Biostatistics 602 - Statistical Inference
Lecture 06
Basu’s Theorem

Hyun Min Kang

January 29th, 2013

Last Lecture

1. What is a complete statistic?
2. Why it is called as "complete statistic"?
3. Can the same statistic be both complete and incomplete statistics, depending on the parameter space?
4. What is the relationship between complete and sufficient statistics?
5. Is a minimal sufficient statistic always complete?

Complete Statistics

Definition

- Let $T = \{f_T(t|\theta), \theta \in \Omega\}$ be a family of pdfs or pmfs for a statistic $T(X)$.
- The family of probability distributions is called complete if $E[g(T)|\theta] = 0$ for all $\theta$ implies $\Pr[g(T) = 0|\theta] = 1$ for all $\theta$.
- In other words, $g(T) = 0$ almost surely.
- Equivalently, $T(X)$ is called a complete statistic.

Example - Poisson distribution

When parameter space is limited - NOT complete

- Suppose $T = \{f_T(t|\lambda) = \frac{\lambda^t e^{-\lambda}}{t!}\}$ for $t \in \{0, 1, 2, \cdots\}$. Let $\lambda \in \Omega = \{1, 2\}$. This family is NOT complete.

With full parameter space - complete

- $X_1, \cdots, X_n \overset{i.i.d.}{\sim} \text{Poisson}(\lambda), \lambda > 0$.
- $T(X) = \sum_{i=1}^{n} X_i$ is a complete statistic.

**Problem**

Let $X$ is a uniform random sample from $\{1, \cdots, \theta\}$ where $\theta \in \Omega = \mathbb{N}$. Is $T(X) = X$ a complete statistic?

**Solution**

Consider a function $g(T)$ such that $E[g(T)|\theta] = 0$ for all $\theta \in \mathbb{N}$. Note that $f_X(x) = \frac{1}{\theta} I(x \in \{1, \cdots, \theta\}) = \frac{1}{\theta} I_{\mathbb{N}_0}(x)$.

$$E[g(T)|\theta] = E[g(X)|\theta] = \sum_{x=1}^{\theta} \frac{1}{\theta} g(x) = \frac{1}{\theta} \sum_{x=1}^{\theta} g(x) = 0$$

$$\sum_{x=1}^{\theta} g(x) = 0$$

for all $\theta \in \mathbb{N}$, which implies

- if $\theta = 1$, $\sum_{x=1}^{\theta} g(x) = g(1) = 0$
- if $\theta = 2$, $\sum_{x=1}^{\theta} g(x) = g(1) + g(2) = g(2) = 0$.

Therefore, $g(x) = 0$ for all $x \in \mathbb{N}$, and $T(X) = X$ is a complete statistic for $\theta \in \Omega = \mathbb{N}$.

Is the previous example barely complete?

**Modified Problem**

Let $X$ is a uniform random sample from $\{1, \cdots, \theta\}$ where $\theta \in \Omega = \mathbb{N} - \{n\}$. Is $T(X) = X$ a complete statistic?

**Solution**

Define a nonzero $g(x)$ as follows

$$g(x) = \begin{cases} 1 & x = n \\ -1 & x = n + 1 \\ 0 & \text{otherwise} \end{cases}$$

$$E[g(T)|\theta] = \frac{1}{\theta} \sum_{x=1}^{\theta} g(x) = \begin{cases} 0 & \theta \neq n \\ \frac{1}{\theta} & \theta = n \end{cases}$$

Because $\Omega$ does not include $n$, $g(x) = 0$ for all $\theta \in \Omega = \mathbb{N} - \{n\}$, and $T(X) = X$ is not a complete statistic.

Last Lecture : Ancillary and Complete Statistics

**Problem**

- Let $X_1, \cdots, X_n \overset{i.i.d.}{\sim} \text{Uniform(}\theta, \theta + 1\text{)}, \ \theta \in \mathbb{R}$.
- Is $T(X) = (X(1), X(n))$ a complete statistic?

**A Simple Proof**

- We know that $R = X(n) - X(1)$ is an ancillary statistic, which do not depend on $\theta$.
- Define $g(T) = X(n) - X(1) - E(R)$. Note that $E(R)$ is constant to $\theta$.
- Then $E[g(T)|\theta] = E(R) - E(R) = 0$, so $T$ is not a complete statistic.
Useful Fact 1: Ancillary and Complete Statistics

**Fact**
For a statistic $T(\mathbf{X})$, if a non-constant function of $T$, say $r(T)$ is ancillary, then $T(\mathbf{X})$ cannot be complete.

**Proof**
Define $g(T) = r(T) - E[r(T)]$, which does not depend on the parameter $\theta$ because $r(T)$ is ancillary. Then $E[g(T)|\theta] = 0$ for a non-zero function $g(T)$, and $T(\mathbf{X})$ is not a complete statistic.

Theorem 6.2.28 - Lehman and Schefle (1950)

**The textbook version**
If a minimal sufficient statistic exists, then any complete statistic is also a minimal sufficient statistic.

**Paraphrased version**
Any complete, and sufficient statistic is also a minimal sufficient statistic.

**The converse is NOT true**
A minimal sufficient statistic is not necessarily complete. (Recall the example in the last lecture.)

Useful Fact 2: Arbitrary Function of Complete Statistics

**Fact**
If $T(\mathbf{X})$ is a complete statistic, then a function of $T$, say $T^* = r(T)$ is also complete.

**Proof**

\[
E[g(T^*)|\theta] = E[g \circ r(T)|\theta]
\]

Assume that $E[g(T^*)|\theta] = 0$ for all $\theta$, then $E[g \circ r(T)|\theta] = 0$ holds for all $\theta$ too. Because $T(\mathbf{X})$ is a complete statistic, $Pr[g \circ r(T) = 0] = 1$, $\forall \theta \in \Omega$. Therefore $Pr[g(T^*) = 0] = 1$, and $T^*$ is a complete statistic.
Proof of Basu’s Theorem

- As $S(X)$ is ancillary, by definition, it does not depend on $\theta$.
- As $T(X)$ is sufficient, by definition, $f_X(X | T(X))$ is independent of $\theta$.
- Because $S(X)$ is a function of $X$, $\Pr(S(X) | T(X))$ is also independent of $\theta$.
- We need to show that $\Pr(S(X) = s | T(X) = t) = \Pr(S(X) = s)$, $\forall t \in T$.

Proof of Basu’s Theorem (cont’d)

\[
\Pr(S(X) = s | \theta) = \sum_{t \in T} \Pr(S(X) = s | T(X) = t) \Pr(T(X) = t | \theta) \tag{1}
\]

\[
\Pr(S(X) = s | \theta) = \Pr(S(X) = s) \sum_{t \in T} \Pr(T(X) = t | \theta) \tag{2}
\]

\[
= \sum_{t \in T} \Pr(S(X) = s) \Pr(T(X) = t | \theta) \tag{3}
\]

Define $g(t) = \Pr(S(X) = s | T(X) = t) - \Pr(S(X) = s)$. Taking (1)-(3),
\[
\sum_{t \in T} [\Pr(S(X) = s | T(X) = t) - \Pr(S(X) = s)] \Pr(T(X) = t | \theta) = 0
\]

Therefore, $S(X)$ is independent of $T(X)$.

Application of Basu’s Theorem

Problem

- $X_1, \cdots, X_n \overset{i.i.d.}{\sim} Uniform(0, \theta)$.
- Calculate $E \left[ \frac{X_{(1)}}{X_{(n)}} \right]$ and $E \left[ \frac{X_{(1)} + X_{(2)}}{X_{(n)}} \right]$.

A strategy for the solution

- We know that $X_{(n)}$ is sufficient statistic.
- We know that $X_{(n)}$ is complete, too.
- We can easily show that $X_{(1)}/X_{(n)}$ is an ancillary statistic.
- Then we can leverage Basu’s Theorem for the calculation.

Showing that $X_{(1)}/X_{(n)}$ is Ancillary

\[
f_X(x | \theta) = \frac{1}{\theta} I(0 < x < \theta)
\]

Let $y = x/\theta$, then $|dx/dy| = \theta$, and $Y \sim Uniform(0, 1)$.

\[
f_Y(y | \theta) = I(0 < y < 1) \frac{X_{(1)}}{X_{(n)}} = \frac{Y_{(1)}}{Y_{(n)}}
\]

Because the distribution of $Y_1, \cdots, Y_n$ does not depend on $\theta$, $X_{(1)}/X_{(n)}$ is an ancillary statistic for $\theta$. 
Applying Basu’s Theorem

- By Basu’s Theorem, $X(1)/X(n)$ is independent of $X(n)$.
- If $X$ and $Y$ are independent, $E(XY) = E(X)E(Y)$.

\[
E[X(1)] = E \left[ \frac{X(1)}{X(n)} \right] = E \left[ \frac{X(1)}{X(n)} \right] E[X(n)]
\]

\[
E \left[ \frac{X(1)}{X(n)} \right] = \frac{E[X(1)]}{E[X(n)]}
= \frac{E[\theta Y(1)]}{E[\theta Y(n)]}
= \frac{E[Y(1)]}{E[Y(n)]}
\]

Obtaining $E[Y(1)]$

- $Y \sim$ Uniform(0, 1)
- $f_Y(y) = I(0 < y < 1)$
- $F_Y(y) = yI(0 < y < 1) + I(y \geq 1)$
- $f_Y(n) = \frac{n!}{(n-1)!} f_Y(y) [F_Y(y)]^{n-1} I(0 < y < 1)$
- $Y(n) \sim$ Beta($n$, 1)
- $E[Y(n)] = \frac{n}{n+1}$

Therefore, $E[\frac{X(1)}{X(n)}] = \frac{E[Y(1)]}{E[Y(n)]} = \frac{1}{n}$

Obtaining $E[Y(2)]$

- $Y \sim$ Uniform(0, 1)
- $f_Y(y) = I(0 < y < 1)$
- $F_Y(y) = yI(0 < y < 1) + I(y \geq 1)$
- $f_Y(n) = \frac{n!}{(n-2)!} [1 - F_Y(y)]^{n-2} f_Y(y) [F_Y(y)] I(0 < y < 1)$
- $Y(2) \sim$ Beta($2$, $n-1$)
- $E[Y(2)] = \frac{2}{n+1}$

Therefore, $E[\frac{X(1)+X(2)}{X(n)}] = \frac{E[Y(1)+Y(2)]}{E[Y(n)]} = \frac{E[Y(1)]+E[Y(2)]}{E[Y(n)]} = \frac{3}{n}$
Summary

Today
- More on complete statistics
- Basu’s Theorem

Next Lecture
- Exponential Family