

- Recap:
- ① Ancillary statistic.
 - ② Complete sufficient statistic is independent of ancillary.
 - ③ Exponential family of distributions.
- Multiparameter exponential family

$$f_{\theta}(x) = h(x) c(\theta) \exp \left\{ \sum_{i=1}^k \omega_i(\theta) t_i(x) \right\}$$

$x_1, \dots, x_n \stackrel{\text{iid}}{\sim} f_{\theta}(x)$ then by factorization theorem

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

Remark: $\left(\sum_{i=1}^n t_1(x_i), \dots, \sum_{i=1}^n t_k(x_i) \right)$ is a complete sufficient statistic if

$\{ (\omega_1(\theta), \dots, \omega_k(\theta)) : \theta \in \mathbb{H} \}$ contains an open set in \mathbb{R}^k .

If you take ~~$(a_1, b_1) \times \dots \times (a_k, b_k)$~~ ~~it's closed~~

• ~~if~~ there has to be some $(a_1, b_1) \times \dots \times (a_k, b_k)$

s.t. $\{ (\omega_1(\theta), \dots, \omega_k(\theta)) : \theta \in \mathbb{H} \}$ contains

the set $(a_1, b_1) \times \dots \times (a_k, b_k)$ fully.

② Example: $x_1, \dots, x_n \stackrel{iid}{\sim} \text{Ber}(p)$

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

$$\theta = p, \quad \omega(\theta) = \log \frac{p}{1-p}$$

$$\left\{ \log \frac{p}{1-p} : p \in (0, 1) \right\} = (-\infty, \infty)$$

thus obviously ~~(5, 6)~~ $(5, 6) \in \left\{ \log \frac{p}{1-p} : p \in (0, 1) \right\}$

which mean

$$f_p(x_1, \dots, x_n) = \exp\left\{\sum_{i=1}^n x_i \log \frac{p}{1-p}\right\} (1-p)^n$$

thus $\sum_{i=1}^n x_i$ is a complete sufficient statistic.

Example: $x_1, \dots, x_n \stackrel{iid}{\sim} N(\mu, \sigma^2)$

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

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

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

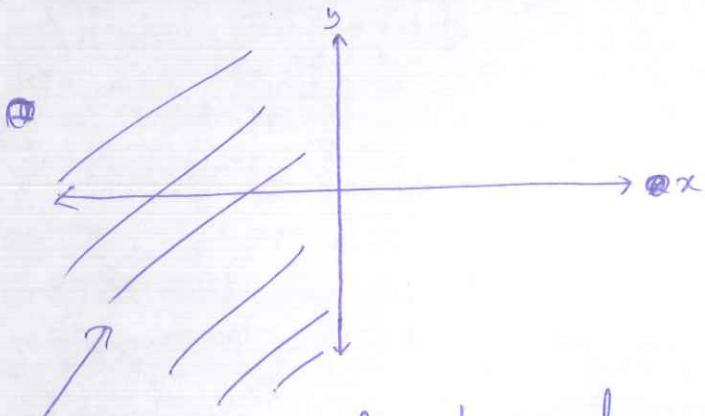
$\underline{\theta} = (\mu, \sigma^2)$, this density is in exponential

family with $c(\underline{\theta}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2} \frac{\mu^2}{\sigma^2}\right)$, $t(x) = 1$

$$\omega_1(\underline{\theta}) = -\frac{1}{2\sigma^2}, \quad \omega_2(\underline{\theta}) = \frac{\mu}{\sigma^2}, \quad t_1(x) = x^2, \quad t_2(x) = x.$$

$$\{(\omega_1(\underline{\theta}), \omega_2(\underline{\theta})) : \underline{\theta} \in \underline{\theta} \in \underline{\theta} \in (-\infty, \infty) \times (0, \infty)\}$$

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



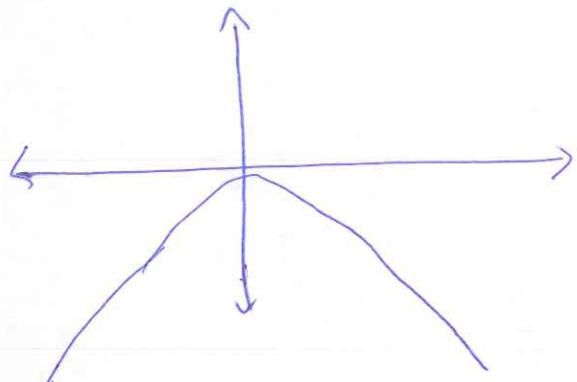
② set and it obviously contains a rectangle of the form $(a_1, b_1) \times (a_2, b_2)$ where both $a_i, b_i < 0$. thus $(\sum_{i=1}^n x_i, \sum_{i=1}^n x_i)$ is a complete sufficient statistic according to the theorem.

Example: $x_1, \dots, x_n \stackrel{iid}{\sim} N(\theta, \theta^2)$, $\theta > 0$.

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

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

$$\left\{ \left(-\frac{1}{2\theta}, \frac{1}{\theta} \right) : \theta \in (0, \infty) \right\}$$



③ (z_1, z_2)

$$z_2^2 = -\frac{1}{6} - 2z_1$$

this is a parabola which does not contain an open rectangle.

④

Likelihood principle

Def: (Likelihood fn.): Let $f_{\theta}(\underline{x})$ be the joint p.m.f. on p.d.f. of the sample $\underline{X} = (X_1, \dots, X_n)$. Then given that $\underline{X} = \underline{x}$ is observed, the function of $\underline{\theta}$ $L(\underline{\theta} | \underline{x}) = f_{\theta}(\underline{x})$ is called the Likelihood fn.

Likelihood principle:

If \underline{x} and \underline{y} are two points such that $L(\underline{\theta} | \underline{x})$ is proportional to $L(\underline{\theta} | \underline{y})$, that is there exists a constant $C(\underline{x}, \underline{y})$ s.t.

$$L(\underline{\theta} | \underline{x}) = C(\underline{x}, \underline{y}) L(\underline{\theta} | \underline{y}) \text{ for all } \underline{\theta},$$

then the conclusion drawn from \underline{x} and \underline{y} are identical. $C(\underline{x}, \underline{y})$ does not depend on $\underline{\theta}$, but may depend on \underline{x} and \underline{y} .

If there are $\underline{\theta}_1$ and $\underline{\theta}_2$ s.t. $L(\underline{\theta}_1 | \underline{x}) = 3 L(\underline{\theta}_2 | \underline{x})$ then $\underline{\theta}_1$ is twice "probable" as a value of $\underline{\theta}$ than $\underline{\theta}_2$.

Example: Let X be the number of successes in 12 Bernoulli trials with success prob. θ . Thus $X \sim \text{Bin}(12, \theta)$. Suppose we have observed 3 successes. Then the likelihood of θ is

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

Let Y be the number of trials required to get 3 successes, $Y \sim \text{Neg Bin}(3, \theta)$

The likelihood of θ here is

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

The two likelihoods are proportional to each other for all θ , by likelihood principle the inference of θ we get from one experiment is the same in the other experiment.

$H_0: \theta = \frac{1}{2}$ vs. $H_1: \theta > \frac{1}{2}$. In the first

experiment the p-value is 0.07 and in the second experiment the p-value is 0.03.

thus in the first experiment (with our convention of rejecting H_0 when p-value < 0.05) ~~the~~ H_0

is not rejected, while in the second experiment H_0 is rejected.

Hence the principle is not well founded.

Techniques to evaluate estimators

In the previous section we studied a few concepts on sufficiency, minimal sufficiency and completeness. Those are tools to evaluate "how good" is the data reduction achieved by an estimator and how much information is lost, if any. In this section, we will use these tools to create "optimal" point estimators for parameters. First we need a metric to evaluate ~~any~~ any estimator.

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 the mean squared error (MSE) of $T(\underline{x})$ is given by

$$E_\theta [(T(\underline{x}) - \tau(\theta))^2]$$

$$\begin{aligned} E_\theta [(T(\underline{x}) - \tau(\theta))^2] &= E_\theta \left[(T(\underline{x}) - E_\theta[T(\underline{x})] + E_\theta[T(\underline{x})] - \tau(\theta))^2 \right] \\ &= E_\theta \left[(T(\underline{x}) - E_\theta(T(\underline{x})))^2 + (E_\theta(T(\underline{x})) - \tau(\theta))^2 \right. \\ &\quad \left. + 2(T(\underline{x}) - E_\theta[T(\underline{x})])(E_\theta[T(\underline{x})] - \tau(\theta)) \right] \end{aligned}$$

$$\begin{aligned} &= E_\theta [(T(\underline{x}) - E_\theta[T(\underline{x})])^2] + (E_\theta[T(\underline{x})] - \tau(\theta))^2 \\ &\quad + 2E_\theta \left\{ (T(\underline{x}) - E_\theta[T(\underline{x})])(E_\theta[T(\underline{x})] - \tau(\theta)) \right\} \end{aligned}$$

$$\begin{aligned}
 &= \text{Var}_{\theta}(\tau(x)) + (E_{\theta}[\tau(x)] - \tau(\theta))^2 \\
 &\quad + 2(E_{\theta}[\tau(x)] - \tau(\theta)) \left\{ E_{\theta}[\tau(x)] - E_{\theta}[\tau(x)] \right\} \\
 &= \text{Var}_{\theta}(\tau(x)) + (E_{\theta}[\tau(x)] - \tau(\theta))^2
 \end{aligned}$$

$$E_{\theta}[\tau(x)] - \tau(\theta) = \text{Bias}_{\theta}(\tau(x))$$

$$\begin{aligned}
 \textcircled{\$} \quad E_{\theta}[(\tau(x) - \tau(\theta))^2] &= \text{MSE}_{\theta}(\tau(x)) \\
 &= \text{Var}_{\theta}(\tau(x)) + [\text{Bias}_{\theta}(\tau(x))]^2
 \end{aligned}$$

\oplus

Friday from 1:30 at BE 358.

Recap:

How to evaluate estimators.

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 MSE of $T(\underline{x})$

$$E_{\theta} [(T(\underline{x}) - \tau(\theta))^2] = \text{Var}_{\theta}(T(\underline{x})) + [\text{Bias}_{\theta}(T(\underline{x}))]^2$$

We will ideally like to have an estimator $T(\underline{x})$ of $\tau(\theta)$ s.t. if we take any other estimator $T_1(\underline{x})$ then $\text{MSE}_{\theta}(T(\underline{x})) \leq \text{MSE}_{\theta}(T_1(\underline{x})) \neq 0$.

$T_1(\underline{x}) = 10$ if $\tau(\theta) = 10$ then this estimator has MSE 0 at that point of θ .

$$\text{Let } C_{\tau} = \left\{ T(\underline{x}) : E_{\theta}(T(\underline{x})) = \tau(\theta) \right\}$$

Note that for any estimator $T(\underline{x}) \in C_{\tau}$ one has $E_{\theta}(T(\underline{x})) = \tau(\theta) \Rightarrow \text{Bias}_{\theta}(T(\underline{x})) = 0$

thus this class of estimators is called the class of all unbiased estimators of $\tau(\theta)$.

$$\text{Note that } \text{MSE}_{\theta}(T(\underline{x})) = \text{Var}_{\theta}(T(\underline{x})) + [\text{Bias}_{\theta}(T(\underline{x}))]^2$$

thus for any $T(\underline{x}) \in C_{\tau}$

$$\text{MSE}_{\theta}(T(\underline{x})) = \text{Var}_{\theta}(T(\underline{x}))$$

~~Thus~~ We would like to find an estimator $T(\bar{x}) \in C_c$ s.t.

$$MSE_0(T(\bar{x})) \leq MSE_0(W(\bar{x})) \quad \text{for } \forall \theta$$

where $W(\bar{x})$ is any other estimator in C_c .

$$\Leftrightarrow \text{Var}_\theta(T(\bar{x})) \leq \text{Var}_\theta(W(\bar{x})) + \delta.$$

It has been found that in the restricted class C_c , there exists an estimator $T(\bar{x})$ which satisfies this property. This estimator is called the uniformly minimum variance unbiased estimator (UMVUE) of $\tau(\theta)$.

Theorem: If $T(\bar{x})$ is the ~~the~~ UMVUE of $\tau(\theta)$, then $T(\bar{x})$ is unique.

Pf: Suppose the statement is not true and $W(\bar{x})$ be another UMVUE of $\tau(\theta)$.

$$\text{Define } T^*(\bar{x}) = \frac{W(\bar{x}) + T(\bar{x})}{2}$$

$$\text{Since } E_\theta[W(\bar{x})] = \tau(\theta) = E_\theta[T(\bar{x})]$$

$$\Rightarrow E_\theta[T^*(\bar{x})] = \tau(\theta)$$

$$\text{Var}_\theta(T^*(\bar{x})) = \text{Var}_\theta\left(\frac{T(\bar{x}) + W(\bar{x})}{2}\right)$$

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

$$\leq \frac{1}{4} \text{Var}_\theta(T(\bar{x})) + \frac{1}{4} \text{Var}_\theta(W(\bar{x})) + \frac{1}{2} \sqrt{\text{Var}_\theta(T(\bar{x})) \text{Var}_\theta(W(\bar{x}))}$$

(by Cauchy-Schwarz inequality)

Since both $T(\underline{x})$ and $W(\underline{x})$ are UMVUE

$$\text{Var}_{\theta} (T(\underline{x})) = \text{Var}_{\theta} (W(\underline{x})) \neq 0.$$

$$\begin{aligned}\text{Var}_{\theta} (T^*(\underline{x})) &\leq \frac{1}{4} \text{Var}_{\theta} (T(\underline{x})) + \frac{1}{4} \text{Var}_{\theta} (T(\underline{x})) \\ &\quad + \frac{1}{2} \sqrt{\text{Var}_{\theta} (T(\underline{x})) \text{Var}_{\theta} (T(\underline{x}))} \\ &= \text{Var}_{\theta} (T(\underline{x}))\end{aligned}$$

If this inequality is strict at least at one θ , then it is clearly a contradiction.

Thus the inequality can't be strict at any point θ . Hence the Cauchy-Schwarz inequality has to be an equality in this case.

Cauchy-Schwarz inequality is an equality if and only if $W(\underline{x}) = a(\theta) T(\underline{x}) + b(\theta)$

$$\text{We all know, } \text{Cov}_{\theta} (T(\underline{x}), W(\underline{x})) = \text{Var}_{\theta} (T(\underline{x}))$$

$$\begin{aligned}\text{Cov}_{\theta} (T(\underline{x}), W(\underline{x})) &= \text{Cov}_{\theta} (T(\underline{x}), a(\theta) T(\underline{x}) + b(\theta)) \\ &= a(\theta) \text{Var}_{\theta} (T(\underline{x}))\end{aligned}$$

$$\Rightarrow a(\theta) \text{Var}_{\theta} (T(\underline{x})) = \text{Var}_{\theta} (T(\underline{x}))$$

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

$$\text{thus } W(\underline{x}) = T(\underline{x}) + b(\theta)$$

$$\text{Now, note that } E_{\theta} [W(\underline{x})] = E_{\theta} [T(\underline{x})] = C(\theta)$$

$$\Rightarrow b(\theta) = 0 \quad \Rightarrow \quad W(\underline{x}) = T(\underline{x})$$

Thus UMVUE is unique.

Theorem: $\textcircled{1}$ $W(\bar{x})$ is the UMVUE of $\tau(\theta)$ if and only if $W(\bar{x})$ is uncorrelated with all unbiased estimators of θ .

$\textcircled{1}$ Note that $W(\bar{x})$ is an unbiased estimator of θ if $E_\theta[W(\bar{x})] = \theta$.

Suppose you start with some statistic and found that it is correlated with some unbiased estimator of θ , then certainly your statistic is not the UMVUE.

Example: $X \sim U(\theta, \theta+1)$,

$$\textcircled{2} \quad W(x) = x - \frac{1}{2} \quad E_\theta[W(x)] = \theta$$

i.e. $W(x)$ is unbiased.

Let $H(x)$ be an unbiased estimator of θ .

$$E_\theta[H(x)] = \theta \Rightarrow \int_0^{\theta+1} H(x) dx = \theta$$

$$\Rightarrow \frac{d}{d\theta} \int_0^{\theta+1} H(x) dx = 0 \Rightarrow H(\theta+1) - H(\theta) = 0$$

Let $H(x) = \textcircled{3} \sin(2\pi x)$. Then this is an unbiased estimator of θ .

$$\text{cov}_\theta \left(x - \frac{1}{2}, \sin(2\pi x) \right) = -\frac{\cos(2\pi\theta)}{2\pi} \neq 0 \text{ for all } \theta.$$

thus $x - \frac{1}{2}$ is not uncorrelated with $\sin(2\pi x)$.

Hence $\bar{x} - \frac{1}{2}$, although an unbiased estimator of θ , is not the UMVUE.

Rao-Blackwell Theorem:

Let $W(\underline{x})$ be any unbiased estimator of $T(\theta)$. Let $T(\underline{x})$ be a sufficient statistic for $T(\theta)$. Define

$$\textcircled{*} \quad \phi(T(\underline{x})) = E[W(\underline{x}) | T(\underline{x})]. \quad \text{Then}$$

- (i) $\phi(T(\underline{x}))$ is an unbiased estimator of $T(\theta)$
- (ii) $\text{Var}_\theta(\phi(T(\underline{x}))) \leq \text{Var}_\theta(W(\underline{x}))$, with equality holding if and only if $\phi(T(\underline{x})) = W(\underline{x})$ w.p. 1.

Pf:- Since $\underline{x} | T(\underline{x})$ is free of θ

$\Rightarrow W(\underline{x}) | T(\underline{x})$ is free of θ

$\Rightarrow E_\theta[W(\underline{x}) | T(\underline{x})]$ is free of θ

Hence $\phi(T(\underline{x})) = E_\theta[W(\underline{x}) | T(\underline{x})]$ is a statistic.

$$(i) \quad E_\theta[\phi(T(\underline{x}))] = E_\theta\{E_\theta[\textcircled{*} W(\underline{x}) | T(\underline{x})]\}$$

$$= E_\theta[W(\underline{x})] = T(\theta)$$

$\Rightarrow \phi(T(\underline{x}))$ is an unbiased estimator of $T(\theta)$.

$$\begin{aligned}
 \text{(ii)} \quad \text{Var}_\theta(W(\underline{x})) &= E_\theta[\text{Var}_\theta(W(\underline{x}) | T(\underline{x}))] \\
 &\quad + \text{Var}_\theta(E_\theta[W(\underline{x}) | T(\underline{x})]) \\
 &= E_\theta[\text{Var}_\theta(W(\underline{x}) | T(\underline{x}))] + \text{Var}_\theta(\phi(T(\underline{x}))) \\
 &\geq \text{Var}_\theta(\phi(T(\underline{x})))
 \end{aligned}$$

Here equality holds if and only if

$$\phi(T(\underline{x})) = W(\underline{x}) \text{ w.p. 1.}$$

Example: $X_1, X_2, X_3 \stackrel{\text{iid}}{\sim} \text{Ber}(p)$.

$$W(\underline{x}) = \frac{X_1 + X_2}{2} \Rightarrow E_p[W(\underline{x})] = p$$

Thus $W(\underline{x})$ is an unbiased estimator of p .

$T(\underline{x}) = X_1 + X_2 + X_3$ is a sufficient statistic.

for p .

$$\begin{aligned}
 \phi(T(\underline{x})) &= \cancel{E}[W(\underline{x})] E_p[W(\underline{x}) | T(\underline{x})] \\
 &= E\left[\frac{X_1 + X_2}{2} \mid X_1 + X_2 + X_3\right]
 \end{aligned}$$

$$\phi(t) = E\left[\frac{X_1 + X_2}{2} \mid X_1 + X_2 + X_3 = t\right]$$

$$= \frac{1}{2} E[X_1 \mid X_1 + X_2 + X_3 = t] + \frac{1}{2} E[X_2 \mid X_1 + X_2 + X_3 = t]$$

... (*)

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

$$\Rightarrow = \frac{1}{2} \frac{t}{3} + \frac{1}{2} \frac{t}{3}$$

$$\phi(t) = \frac{t}{3} \Rightarrow \phi(\tau(\underline{x})) = \frac{\underline{x_1+x_2+x_3}}{3}$$

Note that we started with $w(\underline{x}) = \frac{\underline{x_1+x_2}}{2}$

$$\text{Var}_p(w(\underline{x})) = p\frac{(1-p)}{4} + p\frac{(1-p)}{4} = p\frac{(1-p)}{2}$$

$$\begin{aligned}\text{Var}_p(\phi(\tau(\underline{x}))) &= \text{Var}_p\left(\frac{\underline{x_1+x_2+x_3}}{3}\right) \\ &= p\frac{(1-p)}{9} + p\frac{(1-p)}{9} + p\frac{(1-p)}{9} = p\frac{(1-p)}{3}\end{aligned}$$

thus $\phi(\tau(\underline{x}))$ has less variance than $w(\underline{x})$.

Recap:

How to find UMVUE.

Rao - Blackwell theorem:

Suppose $W(\underline{x})$ is any unbiased estimator of $\tau(\theta)$.
 $T(\underline{x})$ be a sufficient statistic. Then

$\phi(T(\underline{x})) = E[W(\underline{x}) | T(\underline{x})]$ will be unbiased for $\tau(\theta)$
and $\text{Var}(\phi(T(\underline{x}))) \leq \text{Var}(W(\underline{x}))$.

Given any unbiased estimator of $\tau(\theta)$, Rao - Blackwell theorem is a way to construct another ^{unbiased} estimator which has a lower variance than the earlier estimator. How much conditioning is needed to reach UMVUE.

Theorem (Lehman - Scheffe):

Suppose $T(\underline{x})$ is a complete sufficient statistic and there exists a function $\phi(T(\underline{x}))$ of $T(\underline{x})$ such that $E[\phi(T(\underline{x}))] = \psi(\theta)$. Then $\phi(T(\underline{x}))$ is the UMVUE for $\psi(\theta)$.

If:- If $T_1(\underline{x})$ is any other unbiased estimator of $\psi(\theta)$. Consider $\phi_1(T(\underline{x})) = E[T_1(\underline{x}) | T(\underline{x})]$, this is a statistic (since $T(\underline{x})$ is a sufficient statistic) and By Rao - Blackwell theorem $\text{Var}(\phi_1(T(\underline{x}))) \leq \text{Var}(T_1(\underline{x}))$

$$E[\phi_1(T(\underline{x}))] = E[\phi(T(\underline{x}))] = \psi(\theta)$$

$$\Rightarrow E[\phi_1(T(\underline{x})) - \phi(T(\underline{x}))] = 0$$

Since $T(\underline{x})$ is complete statistic

$$\Rightarrow \phi_1(T(\underline{x})) = \phi(T(\underline{x})) \text{ w.p. 1.}$$

~~Step~~ Algorithm 1:

- ① First find a complete sufficient statistic from a class of distributions. Let it be $T(\underline{x})$. Use our earlier results to do this.
- ② Try to find a statistic $\phi(T(\underline{x}))$ s.t.

$$E[\phi(T(\underline{x}))] = \psi(\theta).$$

When ~~this~~ is found the function ϕ is not immediate, you will take the following step.

- i) find any unbiased estimator of $\psi(\theta)$.
- ii) let that unbiased estimator be $W(\underline{x})$.
- iii) Then calculate $E[W(\underline{x}) | T(\underline{x})]$

Now $E[W(\underline{x}) | T(\underline{x})]$ is unbiased for $\psi(\theta)$ and it is a function of $T(\underline{x})$. Thus it is going to be the UMVUE.

example: $x_1, \dots, x_n \stackrel{iid}{\sim} \text{Ber}(p)$. What is the ~~@~~ UMVUE for p ?

Here complete ~~(ii)~~

$$f_p(x) = \left(\frac{p}{1-p}\right)^{\sum_{i=1}^n x_i} (1-p)^n = (1-p)^n \exp\left\{\sum_{i=1}^n x_i \log \frac{p}{1-p}\right\}$$

$\sum_{i=1}^n x_i$ is a complete sufficient statistic by the result of exponential family.

thus $E\left[\frac{\sum_{i=1}^n x_i}{n}\right] = p$. Now $\frac{\sum_{i=1}^n x_i}{n}$ is a function of ~~the~~ a complete sufficient statistic and is unbiased for p . Hence it is the UMVUE for p .

Find UMVUE for p^2 ?

~~We want to find a function of~~ $\sum_{i=1}^n x_i$ that is unbiased for p^2 .

$$\tau(x) = \sum_{i=1}^n x_i \sim \text{Bin}(n, p).$$

$$\begin{aligned} E[\tau(x)(\tau(x)-1)] &= E[\tau(x)^2] - E[\tau(x)] \\ &= \text{Var}(\tau(x)) + \{E[\tau(x)]\}^2 - E[\tau(x)] \\ &= np(1-p) + (np)^2 - np = np - np^2 + n^2p^2 - np \\ &= n(n-1)p^2 \end{aligned}$$

$$\Rightarrow E\left[\frac{\tau(x)(\tau(x)-1)}{n(n-1)}\right] = p^2$$

So we found an unbiased estimator of p^2

③

that is a function of $T(\underline{x})$. Thus By Lehman-Scheffe theorem $\frac{T(\underline{x})(T(\underline{x})-1)}{n(n-1)}$ is the UMVUE for p^2 .

Another way using Rao-Blackwell:

$$W(\underline{x}) = I(x_1=1, x_2=1) \quad (\text{if } x_1=0 \text{ or } x_2=0)$$

$$= 1 \quad \text{if } x_1=1, x_2=1$$

$$E[W(\underline{x})] = 0 \cdot P(x_1=1, x_2=1) + 0 \cdot P(\text{one of } x_1 \text{ and } x_2 \neq 1)$$

$$= P(x_1=1) P(x_2=1) = p^n$$

Find $E[W(\underline{x}) \mid \sum_{i=1}^n x_i = t]$

$$E[W(\underline{x}) \mid \sum_{i=1}^n x_i = t] = 1 \cdot P(x_1=1, x_2=1 \mid \sum_{i=1}^n x_i = t)$$

$$+ 0 \cdot P(\text{one of } x_1 \text{ and } x_2 \neq 1 \mid \sum_{i=1}^n x_i = t)$$

$$= P(x_1=1, x_2=1 \mid \sum_{i=1}^n x_i = t)$$

$$= \frac{P(x_1=1, x_2=1, \sum_{i=1}^n x_i = t)}{P(\sum_{i=1}^n x_i = t)}$$

$$= \frac{P(x_1=1, x_2=1, \sum_{i=3}^n x_i = t-2)}{P(\sum_{i=1}^n x_i = t)}$$

$$\begin{aligned}
 &= \frac{P(X_1=1) P(X_2=1) P\left(\sum_{i=3}^n X_i = t-2\right)}{P\left(\sum_{i=1}^n X_i = t\right)} \\
 &= \frac{\cancel{p} \times \cancel{p} \times \binom{n-2}{t-2} \cancel{p^{t-2}} \cancel{(1-p)^{n-t}}}{\binom{n}{t} \cancel{p^t} \cancel{(1-p)^{n-t}}} \\
 &= \frac{\binom{n-2}{t-2}}{\binom{n}{t}} = \frac{(n-2)!}{(t-2)!(n-t)!} \times \frac{t!(n-t)!}{n!} \\
 &= \frac{t(t-1)}{n(n-1)} \quad \begin{bmatrix} t! = t(t-1)(t-2)! \\ n! = n(n-1)(n-2)! \end{bmatrix}
 \end{aligned}$$

$$E[W(\underline{x}) | T(\underline{x})] = \frac{T(\underline{x})(T(\underline{x})-1)}{n(n-1)}$$

Example: Consider $X_1, \dots, X_n \stackrel{iid}{\sim} N(\mu, \sigma^2)$.
 $(\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2)$ is a complete sufficient statistic.

What is the UMVUE of μ .

Now $E\left[\frac{\sum_{i=1}^n X_i}{n}\right] = \mu$ thus \bar{X} is ~~a~~ the UMVUE for μ .

What is the UMVUE for $\mu^2 + \sigma^2$?

$E\left[\frac{\sum_{i=1}^n X_i^2}{n}\right] = \mu^2 + \sigma^2$ $\frac{\sum_{i=1}^n X_i^2}{n}$ is a function of ~~a~~ a complete sufficient statistic. Hence it is the UMVUE for $\mu^2 + \sigma^2$.

Example: $x_1, \dots, x_n \stackrel{iid}{\sim} \text{Pois}(\lambda)$. What is the UMVUE of $P(x_1=0)$?

$$P(x_1=0) = e^{-\lambda}$$

By the exponential family result $\sum_{i=1}^n x_i$ is a complete sufficient statistic.

$$W(\underline{x}) = I(x_1=0)$$

$$E[W(\underline{x})] = P(x_1=0)$$

$E[W(\underline{x}) | \sum_{i=1}^n x_i]$: we need to find

$$\begin{aligned} E[W(\underline{x}) | \sum_{i=1}^n x_i = t] &= P(x_1=0 | \sum_{i=1}^n x_i = t) \\ &= \frac{P(x_1=0, \sum_{i=1}^n x_i = t)}{P(\sum_{i=1}^n x_i = t)} \end{aligned}$$

What is the distribution of $\sum_{i=1}^n x_i$?

$$\sum_{i=1}^n x_i \sim \text{Pois}(n\lambda)$$

$$= \frac{P(x_1=0, \sum_{i=2}^n x_i = t)}{P(\sum_{i=1}^n x_i = t)}$$

$$= \frac{P(x_1=0) P(\sum_{i=2}^n x_i = t)}{P(\sum_{i=1}^n x_i = t)}$$

$$= \frac{(n-1)^t}{n^t}$$

$$= \frac{\cancel{e^{-n\lambda}} \frac{e^{-(n-1)\lambda} [(n-1)\lambda]^t}{t!}}{\cancel{e^{-n\lambda}} \frac{(n\lambda)^t}{t!}}$$

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

Thus ~~$E[W(\underline{x}) | \mathcal{F}(\underline{x})]$~~ ~~$E(W)$~~ ~~$\sum_{i=1}^n x_i$~~

$$E[W(\underline{x}) | \sum_{i=1}^n x_i] = \left(1 - \frac{1}{n}\right)^{\sum_{i=1}^n x_i}$$

This is the UMVUE for $P(X_1=0) = e^{-\lambda}$.