Example: Sufficient Statistics for Random Samples from Various Families of Normal Distributions

Let $X = (X_1, \ldots, X_n)$ where $X_1, \ldots, X_n$ are iid $N(\mu, \sigma^2)$.

Consider different families of normal distributions.

$\Theta_1 = \{(\mu, \sigma^2) : \sigma^2 > 0\}$ (all normal distns)

$\Theta_2 = \{(\mu, \sigma^2) : \sigma^2 = \sigma_0^2\}$ (known variance)

$\Theta_3 = \{(\mu, \sigma^2) : \mu = \mu_0\,,\ \sigma^2 > 0\}$ (known mean)

(Draw $\Theta_1, \Theta_2, \Theta_3$ in (μ, σ^2) plane.)

For each space, the "obvious" sufficient statistic is different.

In all cases, the joint pdf of X is given by

$$\begin{aligned} f(x \mid \mu, \sigma^2) &= \prod_{i=1}^{n} (2\pi\sigma^2)^{-1/2} \exp\left\{-\frac{(x_i-\mu)^2}{2\sigma^2}\right\} \\ &= (2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}\sum_i (x_i-\mu)^2\right\}\,. \qquad (\dagger) \end{aligned}$$

Θ_3:

Here $\mu = \mu_0$, (a known value) so the "unknown" parameter is $\theta = \sigma^2$.

The joint pdf may be factored as

$$
\begin{aligned}
f(x \,|\, \sigma^2) &= \left[(2\pi\sigma^2)^{-n/2} \exp\left\{ -\frac{1}{2\sigma^2} \sum_i (x_i - \mu_0)^2 \right\} \right] \cdot 1 \\
&= g\left(\textstyle\sum_i (x_i - \mu_0)^2 \,,\, \sigma^2 \right) h(x) \\
&= g\left(T_3(x) \,,\, \sigma^2 \right) h(x)
\end{aligned}
$$

where

$$T_3(x) \equiv \sum_i (x_i - \mu_0)^2$$

so that $T_3 = T_3(X) = \sum_i (X_i - \mu_0)^2$ is a SS for Θ_3.

Note: T_3 is not even a statistic if μ is unknown (i.e., not fixed).

For the rest (Θ_1 and Θ_2), we modify (†) by substituting

$$\sum_{i=1}^{n} (x_i - \mu)^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2 + n(\bar{x} - \mu)^2$$

$$\text{where } \bar{x} = n^{-1} \sum_{i=1}^{n} x_i \,.$$

(This is an identity valid for all $x_1, \ldots, x_n$ and μ.)

Substituting in (†) and breaking up the exponential yields

$$
\begin{aligned}
& f(x \,|\, \mu, \sigma^2) \\
&= (2\pi\sigma^2)^{-n/2} \exp\left\{ -\frac{\sum_i (x_i - \bar{x})^2}{2\sigma^2} \right\} \exp\left\{ -\frac{n(\bar{x} - \mu)^2}{2\sigma^2} \right\} . \qquad (\ddagger)
\end{aligned}
$$

Θ_2:

Here $\sigma^2 = \sigma_0^2$, (a known value) so the "unknown" parameter is $\theta = \mu$.

Factoring the joint pdf (‡) as

$$
\begin{aligned}
& f(x \mid \mu) \\
&= \left[(2\pi\sigma_0^2)^{-n/2} \exp\left\{-\frac{\sum_i (x_i - \bar{x})^2}{2\sigma_0^2}\right\}\right] \left[\exp\left\{-\frac{n(\bar{x} - \mu)^2}{2\sigma_0^2}\right\}\right] \\
&= h(x)\, g(\bar{x}, \mu) = h(x)\, g(T_2(x), \mu) \\
&\quad \text{where } T_2(x) \equiv \bar{x}
\end{aligned}
$$

shows that $T_2 = T_2(X) = \bar{X}$ is a SS for Θ_2.

Θ_1:

Here both μ and σ^2 are unknown so $\theta = (\mu, \sigma^2)$.

It is clear that (‡) may be written as

$$
\begin{aligned}
f(x \mid \mu, \sigma^2) &= g\left(\bar{x}, \textstyle\sum_i (x_i - \bar{x})^2, \mu, \sigma^2\right) \cdot 1 \\
&= g(T_1(x), \theta)\, h(x) \\
&\quad \text{where} \quad T_1(x) = \left(\bar{x}, \textstyle\sum_i (x_i - \bar{x})^2\right)
\end{aligned}
$$

so that $T_1 = T_1(X) = \left(\bar{X}, \sum_i (X_i - \bar{X})^2\right)$ is a SS for Θ_1.

Note: T_1 is also a SS for Θ_2 and Θ_3. Neither T_2 or T_3 is a SS for Θ_1.

General Facts about SS

① If $T = T(X)$ is a SS for $\theta \in \Theta_A$,
and $\Theta_B \subset \Theta_A$,
then T is SS for $\theta \in \Theta_B$.

Proof: If $\mathcal{L}(X|T)$ is constant for $\theta \in \Theta_A$, then it is constant for $\theta \in \Theta_B$.

② If T is a SS (for $\theta \in \Theta$) and
$T = \phi(U)$ where $U = U(X)$,
then U is also a SS (for $\theta \in \Theta$).

Proof: (using FC)

$$T \text{ is SS} \Rightarrow f(x|\theta) = g(T(x)|\theta)\,h(x)$$

$$= g(\phi(U(x))|\theta)\,h(x)$$

$$= g^*(U(x)|\theta)\,h(x)$$

where $g^*(u|\theta) = g(\phi(u)|\theta)$

$\Rightarrow U(X)$ is SS.

(3) If $T = T(X)$ is a suff. stat. (for $\theta \in \Theta$), then $U = (S, T)$ is also a suff. stat. for any $S = S(X)$.

Proof: Immediate consequence of (2) by taking $\varphi(s,t) = t$.
With this choice of φ we have
$T = \varphi(U) \Rightarrow U$ is SS.

(4) If $T = T(X)$ and $U = U(X)$ are related by $T = \varphi(U)$ where ϕ is 1-1 function, then T is SS iff U is SS.

<u>Application</u> to random samples from various families of normal distributions:

<u>Recall</u>:

$T_1 = (\bar{x}, \Sigma(x_i - \bar{x})^2)$ is SS for $\Theta_1 = \{(\mu, \sigma^2) : \sigma^2 > 0\}$.

$T_2 = \bar{x}$ is SS for $\Theta_2 = \{(\mu, \sigma^2) : \sigma^2 = \sigma_0^2\}$.

$T_3 = \Sigma(x_i - \mu_0)^2$ is SS for
$\Theta_3 = \{(\mu, \sigma^2) : \mu = \mu_0, \sigma^2 > 0\}$.

T_1 is SS for Θ_1

$\Rightarrow$ T_1 is SS for Θ_2 and for Θ_3

since $\Theta_1 \supset \Theta_2$ and $\Theta_1 \supset \Theta_3$.

(Here we use Fact ①.)

T_2 is SS for Θ_2

$\Rightarrow$ T_1 is SS for Θ_2.

(by using Fact ③)

T_3 is SS for Θ_3 and

$$T_3 = \sum (X_i - \mu_0)^2 = \sum (X_i - \bar{X})^2 + n(\bar{X} - \mu_0)^2$$
$$= \varphi(T_1)$$

$\Rightarrow$ T_1 is SS for Θ_3 (by using Fact ②)

T_1 is SS for Θ_1

$\Rightarrow$ $\left(\bar{X}, \frac{1}{n-1} \sum (X_i - \bar{X})^2\right)$ is SS for Θ_1

$\Rightarrow$ $\left(\sum X_i, \sum X_i^2\right)$ is SS for Θ_1

since both of these are 1-1 functions of T_1. (Here we use Fact ④.)

A <u>minimal sufficient statistic</u> is a function of any other sufficient statistic.

$T = T(X)$ is <u>minimal</u> <u>suff.</u> if for every suff. stat. $S = S(X)$ there exists a function ψ such that $T = \psi(S)$, that is, $T(X) = \psi(S(X))$.

<u>Lehmann - Scheffe Theorem</u>

$X \sim P_\theta$, $\theta \in \Theta$.

$T(X)$ is a minimal sufficient statistic

iff

for all x, y

$T(x) = T(y)$ iff $\dfrac{f(x|\theta)}{f(y|\theta)}$ is constant as a function of θ.

Remark on the definition of Minimal Sufficient Statistics:

It is difficult to show a statistic is MSS directly from the definition. For proving MSS, we usually use the Lehmann-Scheffe Theorem.

However, it is often very easy to prove a statistic is **not** MSS using the definition. If S and T are two different sufficient statistics, and T can**not** be written as a function of S, then T is **not** minimal.

Example: Consider the three families of normal distributions used earlier.

T_1 and T_2 are both SS for Θ_2, but T_1 clearly cannot be written as a function of T_2. Thus T_1 is not a MSS for Θ_2.

Similarly, T_1 and T_3 are both SS for Θ_3, but T_1 clearly cannot be written as a function of T_3. Thus T_1 is not a MSS for Θ_3.

Comments

① In situations where the support of $f(x|\theta)$ depends on θ, a better statement (which avoids awkward $\frac{0}{0}$'s) is: For all x, y,

$$T(x) = T(y) \text{ iff } f(x,\theta) = c(x,y) f(y,\theta) \text{ for all } \theta.$$

② The "iff" can be broken down as two results.

(a) If $T(X)$ is sufficient, then for all x, y,

$$T(x) = T(y) \text{ implies } \frac{f(x|\theta)}{f(y|\theta)} \text{ constant in } \theta.$$

(b) A sufficient statistic $T(X)$ is minimal if for all x, y,

$$\frac{f(x|\theta)}{f(y|\theta)} \text{ constant in } \theta \text{ implies } T(x) = T(y).$$

Examples for Lehmann-Scheffe Theorem

Example: $X = (X_1, \ldots, X_n)$ iid $N(\mu, \sigma^2)$.

$T(X) = (\bar{x}, s^2)$ is MSS for (μ, σ^2).

(See Text) $\rightarrow$ $s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$

Example: $X = (X_1, \ldots, X_n)$ iid Uniform(α, β).

$\Theta = \{ (\alpha, \beta) : -\infty < \alpha < \beta < \infty \}$.

$T(X) = (X_{(1)}, X_{(n)})$ is MSS for (α, β).

$(X_{(1)} = \min X_i,\ X_{(n)} = \max X_i.)$

We must verify: for all x, y

$T(x) = T(y)$ iff $\exists\, c \neq 0$ such that

$$f(x|\theta) = c\, f(y|\theta) \quad \forall \theta.$$

(c does not involve θ, but can depend on x, y.)

In this case

$$f(x|\theta) = \prod_{i=1}^{n} \frac{1}{\beta - \alpha} I(\alpha \le x_i \le \beta)$$

$$= \frac{1}{(\beta - \alpha)^n} I(x_{(1)} \ge \alpha)\, I(x_{(n)} \le \beta)$$

Similarly

$$f(y|\theta) = \frac{1}{(\beta-\alpha)^n} I(y_{(1)} \geq \alpha)\, I(y_{(n)} \leq \beta).$$

Clearly

$(x_{(1)}, x_{(n)}) = (y_{(1)}, y_{(n)})$ implies

$$f(x|\theta) = f(y|\theta) \ \forall\, \theta \in \Theta.$$

Can take $c=1$.

This gives one direction. What about the other?

Define $A(x) = \{\theta : f(x|\theta) > 0\}$.

(α, β) with $\alpha < \beta$

Assume $\exists\, c \neq 0$ such $f(x|\theta) = c f(y|\theta)$ for all θ.

Then we must have $A(x) = A(y)$.

But $A(x) = \{(\alpha, \beta) : \alpha \leq x_{(1)}, \beta \geq x_{(n)}\}$ for any x.

Thus $A(x) = A(y)$ implies $(x_{(1)}, x_{(n)}) = (y_{(1)}, y_{(n)})$.

QED

Note: This style of argument can only work for examples similar to the uniform distn. where the support depends upon the parameter value.

<u>Example</u>: $X = (X_1, \ldots, X_n)$ iid Uniform$(\theta, \theta+1)$.

$T(X) = (X_{(1)}, X_{(n)})$ is MSS for θ.

(See text.)

Comments: The dimension of the MSS does <u>not</u> have to be the same as the dimension of the parameter.

→ "Shrinking" the parameter space does <u>not</u> always change the MSS.

When $X = (X_1, \ldots, X_n)$ iid Uniform(α, β),

$\Theta_1 = \{(\alpha, \beta) : \alpha < \beta\}$ and

$\Theta_2 = \{(\alpha, \beta) : \beta = \alpha + 1\}$

have the same MSS.

Example: (Random Sample Model)

Suppose $\underset{\sim}{X} = (X_1, \dots, X_n)$ iid $\psi(x|\theta)$ (pdf or pmf) where $\psi(x|\theta)$ is an arbitrary family of pdf's (pmf's).

Then $T(\underset{\sim}{X}) = (\underbrace{X_{(1)}, X_{(2)}, \dots, X_{(n)}}_{\text{the order statistics (data arranged in increasing order)}})$

is a suff. stat. for θ, but may not be minimal.

Proof: Use FC.

$$f(\underset{\sim}{x}|\theta) = \prod_{i=1}^{n} \psi(x_i|\theta) = \prod_{i=1}^{n} \psi(x_{(i)}|\theta) \cdot 1$$

$$= g(T(\underset{\sim}{x})|\theta)\, h(\underset{\sim}{x}).$$

Note: (assume $x_{(1)} < x_{(2)} < \dots < x_{(n)}$)

$P(\underset{\sim}{X} = \underset{\sim}{x} | T(\underset{\sim}{X}) = t) = \frac{1}{n!}$ if $\underset{\sim}{x}$ is any rearrangement of $x_{(1)}, \dots, x_{(n)}$.

($= 0$ otherwise).

All possible orderings are equally likely.

To generate from $\mathcal{L}(\underset{\sim}{X} | T)$,
place the values $\underbrace{x_{(1)}, \ldots, x_{(n)}}_{T(\underset{\sim}{x})}$ in a hat
and draw them out one by one.

<u>Comment:</u> For random sample models, the order statistics are often the MSS.

<u>Example:</u> $\underset{\sim}{X} = (X_1, \ldots, X_n)$ iid $\psi(x|\theta)$
with $\psi(x|\theta) = \frac{1}{\pi} \frac{1}{1+(x-\theta)^2}$ $\left(\begin{array}{c}\text{Cauchy}\\ \text{Location}\\ \text{Family}\end{array}\right)$.

Look at

$$\frac{f(\underset{\sim}{x}|\theta)}{f(\underset{\sim}{y}|\theta)} = \frac{\prod_{i=1}^{n} \frac{1}{\pi} \frac{1}{1+(x_i-\theta)^2}}{\prod_{i=1}^{n} \frac{1}{\pi} \frac{1}{1+(y_i-\theta)^2}} = \frac{\prod_{i=1}^{n} (1+(y_{(i)}-\theta)^2)}{\prod_{i=1}^{n} (1+(x_{(i)}-\theta)^2)}$$

No further (obvious) simplifications occur.
The ratio is constant (in θ) iff $x_{(i)} = y_{(i)} \; \forall i$.
But this is hard to show rigorously.

Example: Suppose $X \sim P_\theta$, $\theta \in \Theta$ and P_θ has a joint pdf (or pmf) $f(x \,|\, \theta)$.

Fact: X is a SS for θ.

Define $T = T(X) = X$. (T is the identity function.)

Proof from FC:

$f(x \,|\, \theta) = f(x \,|\, \theta) \cdot 1 = g(T(x) \,|\, \theta) \cdot h(x)$

where $g \equiv f$ and $h(x) \equiv 1$. Thus T is SS.

Proof from definition of SS:

$\mathcal{L}(X \,|\, T(X) = t) = \mathcal{L}(X \,|\, X = t) = \delta_t$

where δ_t is the probability measure (distn) which places all its mass at the point (data set) t.

(This fact is not useful, but only intended to illustrate the definitions.)

Example continued: Further suppose $X = (X_1, \ldots, X_n)$ where $X_1, \ldots, X_n$ are iid from the pdf (pmf) $\psi(x \,|\, \theta)$.

Fact: $T(X) = X = (X_1, \ldots, X_n)$ is **not** a MSS.

Proof from definition of MSS:

Let $S = S(X) = (X_{(1)}, X_{(2)}, \ldots, X_{(n)})$ (the order statistics).

Since we have a random sample model, S is a SS.

But clearly T is **not** a function of S. (You can**not** recover the original ordering of the data given only the order statistics.)

Thus T is **not** a MSS.