7.3.3 Sufficiency and Unbiasedness

We know from chapter 4

$$\mathbb{E}[X] = \mathbb{E}_Y\left[\mathbb{E}_{X|Y}[X|Y]\right]$$

Law of iterated expectation

$$\mathrm{Var}(X) = \mathrm{Var}\left(\mathbb{E}[X|Y]\right) + \mathbb{E}\left[\mathrm{Var}(X|Y)\right]$$

Law of total variance

Theorem 7.3.17 (Rao-Blackwell)

Let W be any unbiased estimator of $\tau(\theta)$, and let T be a sufficient statistic for θ. Define $\phi(T) = \mathbb{E}[W|T]$. Then $\mathbb{E}_\theta[\phi(T)] = \tau(\theta)$ and $\mathrm{Var}_\theta(\phi(T)) \leq \mathrm{Var}_\theta(W)$ for all θ; that is

$\phi(T)$ is a uniformly better unbiased estimator of $\tau(\theta)$

Proof:

W is a unbaised estimator of $\tau(\theta)$

$$\mathbb{E}[W] = \tau(\theta)$$

we know that

$$\tau(\theta) = \mathbb{E}[W] = \mathbb{E}\Big[\mathbb{E}[W|T]\Big]$$

$$= \mathbb{E}[\phi(T)]$$

$\Rightarrow$ $\phi(T)$ is a unbaised Estimator of $\tau(\theta)$

Also

$$\text{var}(W) = \text{Var}\left(\mathbb{E}[W|T]\right) + \mathbb{E}\Big[\text{var}(W|T)\Big]$$

$$= \text{var}(\phi(T)) + \underbrace{\mathbb{E}\Big[\text{var}(W|T)\Big]}_{\geq 0}$$

$\Rightarrow \quad Var(\omega) \geq Var(\phi(T))$

$\Rightarrow$ $\phi(T)$ is uniformly better than ω

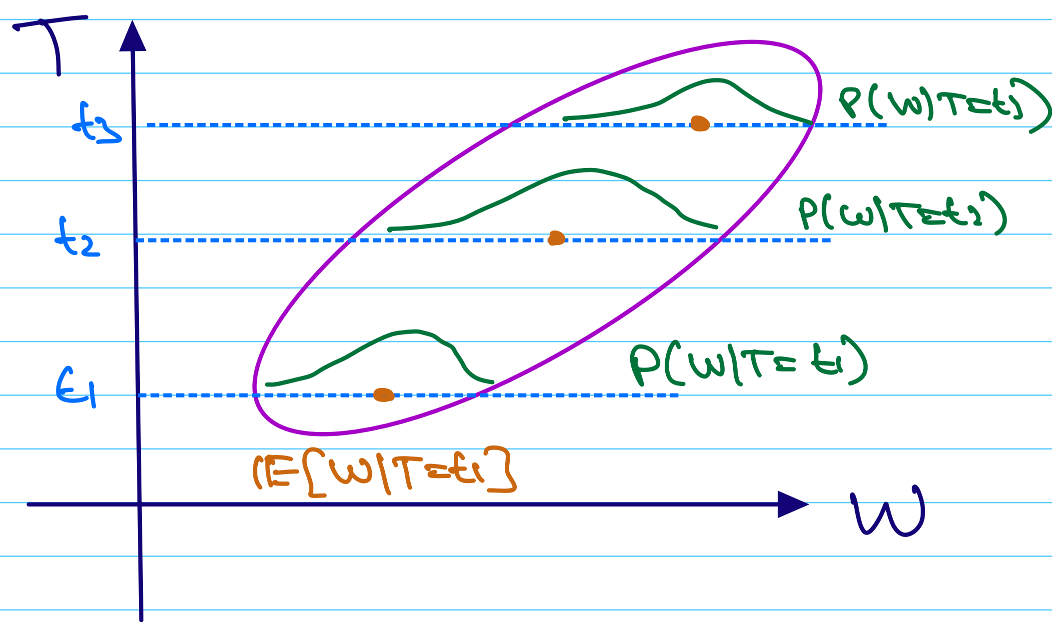
$\Rightarrow$ $\phi(T) = \mathbb{E}[W|T]$, because T is a sufficient statistic, then $\mathbb{E}[W|T]$ does not depend on θ

$\Rightarrow$ $\phi(T)$ is indeed function of samples not θ

$\Rightarrow$ $\phi(T)$ is an sufficient statistic, unbaised, uniformly better Estimator compared to W

$\Rightarrow$ $\phi(T)$ is some function of sufficient statistic T.

$\Rightarrow$ so we need to consider only statistics that are function's of a sufficient statistic in our search for best unbiased estimator's.



$\mathbb{E}[W|T]$ is a random variable

we are taking into account of only the Expected value at $T=t$, and abandoning the distribution at $T=t$ (i.e abandoning variance)

Key Point's (intution)

① W is not a sufficient statistic

① This mean's W carries extra,

unnecessary information apart from the information about the unknown parameter θ

(ii) it does not use all the information about θ, in the random sample, making it an inefficient estimator

② W is an unbiased Estimator

(i) It known's the direction to get unknown parameter θ
But it has high variance (unnecessary data) and not using all the info about θ in random sample

③ T is a sufficient statistic But unbiased.

(i) It captures all the information about θ, However it is biased, it doesnot know How to calculate θ

④ Intution:

W "know's the direction" but "cannot work" (ie has high variance)

T "can work" but "does not know the direction" (i.e Biased)

Solution: Use W's direction to correct T, leading to a new estimator that is unbaised and efficient.

intutively, if we have an unbiased estimator (which gives the correct answer on average but might have high variability), conditioning it on a sufficient statistic (which contain's all relevant information about the parameter) will make it more stable (reducing variability) without introducing Bias.

Since this conditioning always improve the estimator, we can focus only on statistics that are function's of sufficient statistics when searching for the best unbaised estimator's. This narrow's down our search, as any other statistic won't improve the estimator further.

Example 7.3.18 (Conditioning on a insufficient Statistic)

$$X_1, X_2 \sim N(\theta, 1)$$

$$\bar{X} = \frac{1}{2}(X_1 + X_2) \Rightarrow \mathbb{E}[\bar{X}] = \theta, \ \operatorname{var} \bar{X} = \frac{1}{2}$$

Consider conditioning on X_1, which is not sufficent

$$\phi(X_1) = \mathbb{E}[\bar{X} \mid X_1]$$

$$= \mathbb{E}\left[\frac{X_1 + X_2}{2} \mid X_1\right]$$

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

$$= \frac{X_1}{2} + \frac{1}{2}\mathbb{E}[X_2]$$

$$\phi(X_1) = \frac{X_1}{2} + \frac{\theta}{2}$$

$$\Rightarrow \phi(X_1) = \frac{X_1 + \theta}{2}$$

$\Rightarrow$ $\phi(X_1)$ is unbaised, $var(\phi(X_1)) \leq Var_\theta(\bar{X})$

$\Rightarrow$ $\phi(X_1)$ is better than $\bar{X}$. However

$\phi(X_1)$ is not an Estimator (dependent on θ)

in essence we can narrow our search by only considering estimator's based on sufficient statistics (function's of sufficient statistics) , as they are the most efficient.

<u>Theorem 7.3.19</u> :- if W is a best unbiased estimator of $\tau(\theta)$, then W is unique.

<u>Proof:</u>

suppose W' is another best unbiased estimator. Now consider

$$W^* = \frac{1}{2}(W + W')$$

$$\mathbb{E}_\theta[W^*] = \tau(\theta) \qquad \text{(unbiased)}$$

Now $\mathrm{Var}_\theta(W^*) = \mathrm{Var}_\theta\left(\frac{1}{2}(W + W')\right)$

$$\Rightarrow \quad Var(W^*) = \frac{1}{4} Var_\theta(W) + \frac{1}{4} Var(W') + \frac{1}{4} cov(W, W')$$

$$\Rightarrow \quad Var(W^*) \leq \frac{1}{4} Var(W) + \frac{1}{4} Var(W') \quad \frac{1}{2} Var(W)^{1/2} \, Var(W')^{1/2}$$

(Becaus both W, W' are best unbaised estimator's $\Rightarrow Var(W) = Var(W')$

$$\Rightarrow \quad \underbrace{Var(W^*) \leq Var(W)}$$

we cannot have strict inequility, because it contradicts the Best unbaised Estimator W, W'

$$\Rightarrow \quad Var(W^*) = Var(W)$$

$$\Rightarrow \quad cov(W, W')^2 = var(W)\, var(W')$$

$$\Rightarrow \quad W' = a(\theta) W + b(\theta)$$

$$cov(W, W') = cov(W, a(\theta)W + b(\theta))$$

$$var(W) = a(\theta)\, var(W)$$

$$\Rightarrow a(\theta) = 1 \,, \quad \mathbb{E}[W] = \tau(\theta)$$

$$\Rightarrow b(\theta) = 0$$

$$\Rightarrow \quad W = W' \quad (\text{unique})$$

Now $\quad \phi_a = W + aU$

U = unbiased Estimator of θ

W = unbiased Estimator of $\tau(\theta)$

$\Rightarrow \mathbb{E}[\phi_a] = \tau(\theta) \Rightarrow$ unbiased Estimator

$$Var(\phi_a) = Var(W) + a^2 Var(U) + 2a \cdot Cov(W, U)$$

$\Rightarrow$ for some $\theta = \theta_0$ if $Cov(W, U) < 0$ we can choose a such that

$$2a\,Cov(W, U) + a^2 Var_{\theta_0}(U) < 0$$

$$a \in \left(0, \frac{-2 Cov_{\theta_0}(W, U)}{Var\, U}\right)$$

$$\Rightarrow Var(\phi_a) < Var(W)$$

$$E[\phi_a] = \tau(\theta) \quad (\text{unbaised})$$

$\Rightarrow$ ϕ_a is beller than $W(a)$ at $\theta = \theta_0$

$\Rightarrow$ Thus, the relationship of W with unbaised estimator's of θ is crucial in evaluating whether W is best unbaised.

This relationship, in fact characterizes best unbiasedness

Theorem 7.3.20: if $E_\theta[W] = \tau(\theta)$, W is the Best unbiased estimator of $\tau(\theta)$ $\Longleftrightarrow$ W is uncorrelated with all unbiased estimator's of 0.

Proof:

if W is Best unbaised estimator of $\tau(\theta)$

$$\Downarrow$$

from above W is uncorrelated with all the Estimators of 0 (U)

$$Cov(W, U) = 0$$

Let W' be an estimator with

$$\mathbb{E}_\theta[W'] = \mathbb{E}_\theta[W] = \tau(\theta)$$

$$W' = W + (W' - W)$$

$$\mathrm{Var}(W') = \mathrm{Var}(W) + \mathrm{Var}(W' - W) + 2\,\mathrm{Cov}(W, W' - W)$$

$$= \mathrm{Var}(W) + \mathrm{Var}(W' - W) + 2\,\mathrm{Cov}\overbrace{(W, U)}^{0}$$

$$\Rightarrow \quad \mathrm{Var}(W') = \mathrm{Var}(W) + \mathrm{Var}(W' - W)$$

$$\Rightarrow \quad \mathrm{Var}(W') \geq \mathrm{Var}(W)$$

$\Rightarrow$ W is best unbiased Estimator of $\tau(\theta)$

An unbiased estimator of θ is essentially pure noise - it doesn't contain any meaningful information about the parameter θ. if you take an estimator W that is already unbiased for $\tau(\theta)$, adding a random noise U $(E_\theta[U]=0)$ should not

improve it. if adding some noise could reduce variance, it means W wasn't the best choice to begin with - there was room for improvement.

① Key idea: The best unbiased estimator should already be optimized in variance, meaning it should not benifit from adding or subtracting anything that doesn't carry information.

② if W is uncorrelated with all unbaised estimator's of 0, it mean's there's no leftover structure in W that could be tweaked using noise

③ if W were co-rrelated with some unbaised estimator's of 0, we could adjust it to get a lower variance Estimator - proving that W wasn't actually the best.

Thus, best unbaised estimator's must be uncorrelated with pure noise, because otherwise, they could still be improved, contradicting their optimality.

To find the Best unbaised Estimator's (BUE), we first need to understand

all unbaised estimator's of zero. if a distribution has the property that the only unbaised Estimator of 0 is itself, $U=0$, then any estimator we consider will automatically be un-correlated with zero $\operatorname{cov}(w, 0)=0$.

This simplifies our search, as it guarantees that any unbiased estimator will be optimal. in short, Completeness, ensures the uniqueness of the best unbaised estimator and end's the search.

Example 7.3.22 (Continuation of Example 7.3.12)

$X_1, X_2, \ldots X_n \sim \text{Uniform}(0, \theta)$

$Y = \max\{X_1, X_2, \ldots, X_n\}$

$\frac{n+1}{n} Y$ is a unbaised Estimator of θ.

we know that Y is a complete statistic

$\Rightarrow$ There are no unbaised Estimator's of θ that are based on Y.

$\Rightarrow \operatorname{cov}\left(\frac{n+1}{n} Y, U\right) = 0$

$\Rightarrow \frac{n+1}{n} Y$ is Best unbiased Estimator of θ.

Theorem 7.3.23: let T be complete sufficient statistic for a parameter θ, and let $\phi(T)$ be any Estimator Based only on T. Then $\phi(T)$ is the unique best unbiased estimator of its expected value.

Example 7.3.24 (Binomial Best unbiased Estimation)

$$X_1, X_2, \ldots X_n \sim \text{binomial}(k, \theta)$$

Estimate the Prob of exactly one success from a Binomial (k, θ), that is, estimate

$$\tau(\theta) = P_\theta(X=1) = k\theta(1-\theta)^{k-1}$$

Sol^n

$$X_i \sim \text{Binomial}(k, \theta)$$

k bernoulli trails with X_i sucess with Probability of θ

$$f_n(X \mid k, \theta) = \prod_{i=1}^{n} \binom{k}{x_i} \theta^{x_i} (1-\theta)^{k-x_i}$$

$$= \theta^{\sum x_i} (1-\theta)^{nk-\sum x_i} \prod_{i=1}^{n} \binom{k}{x_i}$$

$$\Rightarrow \underbrace{\left[\prod_{i=1}^{n} \binom{k}{x_i}\right]}_{h(x)} \cdot \underbrace{\left[\left(\frac{\theta}{1-\theta}\right)^{\sum x_i} \cdot (1-\theta)^{k}\right]}_{g(\sum x_i \mid \theta)}$$

$\Rightarrow$ $\sum x_i$ = sufficient statistic

$\sum x_i \sim$ binomial (nk, θ)

(complete sufficient static)

$\Rightarrow$ Now find a unbaised Estimator and apply Rao-Blackwell theorem

$$h(x_i) = \begin{cases} 1 & \text{if } x_i = 1 \\ 0 & \text{o.w} \end{cases}$$

$$E_\theta[h(x_1)] = \sum_{x_1=0}^{k} h(x_1) \binom{k}{x_1} \theta^{x_1} (1-\theta)^{k-x_1}$$

$$= \binom{k}{1} \theta^1 (1-\theta)^{k-1}$$

$$E_\theta[h(x_1)] = k\theta(1-\theta)^{k-1} \quad \text{unbiased.}$$

$\Rightarrow$ Apply Rao-Blackell

$$\phi\left(\sum_{i=1}^{n} x_i\right) = \mathbb{E}\left[h(x_1) \,\middle|\, \sum_{i=1}^{n} x_i\right]$$

$$\phi(t) = \mathbb{E}\left[h(x_1) \,\middle|\, \sum_{i=1}^{n} x_i = t\right]$$

$$\Rightarrow \quad \phi(t) = P\left(x_1 = 1 \,\middle|\, \sum_{i=1}^{n} x_i = t\right)$$

$$= \frac{\mathbb{P}_\theta\left(x_1 = 1, \sum_{i=1}^{n} x_i = t\right)}{P_\theta\left(\sum_{i=1}^{n} x_i = t\right)}$$

$$= \frac{\mathbb{P}_\theta(X_1 = 1)\, \mathbb{P}_\theta\left(\sum_{i=2}^{n} X_i = t-1\right)}{\mathbb{P}_\theta\left(\sum_{i=1}^{n} X_i = t\right)}$$

$$= \frac{\binom{k}{1} \theta^1 (1-\theta)^{k-1} \binom{(n-1)k}{t-1} \theta^{t-1} (1-\theta)^{(n-1)k-t+1}}{\binom{kn}{t} \theta^t (1-\theta)^{kn-t}}$$

$$= \frac{\binom{k(n-1)}{t-1}}{\binom{kn}{t}}$$

$$\Rightarrow \phi\left(\sum_{i=1}^{n} X_i\right) = k \frac{\binom{k(n-1)}{\sum X_i - 1}}{\binom{kn}{\sum X_i}}$$