Recap:

① Complete statistic

Completeness is an extra condition over minimal sufficiency.

Our goal:

Basu's Theorem: If $T(\underline{X})$ is a complete sufficient statistic, then $T(\underline{X})$ is independent of any ancillary statistic.

proof:- Proof in the discrete case.

Let $S(\underline{X})$ be any ancillary statistic.

Then $P(S(\underline{X})=s)$ does not depend on θ.

$P(S(\underline{X})=s \mid T(\underline{X})=t)$ is also independent of θ.

as $T(\underline{X})$ is sufficient.

$$P(S(\underline{X})=s) = \sum_t P(S(\underline{X})=s \mid T(\underline{X})=t)\, P(T(\underline{X})=t) \quad \cdots ①$$

$$P(S(\underline{X})=s) = \sum_t P(S(\underline{X})=s)\, P(T(\underline{X})=t) \quad \cdots ②$$

Now compare ① & ②,

$$\sum_t P(S(\underline{X})=s \mid T(\underline{X})=t)\, P(T(\underline{X})=t) = \sum_t P(S(\underline{X})=s)\, P(T(\underline{X})=t)$$

$$\Rightarrow \sum_t \left\{ P(S(\underline{X})=s \mid T(\underline{X})=t) - P(S(\underline{X})=s) \right\} P(T(\underline{X})=t) = 0$$

if $g(t) = P(S(\underline{X}) = s \mid T(\underline{X}) = t) - P(S(\underline{X}) = s)$

$\Rightarrow \sum_t g(t)\, P(T(\underline{X}) = t) = 0$

$\Rightarrow E[g(T(\underline{X}))] = 0$ true for all θ

Since $T(\underline{X})$ is complete

$\Rightarrow P(g(T(\underline{X})) = 0) = 1$

with prob. 1, $P(S(\underline{X}) = s \mid T(\underline{X}) = t) = P(S(\underline{X}) = s)$

$\Rightarrow S(\underline{X})$ is independent of $T(\underline{X})$.

Example: $X_1, \dots, X_n \overset{iid}{\sim} Exp(\theta)$. Find $E\left[\frac{X_n}{\sum_{i=1}^{n} X_i}\right]$

$Exp(\theta)$ is a scale family.

any $\frac{X_i}{X_n}$ is an ancillary statistic.

clearly, $\frac{X_n}{\sum_{i=1}^{n} X_i} = \frac{1}{\sum_{i=1}^{n} \frac{X_i}{X_n}}$

Since, $\frac{X_i}{X_n}$ is ancillary $\Rightarrow \frac{1}{\sum_{i=1}^{n} \frac{X_i}{X_n}}$ is also

an ancillary statistic.

In this case $\sum_{i=1}^{n} X_i$ is complete sufficient for θ.

We know, by Basu's theorem,

$\frac{X_n}{\sum_{i=1}^{n} X_i}$ and $\sum_{i=1}^{n} X_i$ are independent.

$$E\left[\frac{X_n}{\sum_{i=1}^{n} X_i}\right] E\left[\sum_{i=1}^{n} X_i\right] = E\left[X_n\right]$$

$$\Rightarrow \quad E\left[\frac{X_n}{\sum_{i=1}^{n} X_i}\right] = \frac{E[X_n]}{\sum_{i=1}^{n} E[X_i]} = \frac{1/\theta}{n/\theta} = \frac{1}{n}$$

Exponential family of distributions

One parameter exponential family has a p.m.f (or p.d.f) given by

$$f_\theta(x) = h(x)\, c(\theta) \exp\left(\omega(\theta)\, t(x)\right), \quad -\infty < x < \infty.$$

Example: $X \sim Bin(n, p)$

$$f_p(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

$$= \binom{n}{x} \left(\frac{p}{1-p}\right)^x (1-p)^n = \binom{n}{x} (1-p)^n \exp\left(x \log \frac{p}{1-p}\right)$$

hence it belongs to an exponential family with $t(x) = x$, $\omega(p) = \log \frac{p}{1-p}$, $c(p) = (1-p)^n$.

Example: $X \sim Pois(\lambda)$

$$f_\lambda(x) = e^{-\lambda} \frac{\lambda^x}{x!}, \quad h(x) = \frac{1}{x!}, \; c(\lambda) = e^{-\lambda}, \; \omega(\lambda) = \log \lambda$$

Example: $X \sim N(\mu, 1)$

$$f_\mu(x) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{(x-\mu)^2}{2}\right\} = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{[x^2 - 2\mu x + \mu^2]}{2}\right\}$$

$$h(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad \omega(\mu) = \mu, \quad c(\mu) = \exp\left(-\mu^2/2\right)$$

$$\int_{-\infty}^{\infty} f_\theta(x) = 1 \Rightarrow \int_{-\infty}^{\infty} h(x)\, c(\theta) \exp\{w(\theta) t(x)\}\, dx = 1$$

$$\Rightarrow \frac{d}{d\theta} \int_{-\infty}^{\infty} h(x)\, c(\theta) \exp\{w(\theta)\, t(x)\}\, dx = 0$$

$$\Rightarrow \int_{-\infty}^{\infty} \frac{d}{d\theta}\left[h(x)\, c(\theta) \exp\{w(\theta)\, t(x)\}\right] dx = 0$$

$$\Rightarrow \int_{-\infty}^{\infty} \left[h(x)\, c'(\theta) \exp\{w(\theta) t(x)\} + h(x)\, c(\theta)\, t(x) w'(\theta) \exp\{w(\theta) t(x)\}\right] dx = 0$$

$$\Rightarrow \int_{-\infty}^{\infty} \underbrace{h(x)\, c(\theta) \exp\{w(\theta)\, t(x)\}} \frac{c'(\theta)}{c(\theta)}\, dx + \int_{-\infty}^{\infty} h(x)\, c(\theta)\, t(x) \exp\{w(\theta) t(x)\}\, w'(\theta)\, dx = 0$$

$$\Rightarrow \frac{c'(\theta)}{c(\theta)} + w'(\theta)\, E[T(X)] = 0$$

$$\Rightarrow E[t(X)] = -\frac{c'(\theta)}{c(\theta)\, w'(\theta)}$$

Similarly taking a second derivative you will be able to evaluate $E[t(X)^2]$ and $\mathrm{Var}(t(X))$ as a function of θ.

Condition that integral w.r.t. x and differential w.r.t. θ can be interchanged is known as a regularity condition which typically exists for any exponential family.

What if $X \sim N(\mu, \sigma^2)$, where both μ and σ^2 are unknown:

Multiparameter exponential family has a density

$$f_{\underline{\theta}}(x) = h(x)\, c(\underline{\theta}) \exp\left\{ \sum_{j=1}^{k} w_j(\underline{\theta})\, t_j(x) \right\}$$

where $\underline{\theta}$ is k-dimensional.

Let $X_1, \ldots, X_n \overset{iid}{\sim} f_{\underline{\theta}}(x)$

what is the sufficient statistic ?

$$f_{\theta}(x_1, \ldots, x_n) = \prod_{i=1}^{n} f_\theta(x_i)$$

$$= \prod_{i=1}^{n} \left\{ h(x_i)\, c(\underline{\theta}) \exp\left(\sum_{j=1}^{k} w_j(\underline{\theta})\, t_j(x_i) \right) \right\}$$

$$= \prod_{i=1}^{n} h(x_i) \left[c(\underline{\theta}) \right]^n \exp\left(\sum_{j=1}^{k} w_j(\theta) \sum_{i=1}^{n} t_j(x_i) \right)$$

By factorization thm. $\left(\sum_{i=1}^{n} t_1(x_i), \ldots, \sum_{i=1}^{n} t_k(x_i) \right)$ is a sufficient statistic.

Exp: $f_{\mu,\sigma^2}(x_1, \ldots, x_n) = \left(\frac{1}{\sqrt{2\pi\sigma}} \right)^n \exp\left\{ - \sum_{i=1}^{n} \frac{(x_i - \mu)^2}{2\sigma^2} \right\}$

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

$\left(\sum_{i=1}^{n} x_i, \sum_{i=1}^{n} x_i^2 \right)$ is a sufficient statistic for (μ, σ^2).

back ①

Remark: It can also be shown that

$\left(\sum_{i=1}^{n} t_1(x_i), \ldots, \sum_{i=1}^{n} t_K(x_i)\right)$ is a complete sufficient statistic if the set $\{(\omega_1(\underline{\theta}), \ldots, \omega_K(\underline{\theta})) : \underline{\theta}\}$ contains an open set in $\mathbb{R}^K$,

We will mainly try to find an ~~interval~~ rectangle $[a_1, b_1] \times \cdots \times [a_K, b_K]$, $a_i < b_i$ s.t. this rectangle is fully contained in the set $\{(\omega_1(\underline{\theta}), \ldots, \omega_K(\underline{\theta})) : \underline{\theta}\}$.

Example: $X_1, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$

$$f_{\mu,\sigma^2}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\left(\frac{x^2}{2\sigma^2} - \frac{2\mu x}{2\sigma^2} + \frac{\mu^2}{2\sigma^2}\right)\right\}$$

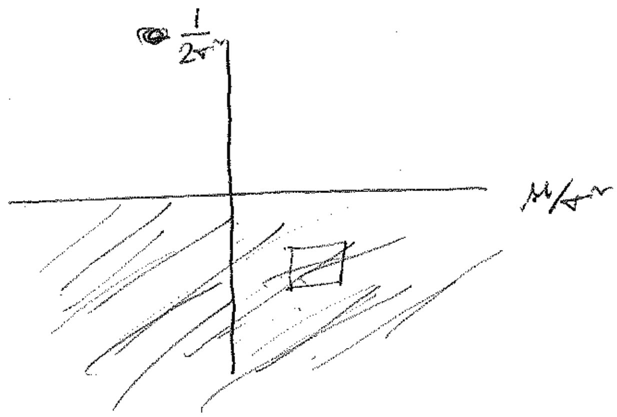
$\omega_1(\mu, \sigma^2) = \frac{\mu}{\sigma^2}$, $t_1(x) = x$

$\omega_2(\mu, \sigma^2) = -\frac{1}{2\sigma^2}$, $t_2(x) = x^2$

$$\{(\omega_1(\mu,\sigma^2), \omega_2(\mu,\sigma^2)) : -\infty < \mu < \infty, \sigma^2 > 0\}$$

$$= \left\{\left(\frac{\mu}{\sigma^2}, -\frac{1}{2\sigma^2}\right) : -\infty < \mu < \infty, \sigma^2 > 0\right\}.$$

back ②



this set clearly contains an open rectangle.

hence, $\left(\sum_{i=1}^{n} t_1(X_i), \sum_{i=1}^{n} t_2(X_i)\right) = \left(\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i^2\right)$

is a complete statistic

<u>Example:</u> $X \sim N(\theta, \theta^2)$.

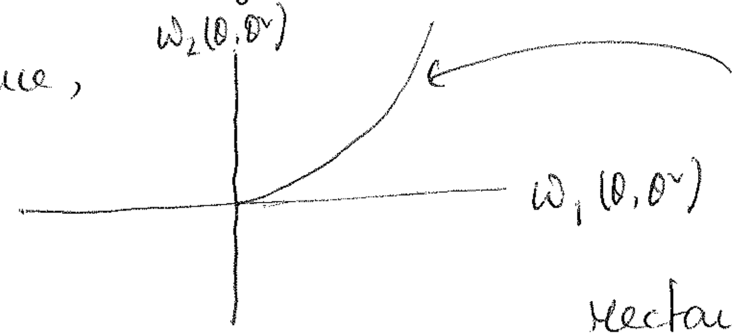
$$f_\theta(x) = \frac{1}{\sqrt{2\pi\theta^2}} \exp\left\{-\frac{(x-\theta)^2}{2\theta^2}\right\}$$

$$= \frac{1}{\sqrt{2\pi\theta^2}} \exp\left\{-\frac{1}{2}\left[\frac{x^2 - 2\theta x + \theta^2}{\theta^2}\right]\right\}$$

~~$\omega_1(\theta) = \frac{1}{\theta}$ and $\omega_2(\theta) =$~~

$\omega_1(\theta, \theta^2) = \frac{1}{\theta}$ and $\omega_2(\theta, \theta^2) = \frac{1}{\theta^2}$

hence,



thus it cannot contain a two dimensional rectangle in it.

hence ~~this is~~ $\left(\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i^2\right)$ is not a complete statistic in this case.

<u>back ③</u>

Comment: If parameters in an exponential family of distributions are not functions of each other, then the above result typically holds.

(1) ~~an~~ sufficient statistic
(2) Minimal " "
(3) Complete " "
(4) Complete and sufficient statistic is independent of any ancillary statistic.
(5) Factorization thm. for determining sufficient stat.
(6) Minimal sufficiency thm. to determine minimal sufficient,
(7) Completeness, we have already seen some techniques

Exponential family gives us a general result to find sufficient, minimal sufficient and complete statistic.

back (4)

Likelihood principle:

To know likelihood principle we need to know the definition of likelihood fn.

Def: Let $f_\theta(\underline{x})$ be the joint pdf or pmf of the sample $\underline{X} = (X_1, \dots, X_n)$. Then given that $\underline{X} = \underline{x}$ is observed, the function of θ defined by $L(\theta|\underline{x}) = f_\theta(\underline{x})$ is called the likelihood fn.

Likelihood principle says that if $\underline{x}$ and $\underline{y}$ are two sample points s.t. $L(\theta|\underline{x})$ is proportional to $L(\theta|\underline{y})$, i.e. $\exists$ a fn. $C(\underline{x}, \underline{y})$ s.t.

$$L(\theta|\underline{x}) = C(\underline{x}, \underline{y})\, L(\theta|\underline{y}), \quad \forall \theta,$$

then the conclusion drawn from $\underline{x}$ and $\underline{y}$ should be identical.

this means if $L(\theta_2|\underline{x}) = 3 \cdot L(\theta_1|\underline{x})$ then θ_2 is thrice probable for the data than θ_1.

Maximum likelihood estimator of a parameter came out of the concept of likelihood principle.

Example: X be the number of sucess in twelve Bernoulli trials with sucess prob. θ.

clearly $X \sim Bin(12, \theta)$

Suppose we observe 3 successes

$$L(\theta | X=3) = \binom{12}{3} \theta^3 (1-\theta)^9.$$

Let Y be the number of trials to get 3 successes.

$$Y \sim NegBin(3, \theta)$$

$$L(\theta | Y=12) = \binom{11}{2} \theta^3 (1-\theta)^9$$

$$L(\theta|X=3) \propto L(\theta|Y=12)$$

~~Likelih~~ Likelihood principle tells us that same inference on θ should be drawn by visualizing this experiment in two different ways.

We will check that

$$H_0: \theta = \frac{1}{2} \quad vs. \quad H_1: \theta > \frac{1}{2}$$

visualizing our experiment ~~as the~~ in the Binomial way gives us p-value of 0·07, while NegBin way gives us p-value 0·03.

In one case H_0 is not rejected, while in the other case it is rejected. Hence inference on θ is different, although ~~both~~ ~~Neg~~ likelihood.

principle ~~should~~ ~~indicates~~ ensures that same inference on θ should be drawn.

Thus this example violates the likelihood principle.

Recap:

① Sufficiency ② minimal sufficiency ③ completeness
④ ancillary.

These are tools to evaluate "how good" is the data reduction achieved by an estimator and how much information is lost, if any.

We will study how to find an "optimal" point estimator for a parameter from a model.

Qn: optimality in what sense?

Definition (Mean squared error): If $\tau(\theta) \neq 0$ is a function of θ and $T(\underline{X})$ is an estimator used to estimate $\tau(\theta)$, then Mean squared Error (MSE) of $T(\underline{X})$ is given by $E_\theta(T(\underline{X})-\tau(\theta))^2$.

Goal: Given any $\tau(\theta)$ (a function of θ), we want to find an estimator $T(\underline{X})$ which has the lowest MSE, uniformly over θ.

In other words, if $H(\underline{X})$ be another estimator of $\tau(\theta)$ then $E_\theta(T(\underline{X})-\tau(\theta))^2 \leq E_\theta(H(\underline{X})-\tau(\theta))^2 \quad \forall \theta$

Let $\tau(\theta)$ is a function of θ taking values 1, 5, 13.

$T_1(\underline{X}) = 1$, $T_2(\underline{X}) = 5$, $T_3(\underline{X}) = 13$.

$E_\theta(T_1(\underline{X})-1)^2 = 0$ for those θ's such that $\tau(\theta)=1$

$E_\theta(T_2(\underline{X})-5)^2 = 0$ for those θ's " " $\tau(\theta)=5$.

We restrict the class of estimators among which we are going to find out estimators with the best MSE.

Let, $C_\tau = \{T : E_\theta(T(\underline{X})) = \tau(\theta)\}$.

These class of estimators are called unbiased estimators.

If $T(\underline{X})$ is an estimator of $\tau(\theta)$, we know

$$E_\theta(T(\underline{X}) - \tau(\theta))^2 = E_\theta\left(T(\underline{X}) - E_\theta(T(\underline{X})) + E_\theta(T(\underline{X})) - \tau(\theta)\right)^2$$

$$= E_\theta\left(T(\underline{X}) - E_\theta(T(\underline{X}))\right)^2 + \left(E_\theta(T(\underline{X})) - \tau(\theta)\right)^2$$

$$= Var_\theta(T(\underline{X})) + Bias_\theta(T(\underline{X}))^2$$

$$Bias_\theta(T(\underline{X})) = E_\theta(T(\underline{X})) - \tau(\theta).$$

For $T \in C_\tau$, $Bias_\theta(T(\underline{X})) = 0$

$\Rightarrow$ MSE of $T(\underline{X}) = Var_\theta(T(\underline{X}))$.

Goal: We started with the goal to find an estimator $T(\underline{X})$ given a function $\tau(\theta)$ s.t. $T(\underline{X})$ has the minimum MSE over all estimators, uniformly over θ.

If we restrict our attention to estimators in C_τ, our goal becomes to find $T(\underline{X}) \in C_\tau$ s.t. $T(\underline{X})$ has the minimum variance over all estimators, uniformly over θ.

We want $T(\underline{X})$ s.t. $Var_\theta(T(\underline{X})) \le Var_\theta(H(\underline{X}))\ \forall \theta$,

where $H(\underline{X}), T(\underline{X}) \in C_\tau$.

If we can find such an estimator, it will be known as the Uniformly minimum variance unbiased estimator, or UMVUE.

Theorem: If $T(\underline{X})$ is UMVUE for $\tau(\theta)$, then $T(\underline{X})$ is unique.

Pf:- Lets say UMVUE is not unique.

So, $\exists$ $H(\underline{X})$ and $T(\underline{X})$ two UMVUES.

clearly, $Var_\theta(T(\underline{X})) = Var_\theta(H(\underline{X})) \quad \forall \theta$.

Create an estimator $T^*(\underline{X}) = \dfrac{H(\underline{X}) + T(\underline{X})}{2}$

$$Var_\theta(T^*(\underline{X})) = Var_\theta\left(\frac{H(\underline{X}) + T(\underline{X})}{2}\right)$$

$$= \frac{1}{4} Var_\theta(H(\underline{X})) + \frac{1}{4} Var_\theta(T(\underline{X})) + \frac{1}{2} Cov_\theta(H(\underline{X}), T(\underline{X}))$$

$$\leq \frac{1}{4} Var_\theta(H(\underline{X})) + \frac{1}{4} Var_\theta(T(\underline{X})) + \frac{1}{2}\sqrt{Var_\theta(H(\underline{X}))\, Var_\theta(T(\underline{X}))}$$

by Cauchy-Schwartz inequality

(for two random variable X_1, X_2, $Cov(X_1, X_2) \leq \sqrt{Var(X_1)Var(X_2)}$

$$= \frac{1}{4} Var_\theta(T(\underline{X})) + \frac{1}{4} Var_\theta(T(\underline{X})) + \frac{1}{2} Var_\theta(T(\underline{X}))$$

$$= Var_\theta(T(\underline{X}))$$

$\Rightarrow$ ~~$T^*(\underline{X})$ has less~~ $Var_\theta(T^*(\underline{X})) \leq Var_\theta(T(\underline{X})) \quad \forall \theta$.

Strict inequality can't happen as $T(\underline{X})$ is UMVUE.

$\Rightarrow Var_\theta(T^*(\underline{X})) = Var_\theta(T(\underline{X})) \quad \forall \theta$.

the equality has to be achieved in the Cauchy-Schwarz step, i.e. $\text{cov}_\theta(H(\underline{X}), T(\underline{X})) = \sqrt{\text{Var}_\theta(T(\underline{X}))\, \text{Var}_\theta(H(\underline{X}))}$

One must have, $H(\underline{X}) = a(\theta)\, T(\underline{X}) + b(\theta)$, for some $a(\underline{\theta})$ and $b(\underline{\theta})$.

$$\text{cov}_\theta(H(\underline{X}), T(\underline{X})) = a(\theta)\, \text{Var}_\theta(T(\underline{X})) \quad - - - - ①$$

But $\text{cov}_\theta(H(\underline{X}), T(\underline{X})) = \sqrt{\text{Var}_\theta(T(\underline{X}))\, \text{Var}_\theta(H(\underline{X}))}$

$$= \text{Var}_\theta(T(\underline{X})) \quad - - - - ②$$

by ① & ②

$\Rightarrow a(\theta) = 1$,

$$E_\theta(H(\underline{X})) = \tau(\theta) \Rightarrow E_\theta(T(\underline{X}) + b(\theta)) = \tau(\theta)$$

$$\Rightarrow b(\theta) + E_\theta(T(\underline{X})) = \tau(\theta)$$

$$\Rightarrow b(\theta) + \cancel{\tau(\theta)} = \cancel{\tau(\theta)} \quad \Rightarrow b(\theta) = 0.$$

This shows us that UMVUE if exists, is unique.

Let $T(\underline{X})$ be any unbiased estimator of $\tau(\theta)$.
Let $U(\underline{X})$ be an estimator s.t. $E_\theta(U(\underline{X})) = 0 \ \forall \theta$.

Lets define $T^*(\underline{X}) = T(\underline{X}) + a\,U(\underline{X})$, a is a const.

$$\text{Var}_\theta(T^*(\underline{X})) = \text{Var}_\theta(T(\underline{X})) + a^2\, \text{Var}_\theta(U(\underline{X})) + 2a\, \text{Cov}_\theta(T(\underline{X}), U(\underline{X}))$$

For some θ_0, if $\text{cov}_{\theta_0}(T(\underline{X}), U(\underline{X})) < 0$
we can choose an $a \in \left(0, -\dfrac{2\,\text{cov}_{\theta_0}(T(\underline{X}), U(\underline{X}))}{\text{Var}_{\theta_0}(U(\underline{X}))}\right)$.

then $2a\, Cov_{\theta_0}(T(\underline{X}), U(\underline{X})) + a^2 Var_{\theta_0}(U(\underline{X})) < 0$.

$\Rightarrow$ ~~Var~~ $Var_{\theta_0}(T^*(\underline{X})) < Var_{\theta_0}(T(\underline{X}))$

If we start from $T(\underline{X})$ which is the UMVUE, and if $Cov_{\theta_0}(T(\underline{X}), U(\underline{X})) < 0$ then we can find an estimator $T^*(\underline{X})$ in the above way which will contradict our assumption that $T(\underline{X})$ is the UMVUE.

This means $Cov_{\theta_0}(T(\underline{X}), U(\underline{X})) = 0 \quad \forall\, \theta_0$.

$\Rightarrow$ $T(\underline{X})$ if is the UMVUE has to be uncorrelated with any unbiased estimator of 0.

This shows that if $T(\underline{X})$ is the UMVUE it is uncorrelated with any unbiased estimator of 0.

If $T(\underline{X})$ is an unbiased estimator that is uncorrelated with any unbiased estimator of 0, does it mean $T(\underline{X})$ is the UMVUE?

Ans: Yes.

If $T^*(\underline{X})$ be any other unbiased estimator of $\tau(\theta)$, then $E_\theta[T(\underline{X}) - T^*(\underline{X})] = 0$

$\Rightarrow T(\underline{X}) - T^*(\underline{X})$ is an unbiased estimator of 0.

$$Var_\theta(T^*(\underline{X})) = Var_\theta(T(\underline{X}) + T^*(\underline{X}) - T(\underline{X}))$$

$$= Var_\theta(T(\underline{X})) + Var_\theta(T^*(\underline{X}) - T(\underline{X})) + 2Cov_\theta(T(\underline{X}), T^*(\underline{X}) - T(\underline{X}))$$

$$= Var_\theta(T(\underline{X})) + Var_\theta(T^*(\underline{X}) - T(\underline{X})) \geq Var_\theta(T(\underline{X}))$$

Example: $X \overset{iid}{\sim} U(\theta, \theta+1)$

$E_\theta\left(X - \frac{1}{2}\right) = \theta \Rightarrow X - \frac{1}{2}$ is an unbiased estimator of θ.

Qn: Is $X - \frac{1}{2}$ UMVUE?

Let $h(X)$ be an unbiased estimator of 0 for this model.

$$\Rightarrow E_\theta[h(X)] = 0 \Rightarrow \int_\theta^{\theta+1} h(x)\,dx = 0$$

$$\Rightarrow \frac{d}{d\theta}\int_\theta^{\theta+1} h(x)\,dx = 0 \Rightarrow h(\theta+1) - h(\theta) = 0$$

$h(x) = \sin(2\pi x)$

$\Rightarrow h(X) = \sin(2\pi X)$ is an unbiased estimator of 0.

$$\mathrm{Cov}\left(X - \frac{1}{2}, \sin(2\pi X)\right) = -\frac{\cos(2\pi\theta)}{2\pi} \neq 0$$

$\Rightarrow X - \frac{1}{2}$ is not uncorrelated with all unbiased estimator of 0.

$\Rightarrow X - \frac{1}{2}$ is not the UMVUE.

Rao - Blackwell theorem:

Let $W(\underline{X})$ be any unbiased estimator of $\tau(\theta)$. Let $T(\underline{X})$ be a sufficient statistic for θ. Let $\phi(T(\underline{X})) = E_\theta[W(\underline{X}) \mid T(\underline{X})]$. Then

① $\phi(T(\underline{X}))$ is an unbiased estimator of θ

② $\mathrm{Var}_\theta(\phi(T(\underline{X}))) \leq \mathrm{Var}_\theta(W(\underline{X}))$ with equality holding if and only if $\phi(T(\underline{X})) = W(\underline{X})$ w.p. 1.

back ①

Pf: Since $T(\underline{X})$ is sufficient $\underline{X} \mid T(\underline{X})$ is free of θ

$\Rightarrow \phi(T(\underline{X})) = E_\theta[W(\underline{X}) \mid T(\underline{X})]$ is free of θ.

$$E_\theta[\phi(T(\underline{X}))] = E_\theta[E_\theta[W(\underline{X}) \mid T(\underline{X})]]$$

$$= E_\theta[W(\underline{X})] = \tau(\theta)$$

$\Rightarrow \phi(T(\underline{X}))$ is unbiased.

$$Var_\theta(W(\underline{X})) = Var_\theta(E_\theta[W(\underline{X}) \mid T(\underline{X})]) + E_\theta[Var_\theta(W(\underline{X}) \mid T(\underline{X}))]$$

$$= Var_\theta(\phi(T(\underline{X}))) + \underbrace{E_\theta[Var_\theta(W(\underline{X}) \mid T(\underline{X}))]}_{\geq 0}$$

$$\geq Var_\theta(\phi(T(\underline{X})))$$

We started with an unbiased estimator $W(\underline{X})$ of $\tau(\theta)$. We ended up finding another unbiased estimator $\phi(T(\underline{X}))$ which has lesser variance.

Ex: $X_1, X_2, X_3 \overset{iid}{\sim} Ber(p)$.

$W(\underline{X}) = \frac{X_1 + X_2}{2}$ $\quad E[W(\underline{X})] = \frac{E[X_1] + E[X_2]}{2} = p$.

sufficient statistic of p is

$T(\underline{X}) = X_1 + X_2 + X_3$.

$$\phi(T(\underline{X})) = E_p\left[\frac{X_1 + X_2}{2} \,\middle|\, X_1 + X_2 + X_3\right]$$

$$E[T(\underline{X}) \mid T(\underline{X})] = T(\underline{X})$$

$$\Rightarrow E[X_1 + X_2 + X_3 \mid X_1 + X_2 + X_3] = X_1 + X_2 + X_3$$

back ②

$$E[X_1 \mid X_1+X_2+X_3] = E[X_2 \mid X_1+X_2+X_3] = E[X_3 \mid X_1+X_2+X_3]$$

$$\Rightarrow E[X_1+X_2+X_3 \mid X_1+X_2+X_3] = X_1+X_2+X_3$$

$$\Rightarrow E[X_1 \mid X_1+X_2+X_3] = E[X_2 \mid X_1+X_2+X_3] = E[X_3 \mid X_1+X_2+X_3]$$

$$= \frac{X_1+X_2+X_3}{3}$$

$$E[W(\underline{X}) \mid T(\underline{X})] = E\left[\frac{X_1+X_2}{2} \,\middle|\, X_1+X_2+X_3\right]$$

$$= \frac{E[X_1 \mid X_1+X_2+X_3] + E[X_2 \mid X_1+X_2+X_3]}{2}$$

$$= \frac{X_1+X_2+X_3}{3}$$

$$Var_p\left(E[W(\underline{X}) \mid T(\underline{X})]\right) = Var_p\left(\frac{X_1+X_2+X_3}{3}\right)$$

$$= \frac{p(1-p)}{3}$$

$$Var_p(W(\underline{X})) = Var_p\left(\frac{X_1+X_2}{2}\right)$$

$$= \frac{p(1-p)}{2}$$

$$\Rightarrow Var_p\left(E[W(\underline{X}) \mid T(\underline{X})]\right) < Var_p(W(\underline{X}))$$

back ③