Biostatistics 602 - Statistical Inference
Lecture 03
Minimal Sufficient Statistics

Hyun Min Kang

January 17th, 2013

Last Lecture - Key Questions

1. How do we show that a statistic is sufficient for θ?
2. What is a necessary and sufficient condition for a statistic to be sufficient for θ?
3. What is an effective strategy to find sufficient statistics using the Factorization Theorem?
4. Is the dimension of a sufficient statistic the always same to the dimension of the parameters?

Recap - Sufficient Statistic

Definition 6.2.1

A statistic $T(\mathbf{X})$ is a *sufficient statistic* for θ if the conditional distribution of sample $\mathbf{X}$ given the value of $T(\mathbf{X})$ does not depend on θ.

Recap - A Theorem for Sufficient Statistics

Theorem 6.2.2

- Let $f_{\mathbf{X}}(\mathbf{x}|\theta)$ is a joint pdf or pmf of X
- and $q(t|\theta)$ is the pdf or pmf of $T(\mathbf{X})$.
- Then $T(\mathbf{X})$ is a sufficient statistic for θ,
- if, for every $\mathbf{x} \in \mathcal{X}$,
- the ratio $f_{\mathbf{X}}(\mathbf{x}|\theta)/q(T(\mathbf{x})|\theta)$ is constant as a function of θ.

Recap - Factorization Theorem

Theorem 6.2.6 - Factorization Theorem

- Let $f_{\mathbf{X}}(\mathbf{x}|\theta)$ denote the joint pdf or pmf of a sample $\mathbf{X}$.
- A statistic $T(\mathbf{X})$ is a sufficient statistic for θ, if and only if
    - There exists function $g(t|\theta)$ and $h(\mathbf{x})$ such that,
    - for all sample points $\mathbf{x}$,
    - and for all parameter points θ,
    - $f_{\mathbf{X}}(\mathbf{x}|\theta) = g(T(\mathbf{x})|\theta)h(\mathbf{x})$.

Minimal Sufficient Statistic

Sufficient statistics are not unique

- $\mathbf{T}(\mathbf{x}) = \mathbf{x}$: The random sample itself is a trivial sufficient statistic for any θ.
- An ordered statistic $(X_{(1)}, \cdots, X_{(n)})$ 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?

Minimal Sufficient Statistic

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 $T'(\mathbf{X})$.

Why is this called "minimal" sufficient statistic?

- 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 $|\mathcal{T}|$ is minimal.
- If size of $\mathcal{T}' = t : t = T'(\mathbf{x}) \text{ for some } \mathbf{x} \in \mathcal{X}$ is not less than $|\mathcal{T}|$, then $|\mathcal{T}|$ can be called as a minimal sufficient statistic.

Theorem for Minimal Sufficient Statistics

Theorem 6.2.13

- $f_{\mathbf{X}}(\mathbf{x})$ be pmf or pdf of a sample $\mathbf{X}$.
- Suppose that there exists a function $T(\mathbf{x})$ such that,
- For every 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 a minimal sufficient statistic for θ.

In other words..

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

Example from the first lecture

Problem

- $X_1, X_2, X_3 \overset{\text{i.i.d.}}{\sim} \text{Bernoulli}(p)$
- Q1: Is $\mathbf{T}_1(\mathbf{X}) = (X_1 + X_2, X_3)$ a sufficient statistic for p?
- Q2: Is $T_2(\mathbf{X}) = X_1 + X_2 + X_3$ a minimal sufficient statistic for p?
- Q3: Is $\mathbf{T}_1(\mathbf{X}) = (X_1 + X_2, X_3)$ a minimal sufficient statistic for p?

Is $\mathbf{T}_1(\mathbf{X}) = (X_1 + X_2, X_3)$ a sufficient statistic?

$$
\begin{aligned}
f_{\mathbf{X}}(\mathbf{x}|p) &= p^{x_1+x_2+x_3}(1-p)^{3-x_1-x_2-x_3} \\
&= p^{x_1+x_2}(1-p)^{2-x_1-x_2}p^{x_3}(1-p)^{1-x_3} \\
h(\mathbf{x}) &= 1 \\
g(t_1, t_2|p) &= p^{t_1}(1-p)^{2-t_1}p^{t_2}(1-p)^{1-t_2} \\
f_{\mathbf{X}}(\mathbf{x}|p) &= g(x_1+x_2, x_3|p)h(\mathbf{x})
\end{aligned}
$$

By Factorization Theorem, $\mathbf{T}_1(\mathbf{X}) = (X_1 + X_2, X_3)$ is a sufficient statistic.

Is $T_2(\mathbf{X}) = (X_1 + X_2 + X_3)$ a minimal sufficient statistic?

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

- If $T_2(\mathbf{x}) = T_2(\mathbf{y})$, 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. $T_2(\mathbf{x}) = T_2(\mathbf{y})$.

Therefore, $T_2(\mathbf{X}) = \sum X_i$ is a minimal sufficient statistic for p by Theorem 6.2.13.

Is $\mathbf{T}_1(\mathbf{X}) = (X_1 + X_2, X_3)$ minimal sufficient?

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

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

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

Therefore, $\mathbf{T}_1(\mathbf{X}) = (A(\mathbf{X}), B(\mathbf{X})) = (X_1 + X_2, X_3)$ is not a minimal sufficient statistic for p by Theorem 6.2.13.

Partition of sample space

<table>
<tr><th>X_1</th><th>X_2</th><th>X_3</th><th>$\mathbf{T}_1(X) = (X_1 + X_2, X_3)$</th><th>$T_2(\mathbf{X}) = X_1 + X_2 + X_3$</th></tr>
<tr><td>0</td><td>0</td><td>0</td><td>(0,0)</td><td>0</td></tr>
<tr><td>0</td><td>0</td><td>1</td><td>(0,1)</td><td rowspan="3">1</td></tr>
<tr><td>0</td><td>1</td><td>0</td><td rowspan="2">(1,0)</td></tr>
<tr><td>1</td><td>0</td><td>0</td></tr>
<tr><td>0</td><td>1</td><td>1</td><td rowspan="2">(1,1)</td><td rowspan="3">2</td></tr>
<tr><td>1</td><td>0</td><td>1</td></tr>
<tr><td>1</td><td>1</td><td>0</td><td>(2,0)</td></tr>
<tr><td>1</td><td>1</td><td>1</td><td>(2,1)</td><td>3</td></tr>
</table>

Background knowledges for proving if and only if

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

1. $a\theta^2 + b\theta + c = 0$ for any $\theta \in \mathbb{R}$
   $\Leftrightarrow a = b = c = 0$.
2. $\sum_{i=1}^k a_i\theta^i = c$ for any $\theta \in \mathbb{R}$
   $\Leftrightarrow a_1 = \cdots = a_k = 0$.
3. $a\theta_1 + b\theta_2 + c = 0$ for all $(\theta_1, \theta_2) \in \mathbb{R}^2$
   $\Leftrightarrow a = b = c = 0$.

Background knowledges for proving if and only if

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. $\frac{I(a<\theta<b)}{I(c<\theta<d)}$ is constant a a function of θ. $\Leftrightarrow a = c$, and $b = d$.
6. θ^t is constant function of θ. $\Leftrightarrow t = 0$.

Uniform Minimal Sufficient Statistic

Example 6.2.15

- $X_1, \cdots, X_n \overset{\text{i.i.d.}}{\sim} \text{Uniform}(\theta, \theta + 1)$, where $-\infty < \theta < \infty$.
- Find a minimal sufficient statistic for θ.

Joint pdf of $\mathbf{X}$

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

Uniform Minimal Sufficient Statistic

Examine $f_{\mathbf{X}}(\mathbf{x}|\theta)/f_{\mathbf{X}}(\mathbf{y}|\theta)$

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

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.

Normal Minimal Sufficient Statistics (Example 6.2.14)

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

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

$$\begin{cases} \sum_{i=1}^n x_i^2 = \sum_{i=1}^n y_i^2 \\ \sum_{i=1}^n x_i = \sum_{i=1}^n y_i \end{cases}$$

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

Are Minimal Sufficient Statistics Unique?

- A short answer is "No"
- For example, $\left(\overline{X},\ s_{\mathbf{X}}^2 = \sum_{i=1}^n (X_i - \overline{X})^2/(n-1)\right)$ is also a minimal sufficient statistic for (μ, σ^2) in normal distribution.
- Important Facts
  1. If $T(\mathbf{X})$ is a minimal sufficient statistic for θ, then its one-to-one function is also a minimal sufficient statistic for θ.
  2. There is always a one-to-one function between any two minimal sufficient statistics. In other words, the partition created by a minimal sufficient statistic is unique

Proving the important facts

Theorem for Fact 1

If $T(\mathbf{X})$ is a minimal sufficient statistic for θ, then its one-to-one function is also a minimal sufficient statistic for θ.

Strategies for Proof

- Let $T^*(\mathbf{X}) = q(T(\mathbf{X}))$ and q is a one-to-one function. Then there exist a q^{-1} such that $T(\mathbf{X}) = q^{-1}(T^*(\mathbf{X}))$
- First is to prove that $T^*(\mathbf{x})$ is a sufficient statistic.
- Next, prove that $T^*(\mathbf{x})$ is also a minimal sufficient statistic.

Proof : $T^*(\mathbf{x})$ is a sufficient statistic

Because $T(\mathbf{X})$ is sufficient, by the Factorization Theorem, there exists h and g such that

$$\begin{aligned} f_{\mathbf{X}}(\mathbf{x}|\theta) &= g(T(\mathbf{x}|\theta))h(\mathbf{x}) \\ &= g(q^{-1}(T^*(\mathbf{x}|\theta)))h(\mathbf{x}) \\ &= (g \circ q^{-1})(T^*(\mathbf{x}|\theta))h(\mathbf{x}) \end{aligned}$$

Therefore, by the Factorization Theorem, T^* is also a sufficient statistic.

Proof : $T^*(\mathbf{x})$ is a minimal sufficient statistic

Because $T(\mathbf{X})$ is minimal sufficient, by definition, for any sufficient statistic $S(\mathbf{X})$, there exist a function w such that $T(\mathbf{X}) = w(S(\mathbf{x}))$.

$$\begin{aligned} T^*(\mathbf{x}) &= q(T(\mathbf{X})) \\ &= q(w(S(\mathbf{X}))) \\ &= (q \circ w)(S(\mathbf{X})) \end{aligned}$$

Thus, $T^*(\mathbf{X})$ is also a function of $S(\mathbf{X})$ always, and by definition, T^* is also a minimal sufficient statistic.

Proving the important facts

Theorem for Fact 2

There is always a one-to-one function between any two minimal sufficient statistics. (In other words, the partition created by minimal sufficient statistics is unique)

Examples

For normal statistics, let $T_1(\mathbf{X}) = (\sum X_i, \sum X_i^2)$ and $T_2(\mathbf{X}) = (\overline{X}, \sum(X_i - \overline{X})^2/(n-1))$. Then, there exists one-to-one functions such that

$$\begin{aligned} \sum X_i &= g_1\left(\overline{X}, \sum(X_i - \overline{X})^2/(n-1)\right) \\ \sum X_i^2 &= g_2\left(\overline{X}, \sum(X_i - \overline{X})^2/(n-1)\right) \\ \overline{X} &= h_1(\sum X_i, \sum X_i^2) \\ \sum(X_i - \overline{X})^2(n-1) &= h_2(\sum X_i, \sum X_i^2) \end{aligned}$$

Proof

Assume that both $T(\mathbf{X})$ and $T^*(\mathbf{X})$ are minimal sufficient. Then by the definition of minimal sufficient statistics, there exist $q(\cdot)$ and $r(\cdot)$ such that

$$\begin{aligned} T(\mathbf{X}) &= q(T^*(\mathbf{X})) \\ T^*(\mathbf{X}) &= r(T(\mathbf{X})) \end{aligned}$$

Therefore, $q = r^{-1}$ holds and they are one-to-one functions.

Summary

Today

- Recap of Factorization Theorem
- Minimal Sufficient Statistics
  - Theorem 6.2.13
  - Two sufficient statistics from binomial distribution
  - Uniform Distribution
  - Normal Distribution
  - Minimal Sufficient Statistics are not unique

Next Lecture

- Ancillary Statistics