\phi(T) - \psi(T)] = 0 \quad \forall \theta \in \Theta$$

by completness of T

$$P_\theta[\psi(T) - \phi(T) = 0] = 1 \quad \forall \theta \in \Theta$$
$$\Rightarrow P_\theta[\psi(T) = \phi(T)] = 1 \quad \forall \theta \in \Theta$$

$$\Rightarrow Var_\theta(\psi(T)) = Var_\theta(\phi(T)) \leq Var_\theta(W) \quad \forall \theta \in \Theta$$

Eg: $X_1, X_2, \dots X_n \overset{iid}{\sim} Ber(p)$

$\sum_{i=1}^{n} x_i$ complete sufficient for p

also, $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ is unbiased for p

$\Rightarrow \bar{x}$ is unique UMVUE of p by Lehmann-Scheff

Eg: $X_1, X_2, \dots X_n \overset{iid}{\sim} Poi(\lambda)$

$\sum_{i=1}^{n} x_i$ is complete sufficient ($\because$ exponent family)

$\Rightarrow \bar{x}$ is unique UMVUE of λ

Eg: X_1 is unbiased estimator of λ and $S = \sum_{i=1}^{n} x_i$ is complete sufficient statistic for λ

$\Rightarrow \mathbb{E}_\lambda[X_1 \mid S] = \mathbb{E}_\lambda[X_1 \mid \Sigma X_i = S]$ ($\because$ corr of Lehmann-Scheffe)

$= \frac{s}{n} = \bar{x}$

Exe: 1) Let $X_1 \dots X_n \overset{iid}{\sim} Exp(\lambda)$ (λ = rate) then show that $\bar{x}$ is UMVUE of $\frac{1}{\lambda}$

2) Suppose $X_1, X_2, \dots X_n \overset{iid}{\sim} Poi(\lambda)$ then find UMVUE of $e^{-\lambda}$
($e^{-\lambda} = P[X_1 = 0]$, so $\mathbb{I}_{(X_1=0)}$ is unbiased estimator)

Ans: 1) $X_1, X_2, \dots X_n \overset{iid}{\sim} Exp(\lambda)$

then $f(x|\lambda) = \lambda e^{-\lambda x}$

$L(\underset{\sim}{x}|\lambda) = \lambda^n e^{-\lambda(\sum_{i=1}^{n} x_i)}$

so, $\sum_{i=1}^{n} x_i$ is complete sufficient for λ

now $\mathbb{E}[\sum_{i=1}^{n} x_i] = \frac{1}{\lambda} + \frac{1}{\lambda} \cdots + \frac{1}{\lambda} = \frac{n}{\lambda}$

$\Rightarrow \bar{x}$ is UMVUE for $\frac{1}{\lambda}$

2) $X_1, X_2, \dots X_n \overset{iid}{\sim} Poi(\lambda)$

now $f(x|\lambda) = \frac{\lambda^x e^{-\lambda}}{x!}$

$= e^{x \ln \lambda} e^{-\lambda} e^{-\ln(x!)}$

now, $L(\underset{\sim}{x}|\lambda) = e^{\Sigma x_i \ln\lambda - n\lambda - \underbrace{\Sigma \ln(x_i!)}_{h(\underset{\sim}{x})}}$

so, Σx_i is complete sufficient for $e^{-\lambda}$

also, $P[X_1 = 0] = e^{-\lambda}$

so, $\mathbb{I}_{X_1=0}$ is unbiased estimator of $e^{-\lambda}$

$$E_\lambda\left[I_{X_1=0} \,\middle|\, \sum_{i=1}^{n} X_i \right] = \text{UMVUE of } e^{-\lambda}$$

$$\text{now } P[X_1 \neq 0 \mid \Sigma X_i] \times 0 + P[X_1 = 0 \mid \sum_{i=1}^{n} X_i] \times 1 = \text{UMVUE of } e^{-\lambda}$$

$$\frac{P[X_1 = 0,\ \Sigma X_i = s]}{P[\Sigma X_i = s]} = e^{-\lambda} \frac{e^{-(n-1)\lambda}(n-1)^s (x)^s}{e^{-n\lambda}(n)^s (x)^s}$$

$$= \left(1 - \frac{1}{n}\right)^s$$

$$\text{so, } E_\lambda\left[I_{X_1=0} \,\middle|\, \sum_{i=1}^{n} X_i \right] = \left(1 - \frac{1}{n}\right)^{\sum_{i=1}^{n} X_i}$$

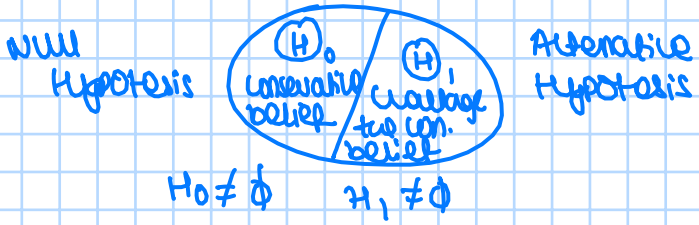
↓ UMVUE of $e^{-\lambda}$

Note: Suppose $X_1, X_2, \ldots X_n \overset{iid}{\sim} \text{Unif}(0,\theta)$, then it can be shown that $X_{(n)}$ is a complete statistic for θ, then from previous result

$$T_C = C X_{(n)}$$

unbiased for $C = \frac{n+1}{n}$ is the unique UMVUE of θ

Testing of Hypothesis:



testing for $H_0: \theta \in \theta_0$ v/s $H_1: \theta \in \theta_1$

Eg: θ be proportion of defective items in a manufacturing process, and we want to find out if process is running smoothly if $\theta < 0.001$

$H_0: \theta \leq 0.0001$ $\quad H_1: \theta > 0.0001$

↑ Null Hypothesis

Eg: Let θ = mean change in patient's blood pressure after a drug

H_0, H_1

$H_0: \theta = 0$ v/s $H_1 = \theta \neq 0$

generally we have

$$X_1, X_2 \ldots X_n \overset{iid}{\sim} f(x/\theta) \quad (\text{pmf/pdf})$$

$$\theta \in \Theta = \theta_0 \cup \theta_1$$

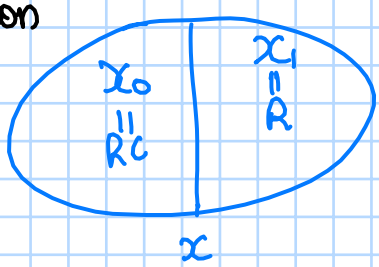
where $\theta_0 \cap \theta_1 = \phi$ and $\theta_0, \theta_1 \neq \phi$

in a hypothesis testing problem, after observing the sample, the experimenter must decide to either accept H_0 or reject it.

in other words, the testing procedure has to partition the sample space χ into

two non-trivial parts $\mathcal{X}_0$ and $\mathcal{X}_1$

Acceptance region



Rejection Region

s.t $\underset{\sim}{x} \in \mathcal{X}_0 \Rightarrow$ accept H_0

$\underset{\sim}{x} \in \mathcal{X}_1 \Rightarrow$ Reject H_0

typically, a hypothesis test (equivalently it's rejection region) is specified in terms of a test statistic

$X_1, X_2, \ldots X_n \overset{iid}{\sim} N(0,1)$

and we want to test for

$H_0 = \theta = 0$ V/S $H_1: \theta \neq 0$

most natural is $\bar{X}$, if $\bar{x}$ is large i.e

$$R = \mathcal{X}_1 = \{\underset{\sim}{x} \in \mathbb{R}^n ; \ \bar{x} > c\}$$

for some $c > 0$

Decision → / Truth ↓	Accept H_0 ($\underset{\sim}{x} \in \mathcal{X}_0 = R^C$)	Reject H_0 ($\underset{\sim}{x} \in \mathcal{X}_1 = R$)
$H_0: \theta \in \theta_0$	correct	type I error
$H_1: \theta \in \theta_1$	type II error	correct

ideally, we should try to reduce P[type I error] and P[type II error] as much as possible, However this involves a tradeoff b/w these two probabilities

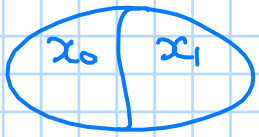
NOTE: Type I error is more severe

So, any good testing procedure will ensure that the probability of type I error $< \alpha$ for $\alpha = 0.05, 0.01$ and try to reduce P[type II error]

4th April:

Testing of Hypothesis:

we have a bias towards null hypothesis, we want to create a test



$\underset{\sim}{x} \in x_0 \rightarrow$ accept H_0
$\underset{\sim}{x} \in x_1 \rightarrow$ Reject H_0 (R^c is Rejection region)

we have seen, in case of $N(0,1)$, $\bar{x}$ is the obvious statistic

Type I error is most severe as our conservative belief true but we end up rejecting it

our goal is to ensure $P[\text{type I}] \leq \alpha$ and $P[\text{type II}]$ small

for any $\theta \in \theta_0$, $P_\theta[\text{type I error}] = P_\theta[\underset{\sim}{x} \in R]$

on other hand, for any $\theta \in \theta_1$, $P_\theta[\underset{\sim}{x} \in R] = 1 - P_\theta[\underset{\sim}{x} \in x_0]$ $\quad R = x_1, R^c = x_0$

$= 1 - P_\theta[\text{type II error}]$

Defn: The function $\beta: \theta \rightarrow [0,1]$ defined by

$\beta(\theta) = P_\theta[X \in R]$, $\theta \in \theta$ is called the power function of a hypothesis testing procedure with rejection region R

we have first showed that

$$\beta(\theta) = \begin{cases} P_\theta[\text{Type I error}] & ; \theta \in \theta_0 \\ 1 - P_\theta[\text{Type II error}] & ; \theta \in \theta_1 \end{cases}$$

fix $\alpha \in [0,1]$, typically $\alpha = 0.01$ or 0.05

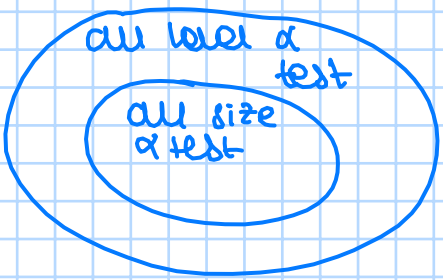
Defn: A test with power function $\beta(\theta)$ is a level α test if

$$\sup_{\theta \in \theta_0} \beta(\theta) = \sup_{\theta \in \theta_0} P_\theta[\underset{\sim}{x} \in R] \leq \alpha$$

Defn: A test with power function $\beta(\theta)$ is a size α test if

$$\sup_{\theta \in \theta_0} \beta(\theta) = \sup_{\theta \in \theta_0} P_\theta[\underset{\sim}{x} \in R] = \alpha$$

Note: Any size α test is also a level α test, therefore, the set of level α test contains the set of size α test



Note: It is possible, we would try to find the level α/ size α test for which the power function $\beta(\theta)$ is maximised for each $\theta \in \Theta_1$

Eg: $X_1, X_2, \dots X_n \overset{iid}{\sim} N(0,1)$, now we want to test for $H_0 : \theta = 0$ v/s $H_1 : \theta > 0$. The most natural test has rejection region of the form

$$R = \{ \underset{\sim}{x} \in \mathbb{R}^n \mid \bar{x} > c \} \text{ for some } c$$

we want to find c s.t the test has size α, $\Theta_0 = \{0\}$

$$\Rightarrow P_{\theta=0}[\bar{X} > c] = \alpha$$

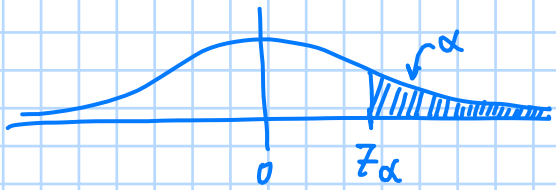
$$\Rightarrow P_0\left[\frac{\bar{X} - 0}{\frac{1}{\sqrt{n}}} > \frac{c - 0}{\frac{1}{\sqrt{n}}} \right] = \alpha$$

under H_0: $X_1, X_2, \dots X_n \overset{iid}{\sim} N(0,1)$

$$\Rightarrow \bar{X} \overset{iid}{\sim} N\left(0, \frac{1}{n}\right)$$

$$\Rightarrow \frac{\bar{X} - 0}{\frac{1}{\sqrt{n}}} \sim N(0,1) \quad \text{let } Z = \frac{\bar{X} - 0}{\frac{1}{\sqrt{n}}} \sim N(0,1)$$

$$\Rightarrow P_0[Z > \sqrt{n}c] = \alpha$$



$\therefore$, $c\sqrt{n} = z_\alpha$ where z_α is s.t $P[Z > z_\alpha] = \alpha = 1 - \Phi(z_\alpha)$ (Φ: $N(0,1)$)

$\Rightarrow c = \frac{z_\alpha}{\sqrt{n}}$, therefore the most natural size α test for $H_0 : \theta = 0$ v/s $H_1 : \theta > 0$ has rejection region $R = \{ \underset{\sim}{x} \in \mathbb{R}^n \mid \bar{x} > \frac{z_\alpha}{\sqrt{n}} \}$

now, if Θ_0 (respectively Θ_1) is a singleton set, then we say that the null (respectively alternative) hypothesis is a basic hypothesis

on other hand if Θ_1 (resp Θ_0) is not a singleton set, then we say that the alternative (resp null) hypothesis is a composite hypothesis

Note: The convention is to keep basic hypothesis (if any) as null hypothesis

If we are testing for one-dimensional parameter θ with parameter space $\Theta \subseteq \mathbb{R}$, then a test of the form

$$H_0 : \theta = \theta_0 \text{ v/s } H_1 : \theta \neq \theta_0$$

is called a two sided test and H_1 is called a two sided alternative

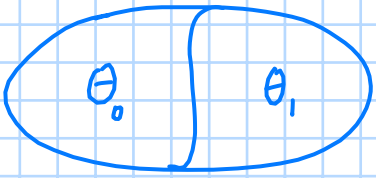
If H_1 is of the form $\theta > \theta_0$ or $\theta < \theta_0$ for some $\theta_0 \in \Theta$, then the test is called one-sided and H_1 is called a one sided alternative

Likelihood Ratio Test:

setup: $X_1, X_2, \dots X_n \overset{iid}{\sim} f(x|\theta)$ (pmf/pdf)

$\theta \in \Theta \subseteq \mathbb{R}$

↑ one-dimension



NOTE: we don't need to have $\Theta \subseteq \mathbb{R}$, but $\Theta \subseteq \mathbb{R}^n$ also fine

the likelihood fution $L(\underset{\sim}{x}|\theta) = \prod_{i=1}^{n} f(x_i|\theta)$

the likelihood ratio test (LRT) statistic for testing $H_0: \theta \in \Theta_0$ v/s $H_1: \theta \in \Theta_1$ is

$$\lambda(\underset{\sim}{x}) = \frac{\sup_{\theta \in \Theta_0} L(\underset{\sim}{x}|\theta)}{\sup_{\theta \in \Theta} L(\underset{\sim}{x}|\theta)} \leq 1$$

Defn: A likelihood ratio test (LRT) is any test that has a region of the form

$R = \{\underset{\sim}{x} : \lambda(\underset{\sim}{x}) < c\}$ where c is a number satisfying $0 < c < 1$

(c has to be determined from the size of the test)

The rationale behind LRTs is as follows, if the data is actually coming from the null distribution (i.e of form $f(x|\theta)$ with $\theta \in \Theta_0$) then the maximum unconstrained likelihood of the sample we may use to the maximum likelihood of the sample when θ is constrained to be inside Θ_0 and hence the observed value of the LRT statistic should be very close to 1. The smaller value of LRT statistic will make $\theta \in \Theta_1$ more likely and hence we should reject $\lambda(\underset{\sim}{x})$ when small

$$\lambda(\underset{\sim}{x}) < c$$

Suppose the (unconstrained) MLE $\hat{\theta} = \hat{\theta}(\underset{\sim}{x})$ exist, i.e

$$L(\underset{\sim}{x}|\hat{\theta}(\underset{\sim}{x})) = \sup_{\theta \in \Theta} L(\underset{\sim}{x}|\theta) \quad \forall\ \underset{\sim}{x} \in \mathcal{X}$$

let us also assume that the constrained MLE $\hat{\theta}_0 = \hat{\theta}_0(\underset{\sim}{x})$ needed in the numerator of the LRT also exist, i.e

$$L(\underset{\sim}{x}|\hat{\theta}_0(\underset{\sim}{x})) = \sup_{\theta \in \Theta_0} L(\underset{\sim}{x}|\theta) \quad \forall\ \underset{\sim}{x} \in \mathcal{X}$$

then the LRT statistic is of the form

$$\lambda(\underset{\sim}{x}) = \frac{\sup_{\theta \in \Theta_0} L(\underset{\sim}{x}|\theta)}{\sup_{\theta \in \Theta} L(\underset{\sim}{x}|\theta)} = \frac{L(\underset{\sim}{x}|\hat{\theta}_0)}{L(\underset{\sim}{x}|\hat{\theta})}$$

↑ constrained MLE

↑ unconstrained MLE

Eg: suppose $X_1, X_2, \dots X_n \overset{iid}{\sim} N(\theta, 1)$, consider testing for $H_0: \theta = \theta_0$ v/s $H_1: \theta \neq \theta_0$ for some pretixed $\theta_0 \in \mathbb{R}$

for this two sided test the likelihood LRT statistic is

$$\lambda(\underset{\sim}{x}) = \sup_{\theta \in \Theta_0} L(\underset{\sim}{x}|\theta) \Big/ \sup_{\theta \in \Theta} L(\underset{\sim}{x}|\theta)$$

$$= \frac{L(\underset{\sim}{x} \mid \theta_0)}{L(\underset{\sim}{x} \mid \hat{\theta})}$$

MLE: $\hat{\theta}$ = MLE of $\theta = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

$$\lambda(\underset{\sim}{x}) = \frac{L(\underset{\sim}{x} \mid \theta_0)}{L(\underset{\sim}{x} \mid \bar{x})} = \frac{(2\pi)^{-n/2} e^{-\sum_{i=1}^{n} \frac{(x_i - \theta_0)^2}{2}}}{(2\pi)^{-n/2} e^{-\sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{2}}}$$

$$= \exp\left\{ -\frac{1}{2}\sum_{i=1}^{n}(x_i - \theta_0)^2 + \frac{1}{2}\sum_{i=1}^{n}(x_i - \bar{x})^2 \right\}$$

$$= \exp\left\{ -\frac{1}{2}\sum_{i=1}^{n}(x_i - \bar{x})^2 - \frac{n}{2}(\bar{x} - \theta_0)^2 + \frac{1}{2}\sum_{i=1}^{n}(x_i - \bar{x})^2 \right\}$$

$$= \exp\left\{ -\frac{n}{2}(\bar{x} - \theta_0)^2 \right\}$$

we know that LRT rejects when $\lambda(\underset{\sim}{x})$ is small
$\Rightarrow (\bar{x} - \theta_0)^2$ large
$\Rightarrow |\bar{x} - \theta_0|$ large

so, rejection region of any LRT is of form

$$R = \{ \underset{\sim}{x} \in \mathbb{R}^n : |\bar{x} - \theta_0| \geq c' \} \text{ for some } c' > 0$$

we want to find size α test LRT,

$$P_{\theta_0}[\underset{\sim}{X} \in R] = \alpha$$

$$P_{\theta_0}[|\bar{X} - \theta_0| \geq c'] = \alpha$$

under H_0 : $X_1, X_2, \ldots X_n \overset{iid}{\sim} N(\theta_0, 1)$

$$\Rightarrow \bar{X} \sim N(\theta_0, \tfrac{1}{n})$$

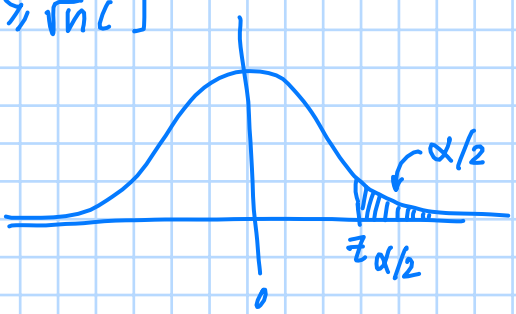
$$\Rightarrow \frac{\bar{X} - \theta_0}{\frac{1}{\sqrt{n}}} \sim N(0,1) \qquad Z = \frac{\bar{X} - \theta_0}{\frac{1}{\sqrt{n}}}$$

$$\Rightarrow P_{\theta_0}[|\bar{X} - \theta_0| \geq c'] = P_{\theta_0}[|Z| \geq \sqrt{n}c']$$

$$\Rightarrow 2P_{\theta_0}[Z \geq \sqrt{n}c'] = \alpha$$

$$\Rightarrow P_{\theta_0}[Z \geq \sqrt{n}c'] = \alpha/2$$

$$\Rightarrow \sqrt{n}c' = z_{\alpha/2}$$



where $z_{\alpha/2}$ is s.t $1 - \Phi(z_{\alpha/2}) = \alpha/2$

$\Rightarrow c' = \frac{1}{\sqrt{n}} z_{\alpha/2}$, so size α LRT rejects when $|\bar{X} - \theta_0| \geq \frac{1}{\sqrt{n}} z_{\alpha/2}$

7th April:

Uniformly most powerful (UMP) test:

Defn: Let C be a class of test for testing $H_0: \theta \in \Theta_0$ v/s $H_1: \theta \in \Theta_1$, a test in class C with power function $\beta(\theta)$ is a uniformly most powerful (UMP) class C test if $\beta(\theta) \geq \beta'(\theta)$ for every $\theta \in \Theta_1$ and every $\beta'(\theta)$ that is the power function of a test in class C

Note: Above may not exist

Eg: If C = the set of all level (size, resp) α test for testing $H_0: \theta \in \Theta_0$ $H_1: \theta \in \Theta_1$, then a UMP (class C-test (if exist) is called a UMP level (size, resp) α test

Note: If Θ_1 is a singleton set (i.e alternative is basic) then we drop the word uniformly, we use most powerful (MP)

now, suppose we have $X_1, X_2, \ldots X_n \overset{iid}{\sim} f(x|\theta)$ (pdf/pmf) and then

$\theta \in \Theta = \{\theta_0, \theta_1\}$

parameter space has 2 distinct points

we consider testing for $H_0: \theta = \theta_0$ v/s $H_1: \theta = \theta_1$

$\Theta_0 = \{\theta_0\}$ and $\Theta_1 = \{\theta_1\}$

we want to find rejection region R s.t $P_{\theta_0}[\underset{\sim}{X} \in R] \leq \alpha$

if discrete, then $\sum_{\underset{\sim}{x} \in R} L(\underset{\sim}{x}|\theta_0) \leq \alpha$ ——①

and $P_{\theta_1}[\underset{\sim}{X} \in R] = \sum_{\underset{\sim}{x} \in R} L(\underset{\sim}{x}|\theta_1)$ should be the largest subject to ①

Theorem: (Neyman-Pearson) $X_1, X_2, \ldots X_n \overset{iid}{\sim} f(x|\theta)$ (pmf/pdf) $\Theta = \{\theta_0, \theta_1\}$ $(\theta_0 \neq \theta_1)$ consider testing for $H_0: \theta = \theta_0$ v/s $H_1: \theta = \theta_1$ using a test with rejection region R that satisfies

(i) $\begin{cases} \underset{\sim}{x} \in R \text{ if } L(\underset{\sim}{x}|\theta_1) > K L(\underset{\sim}{x}|\theta_0) \\ \text{and } \underset{\sim}{x} \in R^c \text{ if } L(\underset{\sim}{x}|\theta_1) < K L(\underset{\sim}{x}|\theta_0) \end{cases}$

for some $K \geq 0$

(ii) $\alpha = P_{\theta_0}[\underset{\sim}{X} \in R]$

then, (a) sufficiency, any test that satisfies (i), (ii) with some $K \geq 0$, is a MP level α test

(b) Necessity, if $\exists$ a test that satisfies (i), (ii) with some $K > 0$, then every MP level α test has size α (i.e (ii)) and every MP level α test satisfies (i) except perhaps on a set A for which $P_{\theta_0}[\underset{\sim}{X} \in A] = P_{\theta_1}[\underset{\sim}{X} \in A] = 0$

Eg: MP test for binomial dist^n, $X \sim Bin(2, \theta)$ sample size = 1, we want to test for $H_0: \theta = \frac{1}{2}$ v/s $H_1: \theta = \frac{3}{4}$

$f(x|\theta) = \binom{2}{x} \theta^x (1-\theta)^{2-x}$ $x \in \{0, 1, 2\}$

$$\frac{f(0|3/4)}{f(0|1/2)} = \frac{\binom{2}{0}\left(\frac{3}{4}\right)^0\left(\frac{1}{4}\right)^2}{\binom{2}{0}\left(\frac{1}{2}\right)^2} = \frac{1}{4}$$

$$\frac{f(1|3/4)}{f(1|1/2)} = \frac{\binom{2}{0}\left(\frac{3}{4}\right)\left(\frac{1}{4}\right)}{\binom{2}{0}\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)} = \frac{3}{4}$$

$$\frac{f(2|3/4)}{f(2|1/2)} = \frac{\binom{2}{0}\left(\frac{3}{4}\right)^2\left(\frac{3}{4}\right)^0}{\binom{2}{0}\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)} = \frac{9}{4}$$

then for $\frac{3}{4} < K < \frac{9}{4}$, then NP lemma says that the test that rejects H_0 if $X=2$ is MP level $\alpha = P_{\frac{1}{2}}[X=2] = \frac{1}{4}$ test

if we choose $\frac{1}{4} < K < \frac{3}{4}$ then NP lemma says test that rejects H_0 if $X \in \{1,2\}$ is MP level $\alpha = P_{\frac{1}{2}}[X \in \{1,2\}] = \frac{3}{4}$ test

$K < \frac{1}{4}$ yield MP test $\alpha = 1$ $K > \frac{9}{4}$ MP test for $\alpha = 0$

If $K = \frac{3}{4}$ then (i) says we must reject $X=2$ and accept H_0 $X=0$ leaving our action for $X=1$ in our hand

if we accept for $X=1$

then MP level $\alpha = P_{\frac{1}{2}}[X=2] = \frac{1}{4}$ test

if we reject for $X=1$

then MP level $\alpha = P_{\frac{1}{2}}[X \in \{1,2\}] = \frac{3}{4}$ test

Note: Not all levels can be attained in the discrete case, but for continuous case, all levels can be attained

Cor: Let $\alpha \in [0,1]$, $x_1 \dots x_n \overset{iid}{\sim} f(x|\theta)$

$\Theta = \{\theta_0, \theta_1\}$ $(\theta_0 \neq \theta_1)$, consider testing for $H_0: \theta = \theta_0$ and $H_1: \theta = \theta_1$, suppose $T = T(\underset{\sim}{X})$ is a sufficient statistic for θ and $g(t|\theta)$ is the pmf/pdf of T, then any test based on T with rejection region $S \subseteq \text{Range}(T)$ is a MP level α test if

(i) $\begin{cases} t \in S & \text{if } g(t|\theta_1) > c\, g(t|\theta_0) \\ t \in S^c & \text{if } g(t|\theta_1) < c\, g(t|\theta_0) \end{cases}$ for some $c \geq 0$

and (ii) $\alpha = P_{\theta_0}[T \in S]$

Eg: $X_1, X_2, \dots X_n \overset{iid}{\sim} N(\theta, \sigma^2)$ with σ^2 known, take $\theta_0, \theta_1 \in \mathbb{R}$ with $\theta_0 > \theta_1$, consider testing $H_0: \theta = \theta_0$ v/s $H_1: \theta = \theta_1$

$T = \bar{X}$ sufficient for θ and $T \sim N(\theta, \frac{\sigma^2}{n})$

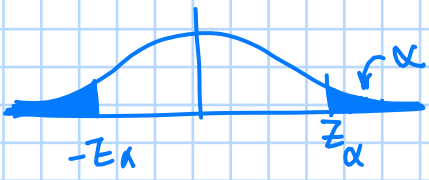
$$\Rightarrow g(t|\theta) = \frac{1}{\sqrt{2\pi}\frac{\sigma}{\sqrt{n}}} \exp\left\{-\frac{(t-\theta)^2}{2\frac{\sigma^2}{n}}\right\} \quad t \in \mathbb{R}$$

$$\Rightarrow g(t|\theta) = \frac{\sqrt{n}}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{n(t-\theta)^2}{2\sigma^2}\right\} \quad t \in \mathbb{R}$$

From below exercise, the MP level α test ($\theta_0 > \theta_1$) has rejection region of form $\bar{X} < c'$

where $\alpha = P_{\theta_0}[\bar{X} < c']$ where under H_0: $\bar{X} \sim N(\theta_0, \frac{\sigma^2}{n})$

$$\Rightarrow \frac{\bar{X} - \theta_0}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1) \quad \text{i.e} \quad P_{\theta_0}\left[\frac{\bar{X} - \theta_0}{\sigma/\sqrt{n}} < \frac{c' - \theta_0}{\sigma/\sqrt{n}}\right] = \alpha$$



$$\frac{c' - \theta_0}{\sigma/\sqrt{n}} = -z_\alpha$$

$$c' = \theta_0 - \frac{\sigma}{\sqrt{n}} z_\alpha$$

So, MP level α test ($\theta_1 < \theta_0$) rejects if

$$\bar{X} < \theta_0 - \frac{z_\alpha}{\sqrt{n}}\sigma$$

Note: The rejection region does not depend on the value of θ_1 as long as $\theta_1 < \theta_0$, this test is MP level α test for testing $H_0: \theta = \theta_0$ v/s $H_1: \theta = \theta_1$

So, by def this test with rejection region $\bar{X} < \theta_0 - \frac{z_\alpha}{\sqrt{n}}\sigma$ is actually UMP level α test for testing $H_0: \theta = \theta_0$ v/s $H_1: \theta < \theta_0$

Exe: Show that the inequality $g(t|\theta_1) > c\, g(t|\theta_0)$ is equivalent to $t < c'$ → done down

Exe: $X_1, X_2, \dots X_n \overset{iid}{\sim} N(\theta, \sigma^2)$ with σ^2 known, fix $\theta_0 \in \mathbb{R}$, find the UMP level α test for testing $H_0: \theta = \theta_0$ v/s $H_1: \theta_1 > \theta_0$ → done down

Eg: $X_1, X_2, \dots X_n \overset{iid}{\sim} N(\theta, \sigma^2)$, σ^2 known, fix $\theta_0 \in \mathbb{R}$, consider testing for $H_0: \theta = \theta_0$ v/s $H_1: \theta \neq \theta_0$, then we will show this does not have UMP level α test (it does not exist)

Consider an alternative parameter point $\theta_1 < \theta_0$, we have already shown that the test that rejects H_0 if $\bar{X} < \theta_0 - z_\alpha \frac{\sigma}{\sqrt{n}}$ has highest possible power at θ_1 among all the level α test (test 1)

Exe: Show that the inequality $g(t|\theta_1) > c\, g(t|\theta_0)$ is equivalent to $t < c'$ $\quad \theta_0 > \theta_1$

Ans:

$$g(t|\theta) = \frac{\sqrt{n}}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{n(t-\theta)^2}{2\sigma^2}\right\}$$

$$g(t|\theta_1) > c\, g(t|\theta_0)$$

$$\Leftrightarrow \frac{\sqrt{n}}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{n(t-\theta_1)^2}{2\sigma^2}\right\} > \frac{\sqrt{n}}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{n(t-\theta_0)^2}{2\sigma^2}\right\}$$

$$\Leftrightarrow \exp\left\{\frac{n}{2\sigma^2}\left((t-\theta_0)^2 - (t-\theta_1)^2\right)\right\} > c$$

$$\Leftrightarrow \frac{n}{2\sigma^2}\left[(t-\theta_0)^2 - (t-\theta_1)^2\right] > \log c$$

$$\Leftrightarrow (\cancel{t}-\theta_0-\cancel{t}+\theta_1)(t-\theta_0+t-\theta_1) > 2\sigma^2 \log c/n$$

$$\Leftrightarrow (2t-(\theta_0+\theta_1)) < \frac{2\sigma^2 \log c/n}{(\theta_0-\theta_1)}$$

$$\Leftrightarrow t < c'$$

Exe: $X_1, X_2, \ldots X_n \overset{iid}{\sim} N(\theta, \sigma^2)$ with σ^2 known, fix $\theta_0 \in \mathbb{R}$, find the UMP level α test for testing $H_0: \theta = \theta_0$ V/S $H_1: \theta_1 > \theta_0$

Ans:

$X_1, X_2, \ldots X_n \overset{iid}{\sim} N(\theta, \sigma^2)$ then

$\bar{x}$ is sufficient statistic for θ

$\bar{x} \sim N(\theta, \frac{\sigma^2}{n})$

now, $g(t|\theta) = \frac{\sqrt{n}}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{n(t-\theta)^2}{2\sigma^2}\right\}$

as $\theta_1 > \theta_0$ $\quad g(t|\theta_1) > c\, g(t|\theta_0) \Leftrightarrow t > c'$

then $P_{\theta_0}[t > c'] = P_{\theta_0}\left[\underbrace{\frac{\bar{x}-\theta_0}{\frac{\sigma}{\sqrt{n}}}}_{N(0,1)} > \frac{c'-\theta_0}{\frac{\sigma}{\sqrt{n}}}\right] = \alpha$

$$z_\alpha = \frac{c'-\theta_0}{\frac{\sigma}{\sqrt{n}}}$$

$$\Rightarrow c' = \frac{\sigma}{\sqrt{n}} z_\alpha + \theta_0$$

$$\Rightarrow \bar{x} > \theta_0 + \frac{\sigma}{\sqrt{n}} z_\alpha \text{ then reject}$$

10th April:

Suppose that the UMP test for $H_0: \theta = \theta_0$ v/s $H_1: \theta \neq \theta_0$ exist, then the UMP test must be most powerful for testing $H_0: \theta = \theta_0$ v/s $H_1: \theta = \theta_1$ (UMP level α test)

therefore, by (ii) Necessity of NP lemma, it follows that this UMP test must have rejection region same as test 1 except for a set of measure 0

a set A satisfying $P_{\theta_0}[\underset{\sim}{x} \in A] = P_{\theta_1}[\underset{\sim}{x} \in A]$

now, consider test 2 that rejects if $\bar{x} > \theta_0 + z_\alpha \frac{\sigma}{\sqrt{n}}$, this is also a size α test (and so level α test), let

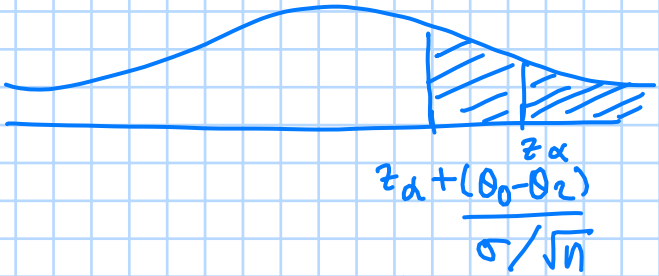
$\beta_1(\theta)$ = power function of test 1

$\beta_2(\theta)$ = power function of test 2

take any $\theta_2 > \theta_0$, observe that $\beta_2(\theta_2) = P_{\theta_2}\left[\bar{x} > \theta_0 + z_\alpha \frac{\sigma}{\sqrt{n}}\right]$

$$= P_{\theta_2}\left[\underbrace{\frac{\bar{x} - \theta_2}{\frac{\sigma}{\sqrt{n}}}}_{N(0,1)} > z_\alpha + \frac{\theta_0 - \theta_2}{\frac{\sigma}{\sqrt{n}}}\right]$$

$\theta_0 - \theta_2 < 0$



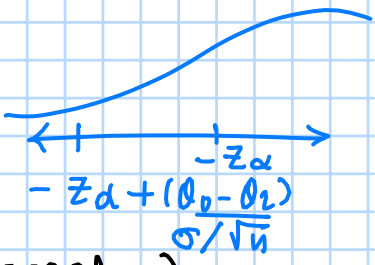
$$> P_{\theta_2}\left[\underbrace{\frac{\bar{x} - \theta_2}{\frac{\sigma}{\sqrt{n}}}}_{N(0,1)} > z_\alpha\right] = \alpha$$

so, $\beta_2(\theta_2) > \alpha$ —— (1)

now, $\beta_1(\theta_2) = P_{\theta_2}\left[\bar{x} < \theta_0 - z_\alpha \frac{\sigma}{\sqrt{n}}\right]$

$$= P_{\theta_2}\left[\frac{\bar{x} - \theta_2}{\frac{\sigma}{\sqrt{n}}} < -z_\alpha + \frac{(\theta_0 - \theta_2)}{\frac{\sigma}{\sqrt{n}}}\right]$$

$$< P_{\theta_2}\left[\frac{\bar{x} - \theta_2}{\frac{\sigma}{\sqrt{n}}} < -z_\alpha\right] = \alpha \quad (\because \text{symmetry})$$



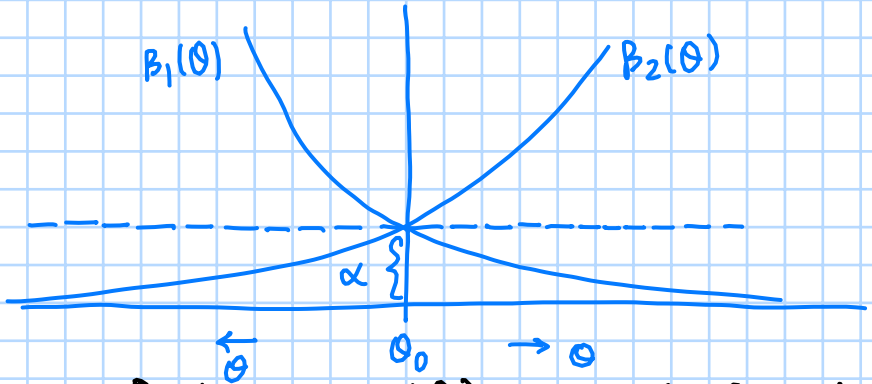
so, $\beta_2(\theta_1) < \alpha$ —— (2)

so, $\beta_1(\theta_2) < \alpha < \beta_2(\theta_2)$ is a contradiction

as test 1 is a.s UMP but $\exists\, \beta_2$ s.t $\beta_1(\theta_2) < \beta_2(\theta_2)$

but $\beta_1(\theta) \geq \beta(\theta)\ \forall \beta$, but this is not true

now, $\beta_2(\theta)$ for test 2 is



the problem is that $\beta_1(\theta) < \alpha$ if $\theta > \theta_0$ for test 1

$\beta_2(\theta) < \alpha$ if $\theta < \theta_0$ for test 2

also, one more problem is that test 1, test 2 is not unbiased

Defn: let $\alpha \in (0,1)$, A test for $H_0: \theta \in \Theta_0$ v/s $H_1: \theta \in \Theta_1$ is called an unbiased level α test if it's power function $\beta(\theta)$ satisfies

$$\beta(\theta) \leq \alpha \quad \forall \theta \in \Theta_0$$

$$\beta(\theta) \geq \alpha \quad \forall \theta \in \Theta_1$$

now, $X_1, X_2 \ldots X_n \overset{iid}{\sim} N(\theta, \sigma^2)$ with σ^2 known, fix $\theta_0 \in \mathbb{R}$

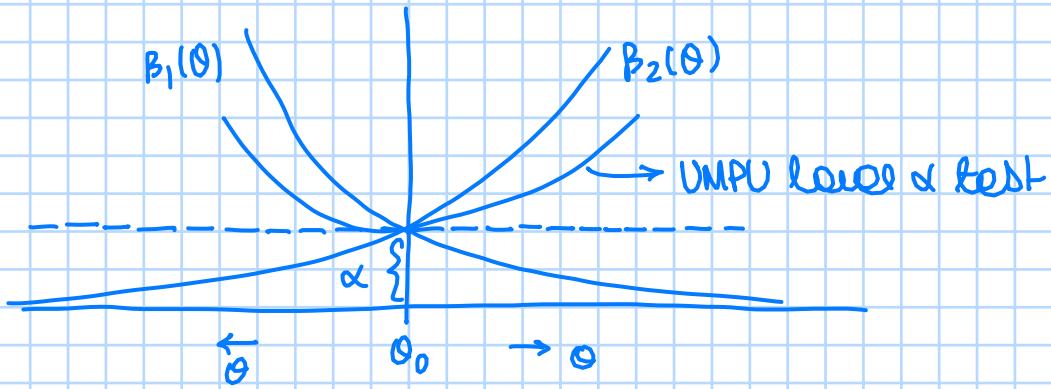
consider testing $H_0: \theta = \theta_0$ v/s $H_1: \theta \neq \theta_0$

the most natural test rejects H_0 when $|\bar{X} - \theta_0|$ is large

Exe: show that the level α test of the above form rejects when $|\bar{X} - \theta_0| > z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$, where $z_{\alpha/2}$ is s.t $\phi(z_{\alpha/2}) = 1 - \alpha/2$ → done down

Note: above test is UMP in the class $\mathcal{C}_\alpha$ of all level α unbiased test for $H_0: \theta = \theta_0$ v/s $H_1: \theta \neq \theta_0$ in a normal population

such a test is called UMPU or UMP-Unbiased level α test



Eg: suppose $X_1, X_2, \ldots X_n \overset{iid}{\sim} N(\theta, \sigma^2)$ with both unknown and fix $\theta_0 \in \mathbb{R}$, suppose we want to test for

$$H_0: \theta = \theta_0 \quad \text{v/s} \quad H_1: \theta \neq \theta_0$$

earlier σ was known, so we were using the fact that under H_0, $\frac{\bar{X} - \theta_0}{\sigma/\sqrt{n}} \sim N(0,1)$

now σ is unknown, and so, we shall estimate it by

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

under H_0, $\frac{\bar{x} - \theta_0}{s/\sqrt{n}} \sim t_{n-1}$

so, the most natural test for testing $H_0: \theta = \theta_0$ v/s $H_1: \theta \neq \theta_0$ when σ is unknown rejects when

$$\frac{|\bar{x} - \theta_0|}{s\sqrt{n}} > t_{n-1, \frac{\alpha}{2}}$$ where $t_{n-1,\alpha/2}$ is a number s.t



$P[T > t_{n-1,\alpha/2}] = \alpha/2$ $\quad t_{n-1,\alpha/2}$ $\quad T \sim t_{n-1}$

Exe: Let $X \sim Exp(\theta)$ with pdf

$$f(x|\theta) = \theta e^{-\theta x} \quad x > 0$$

$\theta > 0$ unknown parameter

(a) Using NP Lemma, find MP test size α for

$H_0: \theta = \theta_0$ v/s $H_1: \theta = \theta_1$

$0 < \theta_0 < \theta_1$

→ done down

(b) Consider the problem of testing for the same null hypothesis as above against the following one-sided composite alternative hypothesis

$H_0: \theta = \theta_0$ v/s $H_1: \theta > \theta_0$

Show that test obtained in (a) is uniformly most powerful for this problem, compute the power function of this test

(c) $H_0: \theta = \theta_0$ v/s $H_1: \theta \neq \theta_0$, investigate if UMP test exist

Exe: $U_1, U_2 \overset{iid}{\sim} Unif(\theta, \theta+1)$ where $\theta \in \mathbb{R}$ unknown, for testing $H_0: \theta = 0$ v/s $H_1: \theta = 0.8$

Test 1: H_0 Reject if $U_1 > 0.81$ → done down

Test 2: H_0 reject when $\max(U_1, U_2) > c$ for some c

both test have same size

(a) find c
(b) find power of both the test

Interval estimation:

$X_1, X_2, \dots X_n \overset{iid}{\sim} f(x|\theta)$ pmf/pdf where θ is scalar

An interval estimator of θ is a pair of fuctions $L(\underset{\sim}{x})$ and $U(\underset{\sim}{x})$ of a sample satisfying $L(\underset{\sim}{x}) \leq U(\underset{\sim}{x}) \ \forall \ \underset{\sim}{x} \in \mathcal{X}$

If $\underset{\sim}{X} = \underset{\sim}{x}$ is observed, then the interval $L(\underset{\sim}{x}) \leq \theta \leq U(\underset{\sim}{x})$ is made

the random interval $[L(\underset{\sim}{X}), U(\underset{\sim}{X})]$ is called an interval estimator of θ ← same as confidence interval

for an interval estimator $[L(\underset{\sim}{X}), U(\underset{\sim}{X})]$ of a parameter θ, the coverage probability is the probability that the random interval covers the parameter value. In other words

$$\text{coverage prob} = P_\theta[\theta \in [L(\underset{\sim}{X}), U(\underset{\sim}{X})]]$$
$$= P_\theta[L(\underset{\sim}{X}) \leq \theta \leq U(\underset{\sim}{X})]$$

Defn: confidence coefficient of the interval estimator $[L(\underset{\sim}{X}), U(\underset{\sim}{X})]$ of θ is $\inf_{\theta \in \Theta} P_\theta[L(\underset{\sim}{X}) \leq \theta \leq U(\underset{\sim}{X})]$

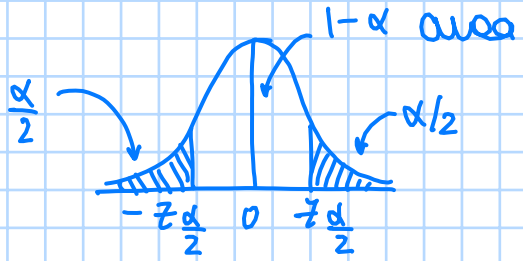
Eg: $X_1, X_2, \dots X_n \overset{iid}{\sim} N(\theta, \sigma^2)$, the most natural confidence interval

we know $\bar{X} \sim N(\theta, \frac{\sigma^2}{n})$

$$\Rightarrow \frac{\bar{X} - \theta}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$

↖ Pivotal quality (distribution is known)

we know that for any $N(0,1)$ r.v



$$P_\theta\left[-z_{\frac{\alpha}{2}} \leq \frac{\bar{X} - \theta}{\frac{\sigma}{\sqrt{n}}} \leq z_{\frac{\alpha}{2}}\right] = 1 - \alpha$$

$$\text{so, } -z_{\alpha/2} \leq \frac{\bar{X} - \theta}{\sigma/\sqrt{n}} \leq z_{\alpha/2} \Rightarrow -\frac{z_{\alpha/2}\sigma}{\sqrt{n}} \leq \bar{X} - \theta \leq z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

$$\Rightarrow -\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq -\theta \leq -\bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

$$\Rightarrow \bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \theta \leq \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

$$\Rightarrow P_\theta\left[\underbrace{\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}}_{L(\underset{\sim}{X})} \leq \theta \leq \underbrace{\bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}}_{U(\underset{\sim}{X})}\right] = 1 - \alpha$$

$$\text{so, } P_\theta[L(\underset{\sim}{X}) \leq \theta \leq U(\underset{\sim}{X})] = 1 - \alpha \ \forall \theta \text{, so } \sim \text{ confidence coeff is } 1 - \alpha$$

or $100(1-\alpha)\%$.

Exe: Show that the level α test of the above form rejects when $|\bar{x}-\theta_0| > z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$, where $z_{\alpha/2}$ is s.t $\phi(z_{\alpha/2}) = 1-\alpha/2$

Ans: Now $R = \{\underset{\sim}{x} \in \mathbb{R}^n \mid |\bar{x}-\theta_0| > c\}$

↑ for some c

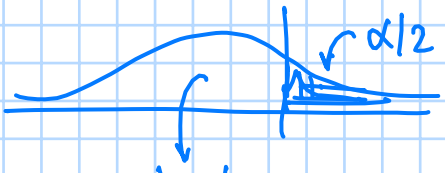
$X_1, X_2, \ldots X_n \overset{iid}{\sim} N(\theta, \sigma^2)$ $\quad \Theta_0 = \{\theta_0\}$

Now, $\sup_{\theta \in \Theta_0} \beta(\theta) = \beta(\theta_0) \leq \alpha$ $\quad \bar{x} \sim N(\theta_0, \frac{\sigma^2}{n})$

$\Rightarrow P_{\theta_0}[|\bar{x}-\theta_0| > c] \leq \alpha$

$\Rightarrow P_{\theta_0}\left[\frac{|\bar{x}-\theta_0|}{\frac{\sigma}{\sqrt{n}}} > \frac{c}{\frac{\sigma}{\sqrt{n}}}\right] \leq \alpha$

$\Rightarrow P_{\theta_0}\left[\frac{\bar{x}-\theta_0}{\frac{\sigma}{\sqrt{n}}} > \frac{c}{\frac{\sigma}{\sqrt{n}}}\right] \leq \frac{\alpha}{2}$ ($\because$ symmetry)



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

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

where $z_{\alpha/2}$ is s.t $\phi(z_{\alpha/2}) = 1-\alpha/2$

Exe: Let $X \sim Exp(\theta)$ with pdf

$$f(x|\theta) = \theta e^{-\theta x} \quad x > 0$$

$\theta > 0$ unknown parameter

(a) Using NP Lemma, find MP test size α for

$H_0: \theta = \theta_0$ v/s $H_1: \theta = \theta_1$

$0 < \theta_0 < \theta_1$

(b) Consider the problem of testing for the same null hypothesis as above against the following one-sided composite alternative hypothesis

$H_0: \theta = \theta_0$ v/s $H_1: \theta > \theta_0$

Show that test obtained in (a) is uniformly most powerful for this problem, compute the power fution of this test

(c) $H_0: \theta = \theta_0$ v/s $H_1: \theta \neq \theta_0$, investigate if UMP test exist

Ans: $X \sim Exp(\theta)$ $f(x|\theta) = \theta e^{-\theta x}$; $x > 0$ $\quad \theta > 0$ unknown

(a) $H_0: \theta = \theta_0$ $\quad H_1: \theta = \theta_1$ $\quad \theta_0 < \theta_1$

$\alpha = P_{\theta_0}[X \in R]$ $\quad \mathcal{X} = (0, \infty)$

$f(x|\theta_0) = \theta_0 e^{-\theta_0 x}$

$f(x|\theta_1) = \theta_1 e^{-\theta_1 x}$

$f(x|\theta_1) > K f(x|\theta_0) \quad x \in R$

$f(x|\theta_1) < K f(x|\theta_0) \quad x \in R^c$

$\theta_1 e^{-\theta_1 x} > K \theta_0 e^{-\theta_0 x}$

$\frac{\theta_1}{\theta_0} > K e^{(\underbrace{\theta_1 - \theta_0}_{>0})x}$ some $K > 0$

$\ln\left(\frac{\theta_1}{\theta_0}\right) > \ln K + (\theta_1 - \theta_0)x$

$\frac{\ln\left(\frac{\theta_1}{\theta_0}\right) - \ln K}{(\theta_1 - \theta_0)} > x$

so $\exists c'$ s.t. $x < c' \Leftrightarrow f(x|\theta_1) > K f(x|\theta_0)$

$x > c' \Leftrightarrow f(x|\theta_1) < K f(x|\theta_0)$

let $R = (0, c')$

$R' = [c', \infty)$

then $\alpha = \mathbb{P}_{\theta_0}[x \in R]$

$= \int_0^{c'} \theta_0 e^{-\theta_0 x} dx$

$= \theta_0 \left| \frac{e^{-\theta_0 x}}{-\theta_0} \right|_0^{c'}$

$= \frac{\theta_0}{-\theta_0}\left[e^{-\theta_0 c'} - e^0 \right]$

$\alpha = 1 - e^{-\theta_0 c'}$

$e^{-\theta_0 c'} = 1 - \alpha$

$-\theta_0 c' = \ln(1-\alpha)$

$c' = -\frac{1}{\theta_0} \ln(1-\alpha)$

$c' = \frac{1}{\theta_0} \ln\left(\frac{1}{1-\alpha}\right)$

so, $(0, \frac{1}{\theta_0} \ln\left(\frac{1}{1-\alpha}\right)) = R$

(b) as R does not depend on θ_1

$H_0: \theta = \theta_0 \quad H_1: \theta = \theta_1$

$R = \left(0, \frac{1}{\theta_0} \ln\left(\frac{1}{1-\alpha}\right)\right)$ is UMP test

$$\beta(\theta) = \begin{cases} P_{\theta_0}[X \in R] & ; \theta = \theta_0 \\ P_\theta[X \in R] & ; \theta > \theta_0 \end{cases}$$

$$= \begin{cases} \alpha & ; \theta = \theta_0 \\ \int_0^{\frac{1}{\theta_0}\ln\left(\frac{1}{1-\alpha}\right)} \theta e^{-\theta x}\, dx & ; \theta > \theta_0 \end{cases}$$

$$= 1 - e^{-\frac{\theta}{\theta_0}\ln\left(\frac{1}{1-\alpha}\right)}$$

$$= 1 - e^{\frac{\theta}{\theta_0}\ln(1-\alpha)}$$

$$= 1 - (1-\alpha)^{\frac{\theta}{\theta_0}}$$

$$\beta(\theta) = \begin{cases} \alpha & ; \theta = \theta_0 \\ 1-(1-\alpha)^{\theta/\theta_0} & ; \theta > \theta_0 \end{cases}$$

(c) if it exist, then it will be MP for

$H_0: \theta = \theta_0 \quad H_1: \theta = \theta_1 \quad \theta_1 > \theta_0$

then by NP lemma

Rejection ratio of $H_0: \theta = \theta_0$, $H_1: \theta \neq \theta_0$ is same as that of (b) (a.e)

let $\beta'(\theta)$ be power function of this test and β as in (b)

then for $\theta_1 < \theta_0$ $\beta'(\theta_1) = 1-(1-\alpha)^{\frac{\theta_1}{\theta_0}}$

$\alpha \in (0,1) \Rightarrow 1-\alpha \in (0,1)$ also $\frac{\theta_1}{\theta_0} < 1$

$\Rightarrow (1-\alpha)^{\theta_1/\theta_0} > 1$

$\Rightarrow \beta'(\theta_1) < 0$ this is a contradiction

So, no such UMP test exist

Exa: $U_1, U_2 \overset{iid}{\sim} \text{Unif}(\theta, \theta+1)$ where $\theta \in \mathbb{R}$ unknown, for testing $H_0: \theta = 0$ vs $H_1: \theta = 0.8$

Test 1: H_0 Reject if $U_1 > 0.81$

Test 2: H_0 reject when $\max(U_1, U_2) > c$ for some c

both test have same size

(a) find c
(b) find power of both the test

Ans: $U_1, U_2 \overset{iid}{\sim} \text{Unif}(\theta, \theta+1)$ $\theta \in \mathbb{R}$ unknown for testing

$$H_0: \theta = 0 \quad H_1: \theta = 0.8$$

Test 1: $R = \{(x_1, x_2) \mid x_1 > 0.81\}$

Test 2: $R = \{(x_1, x_2) \mid \max(x_1, x_2) > c\}$

Case as both have same size

$$\beta_1(\theta_0) = P_{\theta_0}[(U_1, U_2) \in R]$$
$$= P[U_2 > 0.81] \quad U_1, U_2 \sim \text{Unif}(0,1)$$
$$= \frac{1 - 0.81}{1}$$
$$= 0.19$$

$$\Rightarrow \beta_2(\theta_0) = 0.19 = P[\max(U_1, U_2) > c] \quad U_1, U_2 \sim \text{Unif}(0,1)$$
$$= 1 - P[\max(U_1, U_2) < c]$$
$$= 1 - P[U_1 < c, U_2 < c]$$
$$= 1 - \frac{c}{1} \times \frac{c}{1} \quad \text{where } c \in (0,1)$$
$$1 - c^2 = 0.19$$
$$0.81 = c^2$$
$$c = 0.9$$

(b) $$\beta_1(\theta) = \begin{cases} 0.19 ; & \theta = 0 \\ P[U_2 > 0.81] ; & \theta = 0.8 \end{cases} = \begin{cases} 0.19 ; & \theta = 0 \\ 0.99 ; & \theta = 0.8 \end{cases}$$

$\rightarrow U_2 \sim \text{Unif}(0.8, 1.8)$

$$\frac{1.8 - 0.81}{1} = 0.99$$

$$\beta_2(\theta) = \begin{cases} 0.19 \\ 1 - P[U_1 < 0.9] P[U_2 < 0.9] ; & \theta = 0.8 \end{cases} ; \theta = 0 = \begin{cases} 0.19 \\ 1 - 0.1 \times 0.1 \end{cases} = \begin{cases} 0.19 \\ 0.99 \end{cases}$$

11th April:

Correspondence between testing and confidence intervals:

Theorem: $X_1, X_2, \dots X_n \overset{iid}{\sim} f(x|\theta)$ (pmf/pdf)

$\theta \in \Theta$ is an unknown parameter

For each $\theta_0 \in \Theta$, let $A(\theta_0)$ be the acceptance region of a level α test for $H_0: \theta = \theta_0$

For each $\underset{\sim}{x} \in \mathcal{X}$, let $C(\underset{\sim}{x}) = \{\theta_0 \in \Theta \mid \underset{\sim}{x} \in A(\theta_0)\}$, then the random set $C(\underset{\sim}{X})$ is a $(1-\alpha)$ confidence set for θ, i.e

$$P_\theta[\theta \in C(\underset{\sim}{X})] \geq 1-\alpha \quad \forall \theta \in \Theta$$

Conversely, let $C(\underset{\sim}{X})$ be a $(1-\alpha)$ confidence set, i.e (1) holds

For any $\theta_0 \in \Theta$, define $A(\theta_0) = \{\underset{\sim}{x} \in \mathcal{X} \mid \theta_0 \in C(\underset{\sim}{x})\}$ then $A(\theta_0)$ is the acceptance region of level α test for $H_0: \theta = \theta_0$

Proof: See that, by construction, $\underset{\sim}{x} \in A(\theta_0) \Leftrightarrow \theta_0 \in C(\underset{\sim}{x})$

Now $A(\theta_0)$ is the acceptance region of a test for $H_0: \theta = \theta_0$

then, $P_{\theta_0}[\underset{\sim}{X} \notin A(\theta_0)] \leq \alpha \quad \forall \theta_0 \in \Theta$

$\Leftrightarrow P_{\theta_0}[\underset{\sim}{X} \in A(\theta_0)] \geq 1-\alpha \quad \forall \theta_0 \in \Theta$

$\Leftrightarrow P_{\theta_0}[\theta_0 \in C(\underset{\sim}{X})] \geq 1-\alpha \quad \forall \theta_0 \in \Theta$

$\Leftrightarrow C(\underset{\sim}{X})$ is a $1-\alpha$ confidence set

Note: If test was size α, then it corresponds to a confidence set $C(\underset{\sim}{X})$ satisfying

$$P_{\theta_0}[\theta_0 \in C(\underset{\sim}{X})] = 1-\alpha \quad \forall \theta_0 \in \Theta$$

Note: For confidence intervals we need to choose a level α test with $H_0: \theta = \theta_0$ v/s $H_1: \theta \neq \theta_0$

Eg: $X_1, X_2, \dots X_n \overset{iid}{\sim} N(\theta, \sigma^2)$ $\theta \in \Theta = \mathbb{R}$ is unknown but σ^2 is known, now consider testing

$H_0: \theta = \theta_0$ v/s $H_1: \theta \neq \theta_0$

We know that acceptance region of the most natural size α test has acceptance region

$$A(\theta_0) = \left\{ \underset{\sim}{x} \mid |\bar{x} - \theta_0| \leq z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right\}$$

$$= \left\{ \underset{\sim}{x} \in \mathbb{R}^n \mid \theta_0 - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \bar{x} \leq \theta_0 + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right\}$$

where $z_{\alpha/2}$ is s.t $\phi(z_{\alpha/2}) = 1-\alpha/2$

So, by inverting this test, we get $C(\underset{\sim}{x}) = \{\theta_0 \in \mathbb{R} \mid \underset{\sim}{x} \in A(\theta_0)\}$

$$= \left\{\theta_0 \in \mathbb{R} \mid \theta_0 - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \bar{x} \leq \theta_0 + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right\}$$

$$= \left\{\theta_0 \in \mathbb{R} \mid \bar{x} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \theta_0 \leq \bar{x} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right\}$$

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

so, the theorem $\Rightarrow C(\underset{\sim}{X}) = \left[\bar{X} - z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}},\ \bar{X} + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right]$ is a confidence interval for θ satisfying

$$P_\theta[\theta \in C(\underset{\sim}{X})] = 1-\alpha \quad \forall\, \theta \in \mathbb{R}$$

Exe: converse to above example, start with CI (confidence interval)

$C(\underset{\sim}{X}) = \left[\bar{X} - z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}},\ \bar{X} + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right]$ satisfying

$P_\theta[\theta \in C(\underset{\sim}{X})] = 1-\alpha \quad \forall\, \alpha \in \mathbb{R}$

by inverting this CI, we get the acceptance region of a size α test for $H_0: \theta = \theta_0$ v/s $H_1: \theta \neq \theta_0$ as follows:

$$A(\theta_0) = \left\{\underset{\sim}{x} \in \mathbb{R}^n \,\middle|\, \theta_0 \in C(\underset{\sim}{x})\right\} \quad \rightarrow \text{above above}$$

now that $\left\{\underset{\sim}{x} \in \mathbb{R}^n \,\middle|\, \theta_0 - z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \leq \bar{x} \leq \theta_0 + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right\}$ so

theorem $\Rightarrow A(\theta_0)$ is acceptance region of a size α test for testing $H_0: \theta = \theta_0$ v/s $H_1: \theta \neq \theta_0$

Note: The $P[|T| > t_{n-1, \frac{\alpha}{2}}]$ test is called the (two-sided) t-test

Exe: $X_1, X_2, \ldots X_n \overset{iid}{\sim} N(\theta, \sigma^2)$ with θ, σ^2 unknown, fix $\theta_0 \in \mathbb{R}$, consider the acceptance region of t-test for testing

$H_0: \theta = \theta_0$ v/s $H_1: \theta \neq \theta_0$ $\rightarrow$ above above

invert this test and obtain a CI $C(\underset{\sim}{X})$ for θ satisfying

$$P_\theta[\theta \in C(\underset{\sim}{X})] = 1-\alpha \quad \forall \theta \in \mathbb{R}$$

Exe: Suppose $X \sim f(x|\theta)$ (pdf) where $f(x|\theta) = \frac{1}{2}e^{-|x-\theta|}$; $x \in \mathbb{R}$

(double exponential density) find a CI $C(\underset{\sim}{X})$ for θ that is symmetric around $\underset{\sim}{X}$ and has confidence coeff $(1-\alpha)$

Now invert this test to get a size α test for $\rightarrow$ above above

$H_0: \theta = 0$ v/s $H_1: \theta \neq 0$

Consistent Estimators:

Defn: A sequence $\{T_n\}_{n \geq 1}$ of point estimators of θ is called consistent for θ if for all $\varepsilon > 0$

$$\lim_{n \to \infty} P_\theta[|T_n - \theta| \leq \varepsilon] = 1$$

so, the idea is that for large n, T_n will be very close to θ with very high probability

Eg: $X_1, X_2, \ldots X_n \overset{iid}{\sim} f(x|\theta)$ (pmf/pdf) with θ = population mean $= \mathbb{E}[X_1]$ (finite)

then WLLN states that the sample mean sequence

$$\bar{X}_n = \frac{X_1 + X_2 + \cdots + X_n}{n} \text{ is consistent for } \theta$$

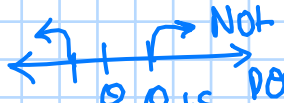
Eg: $X_1, X_2, \dots X_n \overset{iid}{\sim} \text{Unif}[0,\theta]$ $\theta \in (0,\infty) = \Theta$

MLE of $\theta = \hat{\theta}_n = X_{(n)}$ $n \geq 1$

claim is that $\{\hat{\theta}_n\}_{n\geq 1}$ is consistant for

fix $\varepsilon > 0$, $P_\theta[|\hat{\theta}_n - \theta| > \varepsilon] \to 0$ as $n \to \infty$, observe that

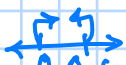
$$P_\theta[|\hat{\theta}_n - \theta| > \varepsilon] = P_\theta[|X_{(n)} - \theta| > \varepsilon]$$



$$= P_\theta[X_{(n)} < \theta - \varepsilon]$$

$$= P_\theta[X_{(n)} \leq \theta - \varepsilon]$$

$(\because \text{cont dist})$



$$= P_\theta[X_1 \leq \theta - \varepsilon]^n \quad (\because \text{done before})$$

$$= \left(\frac{\theta-\varepsilon}{\theta}\right)^n \quad \text{as } \left(\frac{\theta-\varepsilon}{\theta}\right) < 1$$

$$\Rightarrow \left(\frac{\theta-\varepsilon}{\theta}\right)^n \to 0 \text{ as } n \to \infty$$

Theorem: If a sequence of estimators $\{\hat{\theta}_n\}$ of θ satisfies

(i) $\mathbb{E}[\hat{\theta}_n] \to \theta$ as $n \to \infty$

(ii) $\text{Var}[\hat{\theta}_n] \to 0$ as $n \to \infty$

then the seq of estimators is consistant for θ

proof: fix $\varepsilon > 0$, to show

$$P[|\hat{\theta}_n - \theta| > \varepsilon] \to 0 \text{ as } n \to \infty$$

Note that, $0 \leq P_\theta[|\hat{\theta}_n - \theta| > \varepsilon] = P_\theta[(\hat{\theta}_n - \theta)^2 > \varepsilon^2]$

$$\overset{\text{markov inequality}}{\leq} \frac{\mathbb{E}[(\hat{\theta}_n - \theta)^2]}{\varepsilon^2}$$

$$= \frac{(\mathbb{E}_\theta(\hat{\theta}_n) - \theta)^2 + \text{Var}_\theta(\hat{\theta}_n)}{\varepsilon^2} \to 0 \quad (\because (i), (ii)) \text{ as } n \to \infty$$

then by sandwich theorem

$$\lim_{n\to\infty} P[|\hat{\theta}_n - \theta| > \varepsilon] = 0$$

Eg: use above theorem to show the MLE $\hat{\theta}_n = X_{(n)}$ is consistant for θ where $X_1, X_2, \dots X_n \overset{iid}{\sim} \text{Unif}(0,\theta)$ → done down

Eg: suppose $X_1, X_2, \dots X_n \overset{iid}{\sim} f(x|\theta)$ (shifted exponential) where

$f(x|\theta) = e^{-(x-\theta)}$ $x \geq \theta$ (shifted standard exponential density)

(a) find MLE $\hat{\theta}_n$ of θ based on $X_1, X_2, \dots X_n$ → done down

(b) show that above MLE $\hat{\theta}_n$ is consistant for θ

Ex: Converse to above example, start with CI (confidence interval)

$C(\underset{\sim}{X}) = [\bar{x} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}]$ satisfying

$P_\theta[\theta \in C(\underset{\sim}{X})] = 1-\alpha \quad \forall \alpha \in \mathbb{R}$

by inverting this CI, we get the acceptance region of a size α test for $H_0: \theta = \theta_0$ v/s $H_1: \theta \neq \theta_0$ as follows:

$A(\theta_0) = \{\underset{\sim}{x} \in \mathbb{R}^n \mid \theta_0 \in C(\underset{\sim}{x})\}$

now that $\{\underset{\sim}{x} \in \mathbb{R}^n \mid \theta_0 - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \bar{x} \leq \theta_0 + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\}$ so

theorm $\Rightarrow A(\theta_0)$ is acceptance region of a size α test for testing $H_0: \theta = \theta_0$ V/S $H_1: \theta \neq \theta_0$

Ans:

$C(\underset{\sim}{X}) = \left[\bar{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right]$

$A(\theta_0) = \{\underset{\sim}{x} \in \mathbb{R}^n \mid \bar{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \theta_0 \leq \bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\}$

$= \{\underset{\sim}{x} \in \mathbb{R}^n \mid |\bar{x} - \theta_0| \leq z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\}$

from above $A(\theta_0)$ is Acceptance region

Ex: $X_1, X_2, \ldots X_n$ iid $N(\theta, \sigma^2)$ with θ, σ^2 unkn, fix $\theta_0 \in \mathbb{R}$, consider the acceptance region of t-test for testing

$H_0: \theta = \theta_0$ v/s $H_1: \theta \neq \theta_0$

invert this test and obtain a CI $C(\underset{\sim}{X})$ for θ satisfying

$P_\theta[\theta \in C(\underset{\sim}{X})] = 1-\alpha \quad \forall \theta \in \mathbb{R}$

Ans: $A(\theta_0): \frac{|\bar{x} - \theta_0|}{s/\sqrt{n}} < t_{n-1, \alpha/2}$

$C(\underset{\sim}{X}) = \{\theta_0 \in \Theta \mid \underset{\sim}{x} \in A(\theta_0)\}$

$= \{\theta_0 \in \Theta \mid \frac{|\bar{x} - \theta_0|}{s/\sqrt{n}} < t_{n-1, \alpha/2}\}$

$|\bar{x} - \theta_0| < s\sqrt{n}\, t_{n-1, \frac{\alpha}{2}}$

$-s\sqrt{n}\, t_{n-1, \frac{\alpha}{2}} < \bar{x} - \theta_0 < s\sqrt{n}\, t_{n-1, \frac{\alpha}{2}}$

$-s\sqrt{n}\, t_{n-1, \frac{\alpha}{2}} < \theta_0 - \bar{x} < +s\sqrt{n}\, t_{n-1, \frac{\alpha}{2}}$

$C(\underset{\sim}{X}) = \left[\bar{x} - s\sqrt{n}\, t_{n-1, \frac{\alpha}{2}}, \bar{x} + s\sqrt{n}\, t_{n-1, \frac{\alpha}{2}}\right]$

$P_\theta[\theta \in C(\underset{\sim}{X})] = 1-\alpha \quad \forall \theta \in \mathbb{R}$ from theorm

Ex: Suppose $X \sim f(x|\theta)$ (pdf) where $f(x|\theta) = \frac{1}{2} e^{-|x-\theta|}$; $x \in \mathbb{R}$

(double exponential density) find a CI $C(X)$ for θ that is symmetric around X and has confidence coeff $(1-\alpha)$

Now invert this test to get a size α test for $H_0: \theta = 0$ v/s $H_1: \theta \neq 0$

Ans:

$$f(x|\theta) = \frac{1}{2} e^{-|x-\theta|}$$

$C(X) = [g(X) - c', g(X) + c']$ as symmetric

$$P_\theta[\theta \in C(X)] = P_\theta[g(X) - c' \leq \theta \leq g(X) + c']$$
$$= P_\theta[X - c' \leq \theta \leq X + c']$$
$$= P_\theta[|X - \theta| \leq c']$$

let $x - \theta = u$ $\quad f(u|\theta) = \frac{1}{2} e^{-u}$

$$= \int_{-c'}^{c'} \frac{1}{2} e^{-|u|} du$$
$$= \int_0^{c'} e^{-u} du$$

$$= 1 - e^{-c'} = 1 - \alpha$$

$e^{-c'} = \alpha \quad -c' = \ln(\alpha) \Rightarrow c' = \ln\left(\frac{1}{\alpha}\right)$

So, $C(X) = \left[X - \ln\left(\frac{1}{\alpha}\right), X + \ln\left(\frac{1}{\alpha}\right)\right]$

$$A(\theta_0) = \{x \mid \theta_0 \in C(X)\}$$

$$= \{x \mid |x - \theta_0| \leq \ln\left(\frac{1}{\alpha}\right)\}$$

So, $R = \{x \mid |x - \theta_0| \geq \ln\left(\frac{1}{\alpha}\right)\}$

size α as $\beta(\theta_0) = P_{\theta_0}[|x - \theta_0| \geq \ln\left(\frac{1}{\alpha}\right)]$

$$= 2 \int_{\ln\left(\frac{1}{\alpha}\right)}^{\infty} \frac{1}{2} e^{-u} du$$

$$= e^{-\ln\left(\frac{1}{\alpha}\right)} = \alpha$$

Ex: use above theorem to show the MLE $\hat{\theta}_n = X_{(n)}$ is consistent for θ where $X_1, X_2, \ldots X_n \overset{iid}{\sim} \text{Unif}(0, \theta)$

Ans: $\hat{\theta}_n = X_{(n)}$ $X_1, X_2, \ldots X_n \overset{iid}{\sim} \text{Unif}(0,\theta)$

$$\mathbb{P}[X_{(n)} \le x] = \left(\mathbb{P}[X_1 \le x]\right)^n$$

$$f_{X_{(n)}}(x|\theta) = \frac{d}{dx}\left(\frac{x}{\theta}\right)^n = \frac{n}{\theta}\left(\frac{x}{\theta}\right)^{n-1}$$

now $\mathbb{E}[\hat{\theta}_n] = \int_0^\theta \frac{n}{\theta} x \left(\frac{x}{\theta}\right)^{n-1} dx$

$$= \frac{1}{\theta^n} n \left| \frac{x^{n+1}}{n+1} \right|_0^\theta = \frac{n}{n+1}\theta$$

$$\text{Var}(\hat{\theta}_n) = \mathbb{E}[\hat{\theta}_n^2] - \left(\mathbb{E}[\hat{\theta}_n]\right)^2$$

$$= \int_0^\theta \frac{n}{\theta} x^2 \left(\frac{x}{\theta}\right)^{n-1} dx - \left(\frac{n}{n+1}\right)^2 \theta^2$$

$$= \frac{n}{\theta^n} \frac{\theta^{n+2}}{n+2} - \left(\frac{n}{n+1}\right)^2 \theta^2$$

$$= \left(\frac{n}{n+2} - \left(\frac{n}{n+1}\right)^2\right)\theta^2$$

$\lim_{n\to\infty} \frac{n}{n+1}\theta = \theta$

$\lim_{n\to\infty} \left(\frac{n}{n+2} - \left(\frac{n}{n+1}\right)^2\right)\theta^2 = 0$

$\Rightarrow$ MLE is consistent

Exe: Suppose $X_1, X_2, \ldots X_n \overset{iid}{\sim} f(x|\theta)$ (shifted exponential) where

$f(x|\theta) = e^{-(x-\theta)}$ $x \ge \theta$ (shifted standard exponential density)

(a) Find MLE $\hat{\theta}_n$ of θ based on $X_1, X_2, \ldots X_n$

(b) Show that above MLE $\hat{\theta}_n$ is consistent for θ

Ans: $f(x|\theta) = e^{-(x-\theta)}$ $x \ge \theta$

(a) $L(\underset{\sim}{x}|\theta) = \prod_{i=1}^n e^{-(x_i-\theta)} = \begin{cases} e^{-\left(\sum_{i=1}^n x_i\right) + n\theta} & ; (x_1, \ldots x_n) \in (\theta, \infty)^n \\ 0 & ; \text{otherwise} \end{cases}$

$\frac{dL(\underset{\sim}{x}|\theta)}{d\theta} = e^{-\left(\sum_{i=1}^n x_i\right)} e^{n\theta}(n) > 0 \rightarrow$ inc function

also we have to maximise $e^{-(\sum x_i) + n\theta}$

by maximising θ

but $\theta < x_1, \theta < x_2, \ldots \theta < x_n \Rightarrow$ to maximise

$$\hat{\theta}_{MLE} = X_{(1)}$$

now, $P[X_{(1)} < x] = 1 - P[X_{(1)} \geq x]$

$$= 1 - \left(\int_x^\infty e^{-(x-\theta)} dx \right)^n$$

$$= 1 - e^{n\theta} \left[e^{-nx} \right]$$

$$= 1 - e^{-n(x-\theta)}$$

now $f_{X_{(1)}}(x|\theta) = n e^{-n(x-\theta)}$

$$\int_\theta^\infty x\, n e^{-n(x-\theta)} dx = n e^{n\theta} \int_\theta^\infty x e^{-nx} dx$$

$$= n e^{n\theta} \left[x \frac{e^{-nx}}{-n} + \int \frac{e^{-nx}}{+n} \right]$$

$$= n e^{n\theta} \left[\frac{\theta}{n} e^{-n\theta} + \frac{1}{n^2} \left[e^{-n\theta} \right] \right]$$

$$= \theta + \frac{1}{n}$$

thus $\lim_{n \to \infty} E[\hat{\theta}_{MLE}] = \theta$

also, $\int_\theta^\infty x^2 n e^{-n(x-\theta)} dx = n e^{n\theta} \int_\theta^\infty x^2 e^{-nx} dx$

$$= n e^{n\theta} \left[-\frac{x^2 e^{-nx}}{n} + \int 2x e^{-nx} \right]$$

$$= n \left[\frac{\theta^2}{n} + 2\left(\frac{\theta}{n^2} + \frac{1}{n^2} \right) \right]$$

$$= \theta^2 + \frac{2\theta}{n} + \frac{2}{n}$$

$$Var(\hat{\theta}_{MLE}) = \theta^2 + \frac{2\theta}{n} + \frac{2}{n} - \theta^2 - \frac{1}{n^2} - 2\frac{\theta}{n}$$

$$= \frac{2}{n} - \frac{1}{n^2} \to 0 \text{ as } n \to \infty$$

14th April:

Fisher information:

Suppose $X \sim f(x|\theta)$ (pmf/pdf) ← one-dim unknown parameter $\theta \in \Theta$

assumption 1: $\frac{\partial}{\partial\theta} f(x|\theta)$ exist $\forall \theta \in \Theta$

score function, $\psi(x,\theta) = \frac{\partial}{\partial\theta} \log f(x|\theta) = \frac{\frac{\partial}{\partial\theta} f(x|\theta)}{f(x|\theta)}$

assumption 2: $\{x : f(x|\theta) > 0\}$ does not depend on θ

assumption 3: Another regularity condition that ensures that we can take the (partial) differential wrt θ inside integral/sum

Lemma: $\mathbb{E}[\psi(x,\theta)] = 0$

proof: (cont case) take any $\theta \in \Theta$

$$\mathbb{E}_\theta[\psi(x,\theta)] = \int_{\mathbb{R}} \psi(x,\theta) f(x|\theta)\, dx$$

$$= \int_{\mathbb{R}} \frac{\frac{\partial}{\partial\theta} f(x|\theta)}{\cancel{f(x|\theta)}} \cancel{f(x|\theta)}\, dx$$

$$= \int_{\mathbb{R}} \frac{\partial}{\partial\theta} f(x|\theta)\, dx$$

$$= \frac{\partial}{\partial\theta} \int_{\mathbb{R}} f(x|\theta)\, dx \quad (\because \text{assumption 3})$$

$$= \frac{\partial}{\partial\theta}(1) = 0$$

similar proof in discrete case

Ex: show the discrete case for above lemma → **Take down**

Defn: (Fisher Information) The fisher information of the distribution with pmf/pdf $f(x|\theta)$ is defined as

$$g(\theta) = \text{var}_\theta[\psi(X,\theta)]$$

$$= \mathbb{E}_\theta[(\psi(X,\theta))^2]$$

Eg: $X \sim \text{Exp}(\lambda) \Rightarrow f(x|\lambda) = \lambda e^{-\lambda x} \quad x > 0$

$$\Rightarrow \log f(x|\lambda) = \log\lambda - \lambda x$$

$$\Rightarrow \psi(x,\lambda) = \frac{\partial}{\partial\lambda} \log f(x|\lambda)$$

$$= \frac{1}{\lambda} - x$$

$$g(\lambda) = \text{var}_\lambda(\psi(X,\lambda)) = \text{var}_\lambda\left(\frac{1}{\lambda} - X\right) = \text{var}(X) = \frac{1}{\lambda^2}$$

Eg: $X \sim \text{Ber}(\theta) \quad \theta \in (0,1) = \Theta$

$$f(x|\theta) = \theta^x(1-\theta)^{1-x}$$
$$\Rightarrow \log f(x|\theta) = x\log\theta + (1-x)\log(1-\theta)$$
$$\Rightarrow \psi(x,\theta) = \frac{d}{d\theta}\log f(x|\theta) = \frac{x}{\theta} - \left(\frac{1-x}{1-\theta}\right)$$

$$\Rightarrow g(\theta) = \text{Var}_\theta\left[\frac{x}{\theta} - \left(\frac{1-x}{1-\theta}\right)\right]$$

$$= \text{Var}\left[\frac{x}{\theta}\right] + \text{Var}\left[\frac{1-x}{1-\theta}\right] - 2\text{Cov}\left[\frac{x}{\theta}, \frac{1-x}{1-\theta}\right]$$

$$= \frac{1}{\theta^2}\text{Var}(X) + \frac{1}{(1-\theta)^2}\text{Var}(X) + \frac{2}{\theta(1-\theta)}\text{Var}(X)$$

$$= \left[\frac{1}{\theta} + \frac{1}{1-\theta}\right]^2 \theta(1-\theta)$$

$$= \frac{1}{\theta(1-\theta)}$$

Exe: Compute fisher information for

① $X \sim \text{Poi}(\theta)$
② $X \sim N(0,1)$
③ $X \sim N(0,\tau^2)$ $\tau^2 = \text{Var}(X)$

→ done down

Cramer-Rao lower bound / Information inequality:

Suppose $X_1, X_2, \dots X_n \overset{iid}{\sim} f(x|\theta)$ (pmf/pdf) with $\theta \in \Theta \subseteq \mathbb{R}$

$\underset{\sim}{X} = (X_1, X_2, \dots X_n)$

assumption: a_1, a_2, a_3 (prov 3 assumptions hold)

Let, $\hat{\theta}(\underset{\sim}{X})$ be a point estimator of θ with finite expectation

$$g(\theta) = E_\theta[\hat{\theta}(\underset{\sim}{X})] \quad \theta \in \Theta$$ and finite variance

assumption: the fisher information $q(\theta)$ is finite, and strictly positive $\forall \theta \in \Theta$

then, $\forall \theta \in \Theta$ $\text{Var}_\theta(\hat{\theta}(\underset{\sim}{X})) \geq \frac{(g'(\theta))^2}{n\,q(\theta)}$

in particular if $\hat{\theta}(\underset{\sim}{X})$ is unbiased for θ, then $\text{Var}_\theta(\hat{\theta}(\underset{\sim}{X})) \geq \frac{1}{n\,q(\theta)}$ $\forall \theta \in \Theta$

Note: if $n=1 \Rightarrow \text{Var}_\theta(\hat{\theta}) \geq \frac{1}{q(\theta)}$ i.e if $q(\theta)$ is small, best unbiased estimator of θ based on X_1 has high variance

In this sense, $q(\theta)$ measures the amount of information X_1 has about θ

Def: An unbiased estimator $\hat{\theta}(\underset{\sim}{X})$ of θ is called efficient for θ if it attains the Cramer Rao lower bound, i.e $\text{Var}_\theta(\hat{\theta}(\underset{\sim}{X})) = \frac{1}{n\,q(\theta)}$ $\forall \theta \in \Theta$

Note: The candidate has to be UMVUE, as if its var > bound, then no such efficient θ

Eg: $X_1, X_2, \dots X_n \overset{iid}{\sim} Ber(\theta)$ $\bar{X}$ = UMVUE of θ

$$\psi(\theta) = \frac{1}{\theta(1-\theta)} > 0 \quad \theta \in (0,1)$$

ψ satisfies all assumptions

$$\frac{1}{\psi(\theta)} = \theta(1-\theta)$$

$$\frac{1}{n\psi(\theta)} = \frac{1}{n}\theta(1-\theta) = \text{CR lower bound (CRLB)}$$

$$Var_\theta(\hat{\theta}) = Var_\theta(\bar{X}) = Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)$$

$$= \frac{Var(X_i)}{n}$$

$$= \frac{\theta(1-\theta)}{n} = \frac{1}{n\psi(\theta)} = CRLB$$

$\Rightarrow \bar{X}$ is an efficient estimator for θ

Eg: $X_1, X_2, \dots X_n \overset{iid}{\sim} Poi(\theta)$

$\bar{X}$ is UMVUE of θ, in this case $\psi(\theta) = \frac{1}{\theta} > 0$ ($\because$ Exercise above)

$$\Rightarrow CRLB = \frac{\theta}{n}$$

also, $Var_\theta(\bar{X}) = \frac{Var(X_1)}{n} = \frac{\theta}{n} = CRLB$

$\Rightarrow \bar{X}$ is an efficient estimator for θ

Eg: $X_1, X_2, \dots X_n \overset{iid}{\sim} Exp\left(\frac{1}{\theta}\right)$ pdf $f(x|\theta) = \frac{1}{\theta} e^{-x/\theta}$; $x > 0$

in this case, $\bar{X}$ is the UMVUE of θ ($\because$ exponential family result)

$$\log f(x|\theta) = -\log\theta - \frac{x}{\theta}$$

$$\Rightarrow \psi(x,\theta) = \frac{d}{d\theta}\log f(x|\theta) = -\frac{1}{\theta} - \frac{x}{\theta^2}$$

$$\Rightarrow \psi(\theta) = Var_\theta[\psi(X,\theta)] = Var_\theta\left(-\frac{1}{\theta} - \frac{X}{\theta^2}\right)$$

$$= \frac{\theta^2}{\theta^4} = \frac{1}{\theta^2} > 0 \quad \forall \theta \in (0,\infty)$$

$$\Rightarrow CRLB = \frac{1}{n\psi(\theta)} = \frac{\theta^2}{n}$$

in other words, $Var(\bar{X}) = \frac{Var(X_1)}{n} = \frac{\theta^2}{n} = CRLB$

$\Rightarrow \bar{X}$ is efficient for θ

Exe: Suppose $X_1, X_2, \dots X_n \overset{iid}{\sim} N(\mu, \tau)$ $\tau = Var(X_1)$

(a) find whether $\bar{x}$ is efficient for μ when $\tau=1$ → done above

(b) find whether $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$ is efficient for τ when $\mu=0$

Note: MLE is asymptotically efficient (efficient when sample size very large under certain regularity conditions)

Quiz-6 : 16th April 11:45 am to 12:15pm , LH102 , Hypothesis Testing

Endsem: 2 Hours , 40 marks, 5:30-7:30

Exe: show the discrete case for above lemma

Ans:

$$\mathbb{E}_{\theta}[\psi(x,\theta)] = \sum_{x\in R}\psi(x,\theta)f(x|\theta)$$

$$= \sum_{x\in R}\frac{df(x|\theta)}{d\theta}$$

$$= \frac{d}{d\theta}\sum_{x\in R}f(x|\theta)$$

$$= \frac{d}{d\theta}(1) = 0$$

Exe: compute fisher information for

① $X \sim Poi(\theta)$
② $X \sim N(\theta,1)$
③ $X \sim N(0,\tau)$ $\tau = Var(X)$

Ans:

① $X \sim POI(\theta)$

$$f(x|\theta) = \frac{\theta^x e^{-\theta}}{x!}$$

$$\frac{d}{d\theta}f(x|\theta) = \frac{x\theta^{x-1}e^{-\theta} - e^x e^{-\theta}}{x!}$$

$$\frac{\frac{d}{d\theta}f(x|\theta)}{f(x|\theta)} = x\theta^{-1} - 1 = \frac{x}{\theta} - 1 = \psi(x,\theta)$$

then $Var\left(\frac{X}{\theta}-1\right) = I(\theta) = \frac{1}{\theta^2}Var(X) = \frac{\theta}{\theta^2} = \frac{1}{\theta}$ (← $Poi(\theta)$)

② $X \sim N(\theta,1)$

$$f(x|\theta) = \frac{1}{\sqrt{2\pi}}e^{-\frac{(x-\theta)^2}{2}}$$

$$\log f(x|\theta) = \log\frac{1}{\sqrt{2\pi}} - \frac{(x-\theta)^2}{2}$$

$$\frac{d}{d\theta}\log f(x|\theta) = +2(x-\theta)(+1)$$

$$\text{var}(+(x-\theta)) = \text{var}(x) = 1$$

③ $x \sim N(0, e)$ $\quad e = \text{var}(x)$

$$f(x|e) = \frac{1}{\sqrt{2\pi e}} e^{-x^2/2e}$$

$$\log f(x|e) = \log \frac{1}{\sqrt{2\pi e}} - \frac{x^2}{2e}$$

$$\frac{d}{de}\log f(x|e) = \frac{d}{de}\log\left(\frac{1}{\sqrt{2\pi e}}\right) + \frac{1}{2e^2}x^2 = \psi(x, e)$$

$$\text{var}(\psi(x,e)) = \text{var}\left(\frac{1}{2e^2}x^2\right) = \frac{1}{4e^4}\text{var}(x^2) = \frac{1}{2e^2}$$

now $x \sim N(0, e)$

$$\Rightarrow \frac{x}{\sqrt{e}} \sim N(0,1)$$

$$\Rightarrow \frac{x^2}{e} \sim \chi_1^2 \quad \text{var}(\chi_1^2) = 2$$

$$\Rightarrow \text{var}\left(\frac{x^2}{e}\right) = 2$$

$$\Rightarrow \frac{\text{var}(x^2)}{e^2} = 2$$

$$\Rightarrow \text{var}(x^2) = 2e^2$$

Exe: suppose $x_1, x_2, \ldots x_n \overset{iid}{\sim} N(\mu, e)$ $\quad e = \text{var}(x_1)$

(a) find whether $\bar{x}$ is efficient for μ when $e = 1$

(b) find whether $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$ is efficient for e when $\mu = 0$

Ans: (a) from ② above

$$\gamma(\mu) = 1 \qquad \bar{x} \sim N\left(\theta, \frac{1}{n}\right)$$

$$\frac{1}{n\gamma(\mu)} = \frac{1}{n} \qquad \text{var}(\bar{x}) = \frac{1}{n}$$

so efficient ($\because \bar{x}$ is unbiased)

(b) s^2 $\quad E[s^2] = e$ so unbiased and

$$\gamma(e) = \frac{1}{2e^2}$$ from above

$$\frac{1}{nI(\theta)} = \frac{2\theta^2}{n}$$

$$\text{Var}(S^2) = \frac{2\theta^2}{n-1}$$

as $\frac{(n-1)S^2}{\theta} \sim \chi^2_{n-1}$

$$E[\chi^2_{n-1}] = n-1$$

$$\text{Var}[\chi^2_{n-1}] = 2(n-1)$$

$$\Rightarrow \text{Var}\left(\left(\frac{n-1}{\theta}\right)S^2\right) = 2(n-1)$$

$$\Rightarrow \frac{(n-1)^2}{\theta^2}\text{Var}(S^2) = 2(n-1)$$

$$\Rightarrow \text{Var}(S^2) = \frac{2\theta^2}{(n-1)}$$

but $\frac{1}{nI(\theta)} = \frac{2\theta^2}{n}$ so S^2 is not efficient