Example:

$X_1, \ldots, X_n$ random sample from $B(1, \theta)$

$0 < \theta < 1$

$g(\theta) = \theta$

$T(\underline{X}) = \sum_1^n X_i$ is suff (also minimal suff)

$\delta(\underline{X}) = X_1$ an unbiased estimator for θ

$$\eta(T) = E(X_1 \mid T)$$

$$= 0 \; P(X_1 = 0 \mid T) + 1 \cdot P(X_1 = 1 \mid T)$$

$$= P(X_1 = 1 \mid T)$$

$$= \frac{P(X_1 = 1, \sum_1^n X_i = t)}{P(\sum_1^n X_i = t)}$$

$$= \frac{P(X_1 = 1, \sum_2^n X_i = t-1)}{P(\sum_1^n X_i = t)}$$

$$= \frac{P(X_1=1)\, P\left(\sum_2^n X_i = t-1\right)}{P\left(\sum_1^n X_i = t\right)} \qquad \left(X_1, \dots X_n \text{ are indep}\right)$$

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

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

$$\Rightarrow \quad \eta(T) = E\left(\delta(\underline{X}) \mid T\right) = \frac{T}{n}$$

By Rao-Blackwell th^{m} $V(\eta(T)) < V(\delta)$

<u>Remark</u>: Start with an other unbiased estimator $\tilde{\delta}(\underline{X}) \to \eta(T)$ would be the same !!

This however may not happen for other suff stat, say, $(X_1, \dots, X_n)$ as suff statistic.

<u>Example</u>: $X_1, \dots, X_n$ i.i.d. random sample from $P(\theta)$, $\theta > 0$

$g(\theta) = e^{-\theta} \; \left(= P(X_1 = 0)\right)$

$T(\underline{X}) = \sum_1^n X_i$ is sufficient for θ (also minimal suff)

Let $\delta(\underline{X}) = \begin{cases} 1, & X_1 = 0 \\ 0, & \text{o/w} \end{cases}$ $\qquad E\,\delta(\underline{X}) = P(X_1 = 0) = e^{-\theta}$

$\Rightarrow \delta(\underline{X})$ is unbiased estimator for $e^{-\theta}$

$$\eta(T) = E\left(\delta(\underline{X}) \mid T\right)$$

$$= P(X_1 = 0 \mid T)$$

$$= \frac{P(X_1 = 0, T = t)}{P(T = t)}$$

$$= \frac{P(X_1 = 0, \sum_2^n X_i = t)}{P(\sum_1^n X_i = t)} \qquad \left(T \sim P(n\theta)\right)$$

$$\sum_2^n X_i \sim P((n-1)\theta)$$

$$= \frac{P(X_1 = 0)\, P(\sum_2^n X_i = t)}{P(\sum_1^n X_i = t)}$$

$$= \frac{(e^{-\theta})\left(e^{-(n-1)\theta}((n-1)\theta)^t / t!\right)}{\frac{e^{-n\theta}(n\theta)^t}{t!}}$$

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

$$\Rightarrow \eta(T) = \left(\frac{n-1}{n}\right)^T.$$

By Rao-Blackwell th^m $V(\eta(T)) < V(\delta)$

<u>Remark</u>: Once again if we start with any other starting unbiased estimator δ and use $T = \sum X_i$, we would get same $\eta(T) = \left(\frac{n-1}{n}\right)^T$

<u>Remark</u>: An additional property of sufficient statistic that ensures existence of unique unbiased estimator based on sufficient statistic, which has minimum variance among all other unbiased estimators, is "Completeness".

Complete Statistic

Def^n: A statistic T is said to be complete if for any real valued f^n g

$$E\, g(T) = 0 \quad \forall\, \theta \in \Theta$$

$$\Rightarrow \; g(t) = 0 \quad \text{with probability } 1$$

(i.e. almost everywhere)

Remark: Completeness of sufficient statistic is important if we are trying to find the "best" unbiased estimator, i.e. the UMVUE.

Let $g(\theta)$ be an estimand and T be complete sufficient statistic.

Suppose $\exists$ an unbiased estimator of $g(\theta)$. then $g(\theta)$ has one and only one unbiased estimator that is a f^n of T.

If $\delta(\underline{X})$ is u.e. of $g(\theta)$, then

$$\eta(T) = E(\delta(\underline{X}) \mid T) \text{ is also u.e. (Rao-Blackwell)}$$

$\eta(T)$ is an u.e. based on T

Let $\eta^*(T)$ be another u.e. of $g(\theta)$ based on T

$$E(\eta(T) - \eta^*(T)) = 0 \quad \forall\, \theta \in \Theta$$

As T is complete, the above implies that

$$\eta(T) = \eta^*(T) \quad \text{with prob } 1$$

i.e. essentially $\exists$ one u.e. of $g(\theta)$ based on T which has the lowest variance among all u.e.s

Remark: If T is thus complete sufficient, unbiased estimator of $g(\theta)$ based on T would be the unique UMVUE – the "best" unbiased estimator.

Remark: If T is complete sufficient

$g(\theta)$: estimand

$\exists$ an u.e. of $g(\theta)$

unique UMVUE for $g(\theta)$ is

$$\eta(T) = E(\delta(X) \mid T)$$

Approaches to prove completeness of suff statistic

(I) s-parameter exponential family argument

Def: X has a dist^n of s-parameter exponential family if it's p.d.f. or p.m.f. is of the form

$$f(x) = h(x) \exp\left(\sum_{i=1}^{s} \eta_i(\theta)\, T_i(x) - \beta(\theta) \right)$$

or $$f(x) = h(x) \exp\left(\sum_{i=1}^{s} \eta_i\, T_i(x) - A(\eta) \right)$$

in the reparametrized form (in terms of η parametrization)

$\{\underset{\sim}{\eta} : \underset{\sim}{\theta} \in \Theta\}$ – is called the natural parameter space