MA 3140: Statistical Inference

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

Minimal Sufficiency

Minimal Sufficiency

- Recall that in any given problem, numerous sufficient statistics are possible.
- **Problem: Which one to use?**
- **Solution: The statistic which leads to maximum reduction of the data while still retaining all the information abut the parameter θ.**

Minimal Sufficient Statistic (MSS)

- A statistic T is minimal sufficient, if
  - T is sufficient, and
  - if S is any other sufficient statistic, then $T = g(S)$ for some function g.

- **How to check if a statistic T is minimal sufficient?**

Theorem - Lehmann and Scheffe (1950, 1955):

Let $f(\boldsymbol{x}, \theta)$ denote the joint pdf/pmf of $\boldsymbol{X}$. Suppose there exists a function $T(\boldsymbol{X})$ such that, for every two sample points $\boldsymbol{x}$ and $\boldsymbol{y}$, the ratio $f(\boldsymbol{x}, \theta)/f(\boldsymbol{y}, \theta)$ does not depend on θ iff $T(\boldsymbol{x}) = T(\boldsymbol{y})$. Then, $T(\boldsymbol{X})$ is a MSS for θ.

Example 1

Let $X_1, \ldots, X_n \sim P(\lambda)$, $\lambda > 0$. Find the MSS for λ.

Solution: The joint pmf is

$$f(\boldsymbol{x}, \lambda) = \frac{e^{-n\lambda} \lambda^{\sum x_i}}{\prod_{i=1}^{n} x_i!}$$

Therefore, $$\frac{f(\boldsymbol{x}, \lambda)}{f(\boldsymbol{x}, \lambda)} = \frac{\prod_{i=1}^{n} y_i!}{\prod_{i=1}^{n} x_i!} \lambda^{\sum x_i - \sum y_i}$$

is independent of λ iff $\sum x_i = \sum y_i$.

Thus, we conclude that $T(\boldsymbol{X}) = \sum X_i$ is minimal sufficient.

Example 2

Let $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$, where both $\mu \in \mathbb{R}$ and $\sigma^2 > 0$ are unknown. Find the MSS for μ and σ^2.

Solution: The joint pdf is

$$f(\boldsymbol{x}, \mu, \sigma^2) = \frac{1}{(\sigma\sqrt{2\pi})^n} e^{-\frac{\sum x_i^2}{2\sigma^2} + \frac{\mu \sum x_i}{\sigma^2} - \frac{n\mu^2}{2\sigma^2}}$$

Therefore, $$\frac{f(\boldsymbol{x}, \mu, \sigma^2)}{f(\boldsymbol{y}, \mu, \sigma^2)} = \frac{e^{-\frac{\sum x_i^2}{2\sigma^2} + \frac{\mu \sum x_i}{\sigma^2}}}{e^{-\frac{\sum y_i^2}{2\sigma^2} + \frac{\mu \sum y_i}{\sigma^2}}}$$

is independent of μ and σ^2 iff $\sum X_i = \sum Y_i$ and $\sum X_i^2 = \sum Y_i^2$.

Thus, $(\sum X_i, \sum X_i^2)$ is minimal sufficient, or we can say that $(\overline{X}, S^2)$ is minimal sufficient.

Remark

- MSS is not unique.
- Any 1-to-1 function of a MSS is also minimal sufficient.
- **Example:**

  Consider $X_1, \dots, X_n \sim U(\theta, \theta+1)$, $-\infty < \theta < \infty$,
  $T_1(\boldsymbol{X}) = (X_{(1)}, X_{(n)})$ is MSS, and so is
  $T_2(\boldsymbol{X}) = (X_{(n)} - X_{(1)}, (X_{(1)} + X_{(n)})/2)$.

Recall

- In any problem there are, in fact, many sufficient statistics.
  - The complete sample is always sufficient.
  - Any one-to-one function of a sufficient statistic is a sufficient statistic.
- **Problem:** Can we determine whether one sufficient statistic is better than the other?
- **Solution:** Find a statistic that achieves the maximum data reduction while still retaining all the information about the parameter θ, i.e., find a MSS.
- **Note:** A MSS is not unique - any one-to-one function of a MSS is a MSS.

Ancillary Statistic

Ancillary Statistic

- A statistic $V(X)$ is said to be ancillary for θ if its distribution does not depend on the parameter θ.
  - It does not contain any information about θ.
  - Its distribution is fixed and known, unrelated to the parameter.
- Paradoxically, an ancillary statistic, when used in conjunction with other statistics, sometimes does contain valuable information about θ.

Example 1: Uniform Ancillary Statistic

Let $X_1, \ldots, X_n \sim U(\theta, \theta + 1)$, $-\infty < \theta < \infty$.

$X_{(1)}, \ldots, X_{(n)}$: order statistics from the sample.

Claim: $R = X_{(n)} - X_{(1)}$ is an ancillary statistic.

Proof: We know that cdf of each X_i is

$$F_\theta(x) = \begin{cases} 0, & x \le \theta \\ x - \theta, & \theta < x < \theta + 1 \\ 1, & \theta + 1 \le x. \end{cases}$$

Example 1 cont'd

The joint pdf of $X_{(1)}$ and $X_{(n)}$ is

$$g_\theta(x_{(1)}, x_{(n)}) = n(n-1)(x_{(n)}-x_{(1)})^{n-2}, \quad \theta < x_{(1)} < x_{(n)} < \theta+1.$$

Make the transformation $R = X_{(n)} - X_{(1)}$ and $M = (X_{(1)} + X_{(n)})/2$.

So, the inverse transformation is $X_{(1)} = (2M - R)/2$ and $X_{(n)} = (2M + R)/2$ with Jacobian 1.

The joint pdf of R and M is

$$h_\theta(r, m) = n(n-1)r^{n-2}, \quad 0 < r < 1, \quad \theta+(r/2) < m < \theta+1-(r/2).$$

Example 1 cont'd

Thus the pdf of R is

$$h_\theta(r) = \int_{\theta+(r/2)}^{\theta+1-(r/2)} n(n-1)r^{n-2}dm = n(n-1)r^{n-2}(1-r), \quad 0 < r < 1.$$

It can be clearly seen that the pdf is same for all θ.

Thus, the distribution of R does not depend on θ, and R is ancillary.

Thanks for your patience!