MA 3140: Statistical Inference

Dr. Sameen Naqvi
Department of Mathematics, IIT Hyderabad
Email id: sameen@math.iith.ac.in

Interpretation of Completeness

Let $X_1, \dots, X_n$ be iid observations from $f(x;\theta)$, where $\theta \in \Theta$ is unknown.

Recall the following definitions:

- A statistic $V = V(\boldsymbol{X})$ is ancillary if its distribution does not depend on θ.
- A statistic $T = T(\boldsymbol{X})$ is sufficient if the conditional distribution of $\boldsymbol{X}$ given T does not depend on θ.
- A sufficient statistic T is complete if

$$E_\theta f(T) = 0,\ \forall\ \theta \Longrightarrow f(t) = 0\ (\forall\ \theta).$$

Interpretation of Completeness

- Also, a statistic $U = U(\boldsymbol{X})$ is an UE of zero if its expectation is zero, irrespective of θ.

- A statistic $V = V(\boldsymbol{X})$ is first-order ancillary if $E_\theta V$ is independent of θ.

  A first-order ancillary is simply an UE of zero plus a constant.

- **Example 1:** Let $X_1, \ldots, X_n$ be iid with density $f(x - \theta)$. Now, compare what happens when f is Normal, with the situation when f is Cauchy or Double Exponential.

Interpretation of Completeness

In the first case, the data can be reduced without loss of information to a small number of sufficient statistics; one for Normal ($\overline{X}$).

In the second case, the set of n OS is minimal sufficient.

The reason for this difference is related to the following property of the ancillary information:

In any location problem, the $n-1$ differences $Y_i = X_{(n)} - X_{(i)}$ $(i = 1, \ldots, n-1)$ constitute an $(n-1)$-dimensional ancillary statistic.

Interpretation of Completeness

In the normal case, the Y's are independent of the MSS and hence carry no information about θ.

On the other hand, when the n OS are minimal, the differences Y_i - although still ancillary - are functions of the MSS.

Thus, sufficiency has not been successful in squeezing out the ancillary material.

Note that the MSS is complete in the normal case, but not in the other case.

This suggests that completeness of a MSS T may be associated with the ability of T completely to rid itself of the ancillary part of the data by making it independent of T.

Interpretation of Completeness

- **Example 2:** Let $X_1, \ldots, X_n$ be iid observations from $U(\theta, \theta + 1)$.

  We have seen that $(X_{(1)}, X_{(n)})$ is sufficient but not complete (Why??).

  Reason: The difference $X_{(n)} - X_{(1)}$ is ancillary, and we can compute its expectation (say, a), i.e.,

  $$E[X_{(n)} - X_{(1)}] = a \Longrightarrow E[X_{(n)} - X_{(1)} - a] = 0.$$

  Thus, $X_{(n)} - X_{(1)} - a$ is an UE of zero which is not identically zero.

Interpretation of Completeness

It means that there exists functions of the sufficient statistic which are not informative about θ.

> This cannot happen in case of a CSS - it is in a sense maximally informative; no functions of it are uninformative.

Interpretation of Completeness

- Completeness also tells that there exists a unique UE of θ.

  Suppose you have 2 different UEs of θ, say $g_1(T)$, $g_2(T)$, based on the sufficient statistic T, i.e.,

  $$Eg_1(T) = Eg_2(T) = \theta$$

  and $P(g_1(T) \neq g_2(T)) > 0$ ($\forall\ \theta$).

  Then, $g_1(T) - g_2(T)$ is an UE of zero, which is not identically zero, proving that T is not complete.

  Thus, completeness of a sufficient statistic tells us that there does exist a unique UE of θ based on T.

Complete Statistics in the Exponential Family

Theorem: Let $X_1, \ldots, X_n$ be iid observations from an exponential family with pdf or pmf of the form

$$f_{\boldsymbol{\theta}}(x) = h(x)c(\boldsymbol{\theta})e^{\sum_{i=1}^{k} Q_i(\boldsymbol{\theta}) T_i(x)}, \quad \boldsymbol{\theta} \in \mathbb{R}^k.$$

Then the statistic

$$T(\boldsymbol{X}) = \left(\sum_{i=1}^{n} T_1(X_i), \ldots, \sum_{i=1}^{n} T_k(X_i) \right)$$

is complete if the parameter space Θ contains an open set in $\mathbb{R}^k$.

Why do we need the condition that the parameter space contain an open set ?

Consider $X \sim N(\theta, \theta^2)$, $\theta > 0$.

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

Here, $Q_1 = \frac{1}{\theta^2}$ and $Q_2 = \frac{1}{\theta}$.
However, the parameter space (θ, θ^2) does not contain a 2-dimensional open set as it consists of only the points on a parabola.

Applications of Basu's Theorem

Example 1: Let $X_1, \dots, X_n$ be iid exponential observations with parameter λ. Suppose you are interested in computing the expected value of

$$g(\boldsymbol{X}) = \frac{X_n}{X_1 + \cdots + X_n}.$$

Solution: First note that the exponential distributions form a scale parameter family and thus, $g(\boldsymbol{X})$ is an ancillary statistic. Second, the exponential distributions also form an exponential family with $T(x) = x$ and so, by previous theorem,

$$T(\boldsymbol{X}) = \sum_{i=1}^{n} X_i$$

is a complete statistic.

Applications of Basu's Theorem contd.

Third, it is easy to see that $T(\boldsymbol{X})$ is a sufficient statistic.

Hence, by Basu's Theorem, $T(\boldsymbol{X})$ and $g(\boldsymbol{X})$ are independent.

Thus, we have

$$\lambda = E_\lambda X_n = E_\lambda T(\boldsymbol{X})g(\boldsymbol{X}) = (E_\lambda T(\boldsymbol{X}))(E_\lambda g(\boldsymbol{X})) = n\lambda(E_\lambda g(\boldsymbol{X})).$$

Hence, for any λ, $E_\lambda g(\boldsymbol{X}) = 1/n$.

Applications of Basu's Theorem contd.

Example 2: Let $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$.

Case 1: $\sigma^2 = \sigma_0^2$ is known.

Define $V = \sum(X_i - \overline{X})^2$ and $T = \overline{X}$.

Recall that $\overline{X}$ is a sufficient statistic for μ. Also, using the previous theorem, it can be seen that the family of $N(\mu, \sigma^2/n)$ distributions, $-\infty < \mu < \infty$, σ^2/n known, is a complete family.

Further, note that V is ancillary for μ.

Thus, $\overline{X}$ and $\sum(X_i - \overline{X})^2$ are independent.

Applications of Basu's Theorem contd.

Case 2: μ and σ^2 are unknown.

We know that $T =???$ is CSS.

Let $V = \frac{\overline{X}-\mu}{S/\sqrt{n}}$.

Then, V is ancillary for σ^2.

Thus, $T =???$ and V are independently distributed.

Rao-Blackwell Theorem

Let W be an unbiased estimator of $g(\theta)$ and T be a sufficient statistic for θ.

Define $\phi(T) = E(W|T)$.

Then,

$$E_\theta \phi(T) = g(\theta)$$

and

$$Var_\theta \phi(T) \leq Var_\theta W, \ \forall \ \theta;$$

i.e., $\phi(T)$ is a uniformly better unbiased estimator of $g(\theta)$.

Therefore, conditioning any UE on a sufficient statistic will result in a uniform improvement.

Rao-Blackwell Theorem contd.

Proof: We have

$$g(\theta) = E_\theta W = E_\theta[E(W|T)] = E_\theta \phi(T) \quad \left(\because EX = E[E(X|Y)]\right)$$

Thus, $\phi(T)$ is unbiased for $g(\theta)$. Also,

$$\begin{aligned} Var_\theta W &= Var_\theta[E(W|T)] + E_\theta[Var(W|T)] \\ &= Var_\theta \phi(T) + E_\theta[Var(W|T)] \\ &\geq Var_\theta \phi(T). \qquad (Var(W|T) \geq 0) \end{aligned}$$

Hence $\phi(T)$ is uniformly better than W.

Lehmann Scheffe Theorem

T: a complete and sufficient statistic

δ: an UE of $g(\theta)$.

Then, there exists a unique UMVUE of $g(\theta)$ given by

$$\eta(T) = E[\delta | T].$$

Note: Here, it does not matter which UE δ is being conditioned; one can thus choose δ so as to make the calculation of conditional expectation as easy as possible.

Lehmann Scheffe Theorem contd.

Proof: Let δ be an unbiased estimator of $g(\theta)$. Then, by Rao-Blackwell Theorem, $\eta(T) = E[\delta|T]$ is such that

$$Var_\theta(\eta(T)) \leq Var_\theta(\delta), \ \forall \ \theta.$$

Let δ^* be another unbiased estimator and consider $\eta^*(T) = E[\delta^*|T]$. Then

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

and by completeness of T, it follows that

$$P_\theta(\eta(T) = \eta^*(T)) = 1, \ \forall \ \theta \in \Theta.$$

So, $\eta(T)$ is the unique UMVUE for $g(\theta)$.

Methods of finding UMVUE

(1.) Guess Method

(2.) Direct Method

(3.) Conditioning

Remark: All these three methods are built on Complete Sufficient Statistics and Unbiased Estimators.

Method 1: Guess Method

- Suppose we know that an UE δ of $g(\theta)$ is a function of a complete sufficient statistic T, then δ is the UMVUE.

Example: Normal UMVUE

Let $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$ and we are interested in estimating $g(\sigma) = \sigma^2$.

We know that UE of σ^2 is $\delta = \frac{1}{n-1} \sum (X_i - \overline{X})^2$ and CSS $T = (\sum X_i, \sum (X_i - \overline{X})^2)$.

Since δ is a function of T, we conclude that δ is UMVUE.

Method 2: Direct Method (Solving for δ)

- If T is a function of a complete sufficient statistic, the UMVUE of $g(\theta)$ is uniquely determined by

$$E_\theta \delta(T) = g(\theta), \quad \forall\ \theta \in \Omega.$$

Example 1: Binomial UMVUE

Let $X_1, \ldots, X_n \sim Bin(1, p)$ and we are interested in estimating $g(p) = p(1-p)$.

We know that $T = \sum X_i \sim Bin(n, p)$ is a CSS. Now,

Method 2: Direct Method (Solving for δ) cont'd

$$E[\delta(T)] = g(p), \ \forall \ 0 < p < 1$$

$$\sum_{t=0}^{n} \delta(t) \binom{n}{t} p^t (1-p)^{n-t} = p(1-p), \ \forall \ 0 < p < 1$$

$$\sum_{t=0}^{n} \delta(t) \binom{n}{t} \left(\frac{p}{1-p}\right)^t = p(1-p)^{-(n-1)}, \ \forall \ 0 < p < 1.$$

If $\frac{p}{1-p} = \rho$ so that $p = \frac{\rho}{1+\rho}$, the above equation can be written as

$$\sum_{t=0}^{n} \delta(t) \binom{n}{t} \rho^t = \rho(1+\rho)^{n-2}, \ \forall \ 0 < p < 1$$

$$= \rho \sum_{t=0}^{n-2} \binom{n-2}{t} \rho^t = \sum_{t=1}^{n-1} \binom{n-2}{t-1} \rho^t, \ \rho \in (0, \infty).$$

Method 2: Direct Method (Solving for δ) cont'd

A comparison of the coefficients on LHS and RHS yields

$$\delta(t) = \frac{t(n-t)}{n(n-1)}, \;\; t = 1, \ldots, n-1.$$

Hence, $\delta(T) = \frac{T(n-T)}{n(n-1)}$, $T = 1, \ldots, n-1$, is the UMVUE.

Example 2: Poisson UMVUE

Let $X_1, \ldots, X_n \sim P(\theta)$ and we are interested in estimating $g(\theta) = e^{-\theta}$.

We know that $T = \sum X_i \sim P(n\theta)$ is a CSS. Now,

Method 2: Direct Method (Solving for δ) cont'd

$$E[\delta(T)] = g(\theta)$$

$$\sum_{t=0}^{\infty} \delta(t) \frac{e^{-n\theta}(n\theta)^t}{t!} = e^{-\theta}$$

$$\sum_{t=0}^{\infty} \delta(t) \frac{(n\theta)^t}{t!} = e^{\theta(n-1)} = \sum_{t=0}^{\infty} \frac{[(n-1)\theta]^t}{t!}$$

Comparing coefficients of θ^t, we get

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

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

Hence, $\delta(T) = \left(1 - \frac{1}{n}\right)^T$ is UMVUE.

Method 2: Direct Method (Solving for δ) cont'd

Example 3: UMVUE for Uniform Distribution

Let $X_1, \ldots, X_n \sim U(0, \theta)$, $\theta > 0$ and $g(\theta) = \theta$.

We know that $T = X_{(n)}$ is a CSS.

Since $E(X_1) = \theta/2$, we find that $2X_1$ is an UE of θ.

Now,

$$E_\theta(T) = \int_0^\infty t \frac{nt^{n-1}}{\theta^n} dx = \frac{n}{n+1}\theta$$
$$\implies E\left(\frac{n}{n+1}T\right) = \theta$$

So, $\frac{n}{n+1}T$ is unbiased for θ, and hence, T is UMVUE for θ.

Thanks for your patience!