5

SUFFICIENCY AND COMPLETENESS

5.1 Sufficient Statistics

We have seen that point estimator of θ reduces to consideration of a statistic that reduces the data to a number (a function). In some sense, this estimator covers all the relevant sample information concerning θ. It is reasonable to restrict attention to such statistic. More generally, the idea of sufficiency involves the reduction of a data set to a set of statistics with no loss of information about the unknown parameter.

Definition 5.1.1. **(Sufficient statistics)**. Let $X_1, \ldots, X_n$ have joint PDF or PMF $f(x_1, \ldots, x_n; \theta)$, $\theta \in Ⓗ$ (may be k-dimension). Let $S_1 = A_1(X_1, \ldots, X_n)$, $S_2 = A_2(X_1, \ldots, X_n)$, …, $S_k = A_k(X_1, \ldots, X_n)$ be k-dimension statistic based on $X_1, \ldots, X_n$. We say $S_1, \ldots, S_k$ is a set of jointly sufficient statistics for θ if the conditional distribution of $(X_1, \ldots, X_n)$ given $S_1 = s_1, \ldots, S_k = s_k$ does not depend on θ. I.e. if

$$f_{X_1,\ldots,X_n|S_1,\ldots,S_k}(x_1, \ldots, x_n|s_1, \ldots, s_k)$$

is free of θ; note that $f_{X_1,\ldots,X_n|S_1,\ldots,S_k}(x_1, \ldots, x_n|s_1, \ldots, s_k) = \mathbb{P}\{X_1 = x_1, \ldots, X_n = x_n|S_1 = s_1, \ldots, S_k = s_k\}$ for discrete case.

The idea is that given $S_1 = s_1, \ldots, S_k = s_k$ are observed, values of $X_1, \ldots, X_n$ no longer give information about θ.

Example 5.1.1.. Let $X_1, \ldots, X_n$ be random sample from $\mathcal{B}in(1, p)$; $f(x; p) = p^x(1-p)^{1-x}, x \in \{0, 1\}$. Is $S = \sum_{i=1}^n X_i \sim \mathcal{B}in(n, p)$ a sufficient statistic? Let's check.

The joint PMF of $X_1, \dots, X_n$ is

$$f_{X_1,\dots,X_n}(x_1,\dots,x_n;p) \overset{i.i.d.}{=} \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i}, \quad \forall x_i \in \{0,1\}$$

The PMF of S is

$$f_S(s;p) = \mathbb{P}\{S = s\} = \binom{n}{s} p^s (1-p)^{n-s}, \quad s = 0, 1, \dots, n$$

Joint distribution of $X_1, \dots, X_n, S$ is

$$f_{X_1,\dots,X_n,S}(x_1,\dots,x_n,s;p) = \mathbb{P}\{X_1 = x_1, \dots, X_n = x_n, S = s\}$$

$$= \mathbb{P}\left\{X_1 = x_1, \dots, X_n = x_n, \sum_{i=1}^n X_i = s\right\}$$

$$= \begin{cases} \mathbb{P}\{X_1 = x_1, \dots, X_n = x_n\} = p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i}, & \text{if } s = \sum_{i=1}^n x_i, \forall x_i \in \{0,1\} \\ 0, & \text{otherwise} \end{cases}$$

So by definition of conditional distribution

$$f_{X_1,\dots,X_n|S}(x_1,\dots,x_n|s) = \frac{f_{X_1,\dots,X_n,S}(x_1,\dots,x_n,s;p)}{f_S(s;p)} = \begin{cases} \frac{1}{\binom{n}{s}}, & s = \sum_{i=1}^n x_i, \forall x_i \in \{0,1\} \\ 0, & \text{otherwise} \end{cases}$$

which does not depend on p. So $S = \sum_{i=1}^n X_i$ a sufficient statistic.

The definition of sufficient statistic is cumbersome to apply. The following theorem is more often used.

Theorem 5.1.1. **(Factorization criterion).** Let $X_1, \dots, X_n$ have joint PDF or PMF $f(x_1, \dots, x_n; \theta)$, $\theta \in \Theta$. Let $S_1 = A_1(X_1, \dots, X_n)$, $S_2 = A_2(X_1, \dots, X_n)$, ..., $S_k = A_k(X_1, \dots, X_n)$ be statistics. Then $S_1, \dots, S_k$ are jointly sufficient for θ if and only if

$$f(x_1,\dots,x_n;\theta) = h(x_1,\dots,x_n) g(A_1(x_1,\dots,x_n),\dots,A_k(x_1,\dots,x_n);\theta)$$

where h does not involve θ and g depends on $x_1, \dots, x_n$ only through the functions $A_1, \dots, A_k$.

Proof of discrete case. Suppose $S_1, \dots, S_k$ are joint sufficient. Let $S = (S_1, \dots, S_k)$, $s = (s_1, \dots, s_k)$. Then by definition

$$f_{X_1,\dots,X_n|S}(x_1, \dots, x_n|s) = \mathbb{P}\left\{X_1 = x_1, \dots, X_n = x_n | S = s\right\}$$

is free of θ. Let $x_1, \dots, x_n$ be fixed and let $s_1 = A_1(x_1, \dots, x_n)$, …, $s_k = A_k(x_1, \dots, x_n)$ be value of S at $x_1, \dots, x_n$. Then

$$\begin{aligned}
&f(x_1, \dots, x_n; \theta) = \mathbb{P}\left\{X_1 = x_1, \dots, X_n = x_n\right\} \\
&= \mathbb{P}\left\{X_1 = x_1, \dots, X_n = x_n, S_1 = s_1, \dots, S_k = s_k\right\} \\
&= \underbrace{\mathbb{P}\left\{X_1 = x_1, \dots, X_n = x_n | S_1 = s_1, \dots, S_k = s_k\right\}}_{\text{does not depend on } \theta \text{ by sufficient statistics}} \mathbb{P}\left\{S_1 = s_1, \dots, S_k = s_k\right\} \\
&= h(x_1, \dots, x_n) f_{S_1,\dots,S_k}(A_1(x_1, \dots, x_n), \dots, A_k(x_1, \dots, x_n); \theta) \\
&= h(x_1, \dots, x_n) g(A_1(x_1, \dots, x_n), \dots, A_k(x_1, \dots, x_n); \theta)
\end{aligned}$$

Now suppose

$$f(x_1, \dots, x_n; \theta) = h(x_1, \dots, x_n) g(A_1(x_1, \dots, x_n), \dots, A_k(x_1, \dots, x_n); \theta)$$

$$\begin{aligned}
&\mathbb{P}\left\{X_1 = x_1, \dots, X_n = x_n | S_1 = s_1, \dots, S_k = s_k\right\} \\
&= \frac{\mathbb{P}\left\{X_1 = x_1, \dots, X_n = x_n, S_1 = s_1, \dots, S_k = s_k\right\}}{\mathbb{P}\left\{S = s_1, \dots, S_k = s_k\right\}} \\
&= \frac{\mathbb{P}\left\{X_1 = x_1, \dots, X_n = x_n, A_1(X_1, \dots, X_n) = s_1, \dots, A_k(X_1, \dots, X_n) = s_k\right\}}{\mathbb{P}\left\{S_1 = s_1, \dots, S_k = s_k\right\}} \\
&= \begin{cases} \frac{\mathbb{P}\{X_1 = x_1, \dots, X_n = x_n\}}{\mathbb{P}\{S_1 = s_1, \dots, S_k = s_k\}}, & s_1 = A_1(x_1, \dots, x_n), \dots, s_k = A_k(x_1, \dots, x_n) \\ 0, & \text{otherwise} \end{cases} \\
&= \begin{cases} \frac{f(x_1, \dots, x_n; \theta)}{\sum_{\{x_1, \dots, x_n : A_j(x_1, \dots, x_n) = s_j, j = 1, \dots, k\}} f(x_1, \dots, x_n; \theta)}, & s_j = A_j(x_1, \dots, x_n), j = 1, \dots, k \\ 0, & \text{otherwise} \end{cases}
\end{aligned}$$

$$= \begin{cases} \frac{h(x_1, \dots, x_n) g(s_1, \dots, s_k; \theta)}{g(s_1, \dots, s_k; \theta) \sum_{\{x_1, \dots, x_n : A_j(x_1, \dots, x_n) = s_j, j = 1, \dots, k\}} h(x_1, \dots, x_n)}, & s_j = A_j(x_1, \dots, x_n), j = 1, \dots, k \\ 0, & \text{otherwise} \end{cases}$$

$$= \begin{cases} \frac{h(x_1,\ldots,x_n)}{\sum_{\{x_1,\ldots,x_n: A_j(x_1,\ldots,x_n)=s_j, j=1,\ldots,k\}} h(x_1,\ldots,x_n)}, & s_j = A_j(x_1,\ldots,x_n), j = 1,\ldots,k \\ 0, & \text{otherwise} \end{cases}$$

which is free of θ. General proof will be discussed in advanced course. $\square$

$\boxed{\textit{Example 5.1.2.}}$. Let $X_1,\ldots,X_n$ be a random sample from $\mathcal{E}xp(\theta,\eta)$

$$f(x;\theta,\eta) = \begin{cases} \frac{1}{\theta}e^{-\frac{x-\eta}{\theta}}, & x > \eta \\ 0, & \text{otherwise} \end{cases}$$

(θ,η) is unknown, $\theta > 0, \eta \in \mathbb{R}$. Find jointly sufficient statistics for (θ,η).

$$f(x_1,\ldots,x_n;\theta,\eta) \overset{i.i.d.}{=} \prod_{i=1}^{n}\left(\frac{1}{\theta}e^{-\frac{x_i-\eta}{\theta}}\mathbb{1}_{(\eta,\infty)}(x_i)^{1}\right)$$

$$= \underbrace{\frac{1}{\theta^n}e^{-\frac{\sum_{i=1}^n x_i - n\eta}{\theta}}\mathbb{1}_{(\eta,\infty)}\left(\min_{1\le i\le n} x_i\right)}_{\text{this is the g part}} = \underbrace{h(x_1,\ldots,x_n)}_{=1} g\left(\sum_{i=1}^{n} x_i, \min_{1\le i\le n} x_i;\theta,\eta\right)$$

By the factorization criterion theorem, $\sum_{i=1}^n X_i$, $\min\limits_{1\le i\le n} X_i$ are jointly sufficient statistics for (θ,η).

$\boxed{\textit{Example 5.1.3.}}$. Let $X_1,\ldots,X_n$ be a random sample from $\mathcal{U}nif(0,\theta)$, $\theta \in \mathbb{R}$ is unknown. Find jointly sufficient statistics for θ.

$$f(x_1,\ldots,x_n;\theta) \overset{i.i.d.}{=} \begin{cases} \frac{1}{\theta^n}, & 0 \le \min\limits_{1\le i\le n} x_i, \ \max\limits_{1\le i\le n} x_i \le \theta \\ 0 & \text{otherwise} \end{cases}$$

$$= \underbrace{\mathbb{1}_{[0,\infty)}\left(\min_{1\le i\le n} x_i\right)}_{=h(x_1,\ldots,x_n)} \underbrace{\frac{1}{\theta^n}\mathbb{1}_{(-\infty,\theta]}\left(\max_{1\le i\le n} x_i\right)}_{=g\left(\max_{1\le i\le n} x_i;\theta\right)}$$

[1]Indicator function: $\mathbb{1}_A(x) = \begin{cases} 1, & x \in A \\ 0, & x \notin A \end{cases}$

By the factorization criterion theorem, $\max\limits_{1 \leq i \leq n} X_i$ is sufficient statistic for θ. Note that for this problem, you might also arrive at

$$f(x_1, \dots, x_n; \theta) = \underbrace{1}_{=h(x_1,\dots,x_n)} \underbrace{\mathbb{1}_{[0,\infty)}\left(\min_{1 \leq i \leq n} x_i\right) \frac{1}{\theta^n} \mathbb{1}_{(-\infty,\theta]}\left(\max_{1 \leq i \leq n} x_i\right)}_{=g\left(\min\limits_{1 \leq i \leq n} x_i, \max\limits_{1 \leq i \leq n} x_i; \theta\right)}$$

By the factorization criterion theorem, $\min\limits_{1 \leq i \leq n} X_i$, $\max\limits_{1 \leq i \leq n} X_i$ are also jointly sufficient statistics for θ. But these are not as sufficient as simply $\max\limits_{1 \leq i \leq n} X_i$.

This brings up concept of minimal sufficient statistic.

Definition 5.1.2. **(Minimal sufficient).** A set of statistics $S_1, \dots, S_k$ is called a set of minimal sufficient statistics for the parameter θ (can be a vector) if they are jointly sufficient for θ and if they are a function of every other set of jointly sufficient statistics.

In other words, $S_1, \dots, S_k$ are minimal sufficient if they are the sufficient statistics that reduce the dimensionality the most. Note that proper application of factorization criterion will often yield minimal sufficient statistics. If not, an alternative is to apply factorization criterion iteratively until can no longer reduce.

5.2 Relation of Sufficiency to Estimators

Theorem 5.2.1. **(Sufficiency and MLE).** If $S_1, \dots, S_k$ are jointly sufficient statistics for θ and $\hat{\theta}$ is a unique MLE of θ, then $\hat{\theta}$ is a function of $S_1 = A_1(X_1, \dots, X_n)$, ..., $S_k = A_k(X_1, \dots, X_n)$.

Proof. By the factorization criterion

$$L(\theta) = f(x_1, \dots, x_n; \theta) = h(x_1, \dots, x_n) g(A_1(x_1, \dots, x_n), \dots, A_k(x_1, \dots, x_n); \theta)$$

To find MLE

$$\ln L(\theta) = \ln h(x_1, \dots, x_n) + \ln g(A_1(x_1, \dots, x_n), \dots, A_k(x_1, \dots, x_n); \theta)$$

If MLE is unique, it therefore is a function of $S_1 = A_1(X_1, \dots, X_n)$, ..., $S_k = A_k(X_1, \dots, X_n)$. $\square$

What good is sufficiency?

$\boxed{\text{Theorem 5.2.2.}}$ **(Rao-Blackwell).** Let $X_1, \dots, X_n$ have joint PDF or PMF, $f(x_1, \dots, x_n; \theta)$, $\theta \in \Theta$ and let $S = (S_1, \dots, S_k)$ be vector of jointly sufficient statistics for θ. If T is any unbiased estimator of $\tau(\theta)$ and if $T^* = \mathbb{E}(T|S)$, then

1. T^* is a statistic, and is an unbiased estimator of $\tau(\theta)$
2. T^* is a function of S
3. $\mathbb{V}arT^* \leq \mathbb{V}arT\ \forall\theta \in \Theta$ with strict "<" for at least one $\theta \in \Theta$

Proof.

1. By sufficiency, the conditional distribution of $X_1, \dots, X_n$ given S is free of θ. Let $T = t(X_1, \dots, X_n)$

$$\mathbb{E}(T|S = s) = \underbrace{\int_{\mathbb{R}} \cdots \int_{\mathbb{R}} t(x_1, \dots, x_n) f_{X_1,\dots,X_n|S}(x_1, \dots, x_n|s)\, dx_1 \dots dx_n}_{\text{some function of } s \text{ that does not depends on } \theta}$$

   So $T^* = \mathbb{E}(T|S)$ is a function of S. Because T^* is a function of S (so a function of X_i) and it does not depends on θ, it is a statistic.

2. Unbiasedness follows from

$$\mathbb{E}(T^*) = \mathbb{E}(\mathbb{E}(T|S)) \stackrel{1}{=} \mathbb{E}T = \tau(\theta), \quad \forall\theta \in \Theta$$

3.

$$\mathbb{V}arT = \mathbb{E}(T - \tau(\theta))^2 = \mathbb{E}(T - T^* + T^* - \tau(\theta))^2$$

$$= \underbrace{\mathbb{E}(T - T^*)^2}_{\geq 0} + 2\mathbb{E}\Big((T - T^*)(T^* - \tau(\theta))\Big) + \underbrace{\mathbb{E}(T^* - \tau(\theta))^2}_{=\mathbb{V}arT^*}$$

[1] Recall law of total expectation: $\mathbb{E}(\mathbb{E}(Y|X)) = \mathbb{E}Y$

Because

$$\mathbb{E}\Big((T-T^*)(T^*-\tau(\theta))\Big) = \mathbb{E}\Big((T-\mathbb{E}(T|S))(\mathbb{E}(T|S)-\tau(\theta))\Big)$$

$$= \mathbb{E}\left(\mathbb{E}\Big((T-\mathbb{E}(T|S))(\mathbb{E}(T|S)-\tau(\theta))\Big)\Big|S\right)$$

$$= \mathbb{E}\left((\mathbb{E}(T|S)-\tau(\theta))\,\underbrace{\mathbb{E}\Big((T-\mathbb{E}(T|S))\Big)}_{=\mathbb{E}(T)-\mathbb{E}(T)=0}\Big|S\right) = 0$$

So $\mathbb{V}arT \geq \mathbb{V}arT^*$.

□

Therefore, we can always improve or reduce variance of an unbiased estimator T by computing $T^* = \mathbb{E}(T|S)$ where S is sufficient for θ. But how far can we take this? We need the concept of completeness to answer it.

5.3 Completeness

Definition 5.3.1. **(Complete).** Let S have joint PDF or PMf $f_S(s;\theta)$ $\theta \in Ⓗ$

1. The family $\{f_S(s;\theta); \theta \in Ⓗ\}$ is called complete if $\mathbb{E}U(S) = 0$ for all $\theta \in Ⓗ$ implies $U(S) = 0$ with probability 1 $\forall\theta \in Ⓗ$.
2. Let $X_1, \ldots, X_n$ have joint PDF or PMF $f(x_1, \ldots, x_n)$, $\theta \in Ⓗ$. Then we say S is a complete sufficient statistic for θ if S is sufficient for θ, and if $\{f_S(s;\theta); \theta \in Ⓗ\}$ is a complete family of densities.

This means that suppose g, h such that $\mathbb{E}g(S) = \tau(\theta)$, $\mathbb{E}h(S) = \tau(\theta)$, $\forall\theta \in Ⓗ$, where S is complete sufficient. Then take $U(S) = g(S) - h(S)$, $\mathbb{E}U(S) = \tau(\theta) - \tau(\theta) = 0$, $\forall\theta \in Ⓗ$, so by completeness $U(S) = g(S) - h(S) = 0$ with probability 1 $\forall\theta \in Ⓗ$. So there is essentially only one unbiased estimator of $\tau(\theta)$ that is a function of S if S is complete sufficient. Rao-Blackwell theorem will them implies that this is the UMVUE. This fact was proven by the following theorem

Theorem 5.3.1. **(Lehmann-Scheffe).** Let $X_1, \ldots, X_n$ have joint PDF or PMF, $f(x_1, \ldots, x_n; \theta)$, $\theta \in \Theta$ and let S be a vector of complete sufficient statistics for θ. If $T^* = l^*(S)$ is an unbiased estimator of $\tau(\theta)$ and a function of S, then T^* is UMVUE.

Proof. Let T be any unbiased estimator of $\tau(\theta)$. By Rao-Blackwell, $\mathbb{E}(T|S)$ is also unbiased for $\tau(\theta)$ and has $\mathbb{V}ar\left(\mathbb{E}(T|S)\right) \leq \mathbb{V}arT$. But $\mathbb{E}(T|S)$ is a function of S, so take $U(S) = T^* - \mathbb{E}(T|S) = l^*(S) - \mathbb{E}(T|S)$. Then $\mathbb{E}(U(S)) = \tau(\theta) - \tau(\theta) = 0$, $\forall \theta \in \Theta$. So by completeness of S, $l^*(S) = \mathbb{E}(T|S)$ with probability 1 $\forall \theta \in \Theta$. And this holds no matter what T we start with. So $\mathbb{V}arT^* = \mathbb{V}ar(l^*(S)) = \mathbb{V}ar\left(\mathbb{E}(T|S)\right) \leq \mathbb{V}arT$ $\forall \theta \in \Theta$ for all unbiased T. □

5.4 Construction of UMVUE

We now have systematic way to find UMVUE

Summary 5.4.1. **(Construct UMVUE).**

Step 1: Find complete sufficient statistics S

Method 1: Use factorization criterion and definition of completeness

Step 1.1: Use factorization criterion to find (minimal) sufficient statistics S

Step 1.2: Check that the sufficient S is complete (assume $\mathbb{E}(U(S)) = 0$ $\forall \theta$ and verify that $U(S) = 0$ with probability 1 $\forall \theta \in \Theta$). If not, go to Step 1.1 and try to find a "more" sufficient set of statistics and repeat Step 1.2.

Method 2: Use Theorem 5.4.1

Step 2: If Step 1 succeeds, find a function of S that is unbiased for $\tau(\theta)$

Method 1: Ad-hoc (e.g. compute $\mathbb{E}(S)$ and adjust to be unbiased)

Method 2: Find some unbiased estimator T and compute $\mathbb{E}(T|S)$

Step 3: By Lehmann-Sheffe theorem, resulting estimator is UMVUE