Conditioning on an unsufficient statistic.

In example. $X_1, X_2 \sim N(\theta, 1)$, then for $\bar{X} = \frac{1}{2}(X_1 + X_2)$ we have : $E[\bar{X}] = \theta$, $Var[\bar{X}] = \frac{1}{2}$.

Let $\phi(X_1) = E[\bar{X} | X_1] = \frac{1}{2}\left[E[X_1|X_1] + E[X_2|X_1]\right] = \frac{X_1}{2} + \frac{\theta}{2}$.

$E[\phi(X_1)] = \theta$. $Var[\phi(X_1)] = \frac{1}{4} < Var[\bar{X}]$ but $\phi(X_1)$ is not a statistic.

Only estimators based on a sufficient statistics need to be considered when trying to find a best unbiased estimator.

Theorem 7.3.19

If W is a best unbiased estimator of $\tau(\theta)$, then W is unique

Proof. Assume W' is another best unbiased estimator and let $W^* = \frac{1}{2}(W + W')$. $E[W^*] = \tau(\theta)$.

Then
$$Var[W^*] = \frac{1}{4} Var[W] + \frac{1}{4} Var[W'] + \frac{1}{2} Cov(W, W')$$
$$\leq \frac{1}{4} Var[W] + \frac{1}{4} Var[W'] + \frac{1}{2}\left[Var[W] \, Var[W']\right]^{\frac{1}{2}}.$$
$$= Var[W]$$

Since W is best unbiased $Var[W^*] = Var[W] \Rightarrow W' = a(\theta) W + b(\theta)$

Also $\underbrace{Cov(W, W')}_{= Var[W]} = Cov(W, a(\theta)W + b(\theta)) = a(\theta) Var[W] \Rightarrow a(\theta) = 1$

and since $E[W'] = a(\theta)\tau(\theta) + b(\theta) = \tau(\theta) + b(\theta) = \tau(\theta)$, $b(\theta) = 0$.

Theorem 6.25

Let $X_1, \dots, X_n$ be iid with pdf/pmf from an exponential family of the form

$$f(x|\underline{\theta}) = h(x)\, c(\underline{\theta})\, e^{\sum_{j=1}^{k} w_j(\underline{\theta}) t_j(x)} \quad \text{where } \underline{\theta} = (\theta_1, \dots, \theta_k)$$

Then $\underline{T}(\underline{X}) = \left(\sum_{i=1}^{n} t_1(X_i), \dots, \sum_{i=1}^{n} t_k(X_i) \right)$ is complete if the parameterspace Θ contains an open set in $\underline{R}^k$

For example

$X_1, \dots, X_n$ iid $\sim N(\theta, \theta^2)$, $i = 1, 2, \dots, n$

$$f(\underline{x}|\underline{\theta}) = (2\pi)^{-\frac{n}{2}} \theta^{-n} e^{-\frac{1}{2\theta^2} \sum_{i=1}^{n} (x_i - \theta)^2}$$

$$= (2\pi)^{-\frac{n}{2}} \theta^{-n} \cdot e^{-\frac{n}{2}} \cdot e^{-\sum_{i=1}^{n} \frac{x_i^2}{2\theta^2} + \sum_{i=1}^{n} x_i \cdot \frac{1}{\theta}}$$

which is a density within the exponential family with $\underline{T}(\underline{x}) = \left(\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i^2 \right)$ and $(w_1(\theta), w_2(\theta)) = \left(\frac{1}{\theta}, -\frac{1}{2\theta^2} \right)$.

~~For example~~

~~$X_1, \dots, X_n$ iid $N(\theta, \theta^2)$, $i = 1, 2, \dots, n$~~

Minimal sufficient?

$$\frac{f(\underline{x}|\theta)}{f(\underline{y}|\theta)} = e^{-\frac{1}{2\theta^2}\left(\sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} y_i^2 \right) + \frac{1}{\theta}\left(\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} y_i \right)}$$

is a constant as a function of $\theta \iff \underline{T}(\underline{x}) = \underline{T}(\underline{y})$

Definition 6.2.21

Let $f(t|\theta)$ be a family of pdfs/pmfs for a statistic $T(\underline{X})$. The family of probability distributions is called complete if $E[g(T)] = 0, \ \forall\theta \Rightarrow P(g(T)=0) = 1$

Equivalently $T(\underline{X})$ is a complete statistic.

$E[g(T)] = 0, \ \forall\theta \Rightarrow P(g(T)=0) = 1$ — the only function $g(T)$ that satisfies $E[g(T)=0], \ \forall\theta$ is the function that is 0 with probability 1, $\forall\theta$.

Why the name

Suppose discrete distributions

$E[X] = \sum_{i=1}^{m} x_i \, P(X=x_i) = \langle \underline{x}, P(\underline{x}|\theta)\rangle$ or the scalar product

If $P(X=x_i|\theta) = v_i$, $i=1,2,\dots,m$ and if v_i, $i=1,2,\dots,m$ span the whole vector space, $\forall\theta$, the only vector, $\underline{x}$, that is orthogonal to all v_i's is the $\underline{0}$-vector.

Completeness assures that the collection of family distributions for all set of θ are sufficiently rich.

Example. T has a Binomial distribution.

$$E[g(T)] = \sum_{t=0}^{m} g(t)\binom{m}{t} p^t (1-p)^{m-t} = (1-p)^m \sum_{t=0}^{m} g(t)\binom{m}{t}\left(\frac{p}{1-p}\right)^t,$$

$$= (1-p)^m \sum_{t=0}^{m} g(t)\binom{m}{t} r(p)^t, \quad 0<p<1 \quad = 0, \ \forall p$$

$\Rightarrow g(t) = 0$, $t = 0, 1, \dots, m$, since the coefficient in the polynomial need to be zero for a polynomial to be zero for all r.

Complete?

$E[S^2] = \theta^2, \quad E[n\bar{x}^2] = n\left(\frac{\theta^2}{n} + \theta^2\right) = (n+1)\theta^2$

Let $g(\underline{T}) = (n+1)S^2 - n\bar{x}^2 \Rightarrow E[g(\underline{T})] = (n+1)\theta^2 - (n+1)\theta^2 = 0, \ \forall\theta$

$\Rightarrow \underline{T}(\underline{X})$ is not complete.

Notice $\frac{1}{2}[w_1(\theta)^2] + w_2(\theta) = 0$. There is a functional relationship between $w_1(\theta)$ and $w_2(\theta)$.

Rule. If $w_1(\theta), \dots, w_k(\theta)$ are linearly independent, then $\underline{T} = \underline{T}(\underline{X})$ is a minimal sufficient statistic

If $w_1(\theta), \dots, w_k(\theta)$ are functionally independent, than $\underline{T} = \underline{T}(\underline{X})$ is a complete sufficient statistics.

Theorem 7.3.23 Lehman-Scheffe.

Let T be a complete sufficient statistic for a parameter θ. If $\phi(T)$ is any estimator based on T, then $\phi(T)$ is the unique best unbiased estimator of $E[\phi(T)]$.

A practical variant of the Theorem.

Let T be a complete sufficient statistic for a parameter θ. If $h(T)$ is an unbiased estimator of $\tau(\theta)$, it is the best unbiased estimator of $\tau(\theta)$ and equal to the $E[W|T]$ for any unbiased estimator W (except on a set of measure 0).

Proof. Define $g(T) = E[W|T]$. Then $Var[g(T)] \leq Var[W]$

But $E[g(T) - h(T)] = \tau(\theta) - \tau(\theta) = 0, \ \forall \theta \Rightarrow g(T) = h(T)$

(except on a set of measure 0), since T is complete.

and $Var[h(T)] = Var[g(T)] \leq Var[W]$

Example.

$X_1, \ldots, X_m \sim$ Binomial (k, θ)

and let $\tau(\theta) = P(X=1) = k\theta(1-\theta)^{k-1}$

$T = \sum_{i=1}^{m} X_i \sim$ Binomial (km, θ) is a complete sufficient statistic for θ.

Let $W(X_1, \ldots, X_m) = \begin{cases} 1 & \text{if } X_1 = 1 \\ 0 & \text{otherwise.} \end{cases}$

$$E[W(X_1, \ldots, X_m)] = 1 \cdot \binom{k}{1}\theta(1-\theta)^{k-1} = k\theta(1-\theta)^{k-1}$$

Then $E[W|T]$ is the unique best unbiased estimator of $k\theta(1-\theta)^{k-1}$.

and $h(t) = E[W | \sum_{i=1}^{m} X_i = t] = 1 \cdot P(X_1 = 1 | \sum_{i=1}^{m} X_i = t)$

$$= \frac{P(X_1 = 1 \cap \sum_{i=1}^{m} X_i = t)}{P(\sum_{i=1}^{m} X_i = t)} = \frac{P(X_1 = 1 \cap \sum_{i=2}^{m} X_i = t-1)}{P(\sum_{i=1}^{m} X_i = t)}$$

$$= \frac{k\theta(1-\theta)^{k-1}\left[\binom{k(m-1)}{t-1}\theta^{t-1}(1-\theta)^{k(m-1)-(t-1)}\right]}{\binom{km}{t}\theta^{t}(1-\theta)^{km-t}} = \frac{k\binom{k(m-1)}{t-1}}{\binom{km}{t}}$$

$$h(T) = \phi\left(\sum_{i=1}^{m} X_i\right) = \frac{k\binom{k(m-1)}{\sum X_i - 1}}{\binom{km}{\sum X_i}}$$