Biostat 602 Winter 2017

Lecture Set 3

Principles of Data Reduction (Minimal Sufficiency)

Minimal Sufficient Statistic

Reading: CB 6.2

- Sufficient statistics are not unique.
- $T(\mathbf{x}) = \mathbf{x}$: The random sample itself is a trivial sufficient statistic for any θ.
- The set of order statistics $T(\mathbf{X}) = (X_{(1)}, \cdots, X_{(n)})$ is always a sufficient statistic for θ, if $X_1, \cdots, X_n$ are iid.
- For any sufficient statistic $T(\mathbf{X})$, its one-to-one function $q(T(\mathbf{X}))$ is also a sufficient statistic for θ.

Question Can we find a sufficient statistic that achieves the maximum data reduction?

Definition 6.2.11

A sufficient statistic $T(\mathbf{X})$ is called a *minimal sufficient statistic* if, for any other sufficient statistic $T'(\mathbf{X})$, $T(\mathbf{X})$ is a function of $T'(\mathbf{X})$.

Remarks

- $T(\mathbf{X})$ is a function of $T'(\mathbf{X}) \implies$ if $T'(\mathbf{x}) = T'(\mathbf{y})$ then $T(\mathbf{x}) = T(\mathbf{y})$.
- The sample space $\mathcal{X}$ consists of every possible sample - *finest* partition
- Given $T(\mathbf{X})$, $\mathcal{X}$ can be partitioned into A_t where $t \in \mathcal{T} = \{t : t = T(\mathbf{X}) \text{ for some } \mathbf{x} \in \mathcal{X}\}$
- Maximum data reduction is achieved when cardinality of $\mathcal{T}$ is minimal.
- If size of $\mathcal{T}' = \{t : t = T'(\mathbf{x}) \text{ for some } \mathbf{x} \in \mathcal{X}\}$ is not less than that of $\mathcal{T}$, then $\mathcal{T}$ is a minimal sufficient statistic. In this case, the partition induced by $\mathcal{T}$ is the *coarsest* possible.

Question 1: If T is *minimal sufficient*, is a one-to-one function of T also *minimal sufficient*?

Question 2: Is there always a one-to-one function between any two *minimal sufficient* statistics?

Note that sufficiency is tied to the parameter under consideration. Consider a random sample $X_1, \ldots, X_n$ from a $N(\mu, \sigma^2)$ population, where σ^2 is **known**. We have seen earlier that in this case, $T(\mathbf{X}) = \overline{X}$ is sufficient for μ. Consider the statistic $\mathbf{T}'(\mathbf{X}) = (T_1(\mathbf{X}), T_2(\mathbf{X})) = (\overline{X}, S^2)$.

- $\mathbf{T}'$ is sufficient for μ (factorization theorem).

- T achieves a coarser data reduction than $\mathbf{T}'$.

- No additional information is gained about μ from $\mathbf{T}'$.

- When σ^2 is not known, T is **not sufficient** for (μ, σ^2). In this case, $\mathbf{T}' = (\overline{X}, S^2)$ is jointly sufficient for (μ, σ^2).

Question Is $(\overline{X}, S^2)$ *minimal sufficient* for (μ, σ^2) (how to check)?

Theorem 6.2.13

Suppose $f_{\mathbf{X}}(\mathbf{x}|\theta)$ be the pdf or pmf of a sample $\mathbf{X}$ parameterized by θ. Suppose there exists a function $T(\mathbf{x})$ such that, for any two sample points $\mathbf{x}$ and $\mathbf{y}$, the ratio $f_{\mathbf{X}}(\mathbf{x}|\theta)/f_{\mathbf{X}}(\mathbf{y}|\theta)$ is constant as a function of θ if and only if $T(\mathbf{x}) = T(\mathbf{y})$. Then $T(\mathbf{x})$ is *minimal sufficient* for θ.

In other words

- $f_{\mathbf{X}}(\mathbf{x}|\theta)/f_{\mathbf{X}}(\mathbf{y}|\theta)$ is constant as a function of $\theta \Longrightarrow T(\mathbf{x}) = T(\mathbf{y})$.
- $T(\mathbf{x}) = T(\mathbf{y}) \Longrightarrow f_{\mathbf{X}}(\mathbf{x}|\theta)/f_{\mathbf{X}}(\mathbf{y}|\theta)$ is constant as a function of θ

Proof:

Example 1: Let X_1, X_2, X_3 be i.i.d. Bernoulli(p). Consider

$$T_1(\mathbf{X}) = X_1 + X_2 + X_3.$$

(a) Is T_1 sufficient for p?

$$\begin{aligned}
f_{\mathbf{X}}(\mathbf{x}|p) &= p^{x_1+x_2+x_3}(1-p)^{3-x_1-x_2-x_3} \\
&= \left(\frac{p}{1-p}\right)^{x_1+x_2+x_3}(1-p)^3
\end{aligned}$$

$$\begin{aligned}
h(\mathbf{x}) &= 1 \\
g(t|p) &= \left(\frac{p}{1-p}\right)^{t}(1-p)^3
\end{aligned}$$

Since

$$f_{\mathbf{X}}(\mathbf{x}|p) = g(x_1 + x_2 + x_3|p)h(\mathbf{x}),$$

by factorization Theorem, T_1 is sufficient for p.

(b) Is T_1 minimal sufficient for p?

$$\begin{aligned}
\frac{f_{\mathbf{X}}(\mathbf{x}|\theta)}{f_{\mathbf{X}}(\mathbf{y}|\theta)} &= \frac{p^{\sum x_i}(1-p)^{3-\sum x_i}}{p^{\sum y_i}(1-p)^{3-\sum y_i}} \\
&= \left(\frac{p}{1-p}\right)^{\sum x_i - \sum y_i}
\end{aligned}$$

– If $T_1(\mathbf{x}) = T_1(\mathbf{y})$, i.e. $\sum x_i = \sum y_i$, then the ratio does not depend on p.
– The ratio above is constant as a function of p only if $\sum x_i = \sum y_i$, i.e. $T_1(\mathbf{x}) = T_1(\mathbf{y})$.

Therefore, $T_1(\mathbf{X}) = \sum X_i$ is a minimal sufficient statistic for p by Theorem 6.2.13.

Example 2: Same premise as in Example 1. Consider

$$\mathbf{T}_2(\mathbf{X}) = (X_1 + X_2, X_3).$$

(a) Is $\mathbf{T_2}$ sufficient for p?

$$\begin{aligned} f_{\mathbf{X}}(\mathbf{x}|p) &= p^{x_1+x_2+x_3}(1-p)^{3-x_1-x_2-x_3} \\ &= p^{x_1+x_2}(1-p)^{2-x_1-x_2}p^{x_3}(1-p)^{1-x_3} \end{aligned}$$

$$h(\mathbf{x}) = 1$$

$$g(t_1, t_2|p) = p^{t_1}(1-p)^{2-t_1}p^{t_2}(1-p)^{1-t_2}$$

$$\text{and} \quad f_{\mathbf{X}}(\mathbf{x}|p) = g(x_1 + x_2, x_3|p)h(\mathbf{x})$$

Hence $\mathbf{T}_2(\mathbf{X}) = (X_1 + X_2, X_3)$ is sufficient for p.

(b) Is $\mathbf{T_2}$ minimal sufficient for p?

Let $A(\mathbf{X}) = X_1 + X_2$, and $B(\mathbf{X}) = X_3$.

$$\begin{aligned} f_{\mathbf{X}}(\mathbf{x}|p) &= p^{x_1+x_2}(1-p)^{2-x_1-x_2}p^{x_3}(1-p)^{1-x_3} \\ &= p^{A(\mathbf{x})+B(\mathbf{x})}(1-p)^{3-A(\mathbf{x})-B(\mathbf{x})} \end{aligned}$$

$$\begin{aligned} \frac{f_{\mathbf{X}}(\mathbf{x}|\theta)}{f_{\mathbf{X}}(\mathbf{y}|\theta)} &= \frac{p^{A(\mathbf{x})+B(\mathbf{x})}(1-p)^{3-A(\mathbf{x})-B(\mathbf{x})}}{p^{A(\mathbf{y})+B(\mathbf{y})}(1-p)^{3-A(\mathbf{x})-B(\mathbf{y})}} \\ &= \left(\frac{p}{1-p}\right)^{A(\mathbf{x})+B(\mathbf{x})-A(\mathbf{y})-B(\mathbf{y})} \end{aligned}$$

- The ratio above is constant as a function of p if (but not only if) $A(\mathbf{x}) = A(\mathbf{y})$ and $B(\mathbf{x}) = B(\mathbf{y})$
- The ratio is still constant as long as $A(\mathbf{x}) + B(\mathbf{x}) = A(\mathbf{y}) + B(\mathbf{y})$, even though $A(\mathbf{x}) \neq A(\mathbf{y})$ and $B(\mathbf{x}) \neq B(\mathbf{y})$

Therefore, $\mathbf{T}_2(\mathbf{X}) = (A(\mathbf{X}), B(\mathbf{X})) = (X_1 + X_2, X_3)$ is not a minimal sufficient statistic for p by Theorem 6.2.13.

Partition of the Sample Space

X_1	X_2	X_3	$\mathbf{T_2}(X) = (X_1 + X_2, X_3)$	$T_1(\mathbf{X}) = X_1 + X_2 + X_3$
0	0	0	(0,0)	0
0	0	1	(0,1)	
0	1	0		3*1
1	0	0	2*(1,0)	
0	1	1		
1	0	1	2*(1, 1)	3*2
1	1	0	(2,0)	
1	1	1	(2,1)	3

Clearly the partition induced by T_1 is coarser than the one induced by $\mathbf{T_2}$.

Some Algebraic Results

Assume that $a, b, c, d, a_1, \cdots, a_n$ are constants.

1. $a\theta^2 + b\theta + c = 0$ for any $\theta \in \mathbb{R} \quad \Leftrightarrow a = b = c = 0.$

2. $\sum_{i=1}^{k} a_i \theta^i = c$ for any $\theta \in \mathbb{R} \quad \Leftrightarrow a_1 = \cdots = a_k = 0,\ c = 0.$

3. $a\theta_1 + b\theta_2 + c = 0$ for all $(\theta_1, \theta_2) \in \mathbb{R}^2 \quad \Leftrightarrow a = b = c = 0.$

4. The following equation is constant

$$\frac{1 + a_1\theta + a_2\theta^2 + \cdots + a_k\theta_k^k}{1 + b_1\theta + b_2\theta^2 + \cdots + b_k\theta_k^k}$$

   $\Leftrightarrow a_1 = b_1, \cdots, a_k = b_k.$

   Note that this does not hold without the constant 1, for example,

$$\frac{\theta + 2\theta^2}{2\theta + 4\theta^2} = \frac{1}{2}$$

5. $\dfrac{I(a < \theta < b)}{I(c < \theta < d)}$ is a constant function of $\theta \quad \Leftrightarrow a = c$, and $b = d$.

6. θ^t is constant function of $\theta \quad \Leftrightarrow t = 0.$

Example 3: Let $X_1, \cdots, X_n$ be iid Uniform$(\theta, \theta+1)$, where $-\infty < \theta < \infty$. Find a minimal sufficient statistic for θ.

Joint pdf of X

$$f_{\mathbf{X}}(\mathbf{x}|\theta) = \prod_{i=1}^{n} I(\theta < x_i < \theta + 1)$$

Hence,

$$\begin{aligned}
\frac{f_{\mathbf{X}}(\mathbf{x}|\theta)}{f_{\mathbf{X}}(\mathbf{y}|\theta)} &= \frac{\prod_{i=1}^{n} I(\theta < x_i < \theta+1)}{\prod_{i=1}^{n} I(\theta < y_i < \theta+1)} \\
&= \frac{I(\theta < x_1 < \theta+1, \cdots, \theta < x_n < \theta+1)}{I(\theta < y_1 < \theta+1, \cdots, \theta < y_n < \theta+1)} \\
&= \frac{I(\theta < x_{(1)} \text{ and } x_{(n)} < \theta+1)}{I(\theta < y_{(1)} \text{ and } y_{(n)} < \theta+1)} \\
&= \frac{I(x_{(n)} - 1 < \theta < x_{(1)})}{I(y_{(n)} - 1 < \theta < y_{(1)})}
\end{aligned}$$

The ratio above is constant if and only if $x_{(1)} = y_{(1)}$ and $x_{(n)} = y_{(n)}$. Therefore, $\mathbf{T}(\mathbf{X}) = (X_{(1)}, X_{(n)})$ is a minimal sufficient statistic for θ.

Example 4(a): Let $X_1, \cdots, X_n$ be iid $\mathcal{N}(\mu, \sigma^2)$, where both parameters are unknown. The parameter is a vector: $\boldsymbol{\theta} = (\mu, \sigma^2)$. The problem is to use find a minimal sufficient statistic for $\boldsymbol{\theta}$.

The joint pdf

$$f_{\mathbf{X}}(\mathbf{x}|\mu, \sigma^2) = (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2\right)$$

$$
\begin{aligned}
\frac{f_{\mathbf{X}}(\mathbf{x}|\mu,\sigma^2)}{f_{\mathbf{X}}(\mathbf{y}|\mu,\sigma^2)} &= \exp\left(-\frac{\sum_{i=1}^n (x_i-\mu)^2}{2\sigma^2}\right) / \exp\left(-\frac{\sum_{i=1}^n (y_i-\mu)^2}{2\sigma^2}\right) \\
&= \exp\left[-\frac{1}{2\sigma^2}\left(\sum_{i=1}^n (x_i^2 - 2\mu x_i + \mu^2) - \sum_{i=1}^n (y_i^2 - 2\mu y_i + \mu^2)\right)\right] \\
&= \exp\left[-\frac{1}{2\sigma^2}\left(\sum_{i=1}^n x_i^2 - \sum_{i=1}^n y_i^2\right) + \frac{\mu}{\sigma^2}\left(\sum_{i=1}^n x_i - \sum_{i=1}^n y_i\right)\right]
\end{aligned}
$$

The ratio above will not depend on (μ,σ^2) if and only if

$$
\left\{\begin{array}{rcl}
\sum_{i=1}^n x_i^2 &=& \sum_{i=1}^n y_i^2 \\
\sum_{i=1}^n x_i &=& \sum_{i=1}^n y_i
\end{array}\right.
$$

Therefore, $\mathbf{T}(\mathbf{X}) = (\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2)$ is a minimal sufficient statistic for (μ,σ^2) by Theorem 6.2.13

Define $\mathbf{T}'(\mathbf{X}) = (\overline{X}, \sum(X_i - \overline{X})^2/(n-1)) = (\overline{X}, S^2)$. Then, there exist one-to-one functions such that

$$
\begin{aligned}
\sum X_i &= g_1\left(\overline{X}, \sum(X_i-\overline{X})^2/(n-1)\right) \\
\sum X_i^2 &= g_2\left(\overline{X}, \sum(X_i-\overline{X})^2/(n-1)\right)
\end{aligned}
$$

and

$$
\begin{aligned}
\overline{X} &= h_1(\sum X_i, \sum X_i^2) \\
\sum(X_i-\overline{X})^2/(n-1) &= h_2(\sum X_i, \sum X_i^2)
\end{aligned}
$$

Thus $\mathbf{T}'$ is minimal sufficient.

Example 4(b): Let $X_1, \cdots, X_n$ be iid $\mathcal{N}(\mu, \sigma^2)$. In each of the following cases, identify a minimal sufficient statistic for the parameter of interest.

- When $\sigma = \sqrt{\mu}$.
- When $\sigma = \mu$.

Example 5: Let $X_1, \cdots, X_n$ be a random sample from $Gamma(\alpha, \beta)$ with pdf

$$f(x|\alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} \exp(-x/\beta).$$

Define $T_1(\mathbf{x}) = \prod_{i=1}^n x_i, \;\; T_2(\mathbf{x}) = \sum_{i=1}^n x_i$. Show that (T_1, T_2) are jointly sufficient for (α, β). Are they minimal sufficient?