Lecture 24: Completeness

Definition 6.2.16 (ancillary statistics)

A statistic $V(X)$ is ancillary iff its distribution does not depend on any unknown quantity. A statistic $V(X)$ is first-order ancillary iff $E[V(X)]$ does not depend on any unknown quantity.

A trivial ancillary statistic is $V(X) \equiv$ a constant.

The following examples show that there exist many nontrivial ancillary statistics (non-constant ancillary statistics).

Examples 6.2.18 and 6.2.19 (location-scale families)

If $X_1, \ldots, X_n$ is a random sample from a location family with location parameter $\mu \in \mathbb{R}$, then, for any pair $(i,j)$, $1 \leq i, j \leq n$, $X_i - X_j$ is ancillary, because $X_i - X_j = (X_i - \mu) - (X_j - \mu)$ and the distribution of $(X_i - \mu, X_j - \mu)$ does not depend on any unknown parameter.

Similarly, $X_{(i)} - X_{(j)}$ is ancillary, where $X_{(1)}, \ldots, X_{(n)}$ are the order statistics, and the sample variance $S^2$ is ancillary.
Note that we do not even need to obtain the form of the distribution of \( X_i - X_j \).

If \( X_1, \ldots, X_n \) is a random sample from a scale family with scale parameter \( \sigma > 0 \), then by the same argument we can show that, for any pair \((i, j)\), \(1 \leq i, j \leq n\), \( X_i / X_j \) and \( X(i) / X(j) \) are ancillary.

If \( X_1, \ldots, X_n \) is a random sample from a location-scale family with parameters \( \mu \in \mathbb{R} \) and \( \sigma > 0 \), then, for any \((i, j, k)\), \(1 \leq i, j, k \leq n\), 

\[
\frac{(X_i - X_k)}{(X_j - X_k)} \quad \text{and} \quad \frac{(X(i) - X(k))}{(X(j) - X(k))}
\]

are ancillary.

If \( V(X) \) is a non-trivial ancillary statistic, then the set \( \{x : V(x) = \nu\} \) does not contain any information about \( \theta \).

If \( T(X) \) is a statistic and \( V(T(X)) \) is a non-trivial ancillary statistic, it indicates that the reduced data set by \( T \) contains a non-trivial part that does not contain any information about \( \theta \) and, hence, a further simplification of \( T \) may still be needed.

A sufficient statistic \( T(X) \) appears to be most successful in reducing the data if no nonconstant function of \( T(X) \) is ancillary or even first-order ancillary, which leads to the following definition.
Definition 6.2.21 (completeness)

Let $X$ be a sample from a family indexed by $\theta \in \Theta$ (the parameter space) and let $E_\theta$ and $P_\theta$ be the expectation and probability, respectively, calculated with respect to a $\theta \in \Theta$.

- A statistic $T(X)$ is complete iff for any function $g$ not depending on $\theta$, $E_\theta[g(T)] = 0$ for all $\theta \in \Theta$ implies $P_\theta(g(T) = 0) = 1$ for all $\theta \in \Theta$.
- A statistic $T$ is boundedly complete iff the previous statement holds for any bounded $g$.
- The family of distributions corresponding to a statistic $T$ is complete (or boundedly complete) iff $T$ is complete (or boundedly complete).

- A complete statistic is boundedly complete.
- If $T$ is complete (or boundedly complete) and $S = \psi(T)$ for a measurable $\psi$, then $S$ is complete (or boundedly complete).
- It can be shown that a complete and sufficient statistic is minimal sufficient (Theorem 6.2.28).
- A minimal sufficient statistic is not necessarily complete.
Example 6.2.15.

In this example, $X_1, \ldots, X_n$ is a random sample from $\text{uniform}(\theta, \theta + 1)$, $\theta \in \mathbb{R}$, and we showed that $T = (X_1, X_n)$ is the minimal sufficient statistic for $\theta$.

We now show that $T$ is not complete.

Note that $V(T) = X_n - X_1 = (X_n - \theta) - (X_1 - \theta)$ is in fact ancillary. Its distribution can be obtained using the result in Example 5.4.7, but we do not need that for arguing that $T$ is not complete.

It is easy to see that $E_{\theta}(V)$ exists and it does not depend on $\theta$ since $V$ is ancillary.

Letting $c = E(V)$, we see that $E_{\theta}(V - c) = 0$ for all $\theta$.

Thus, we have a function $g(x, y) = x - y - c$ such that

$E_{\theta}[g(X_1, X_n)] = E_{\theta}(V - c) = 0$ for all $\theta$ but\n
$P_{\theta}(g(X_1, X_n) = 0) = P_{\theta}(V = c) \neq 0$.

This shows that $T$ is not complete.

If a minimal sufficient statistic $T$ is not complete, then

- there is a non-trivial first order ancillary statistic $V(T)$;
- there does not exist any complete statistic.
Example 6.2.22.

Let $T \sim \text{binomial}(n, \theta)$, $0 < \theta < 1$, (note that $T$ is a sufficient statistic based on a random sample of size $n$ from $\text{binomial}(1, \theta)$).

If $g$ is a function such that $E_\theta[g(T)] = 0$ for all $\theta$, then

\[
0 = E_\theta[g(T)] = \sum_{t=0}^{n} g(t) \binom{n}{t} \theta^t (1 - \theta)^{n-t}
\]

\[
= (1 - \theta)^n \sum_{t=0}^{n} g(t) \binom{n}{t} \left( \frac{\theta}{1 - \theta} \right)^t \quad \text{all } \theta \in (0, 1)
\]

Since the factor $(1 - \theta)^n \neq 0$, we must have

\[
0 = \sum_{t=0}^{n} g(t) \binom{n}{t} \varphi^t \quad \text{all } \varphi > 0, \quad \varphi = \frac{\theta}{1 - \theta}
\]

The last expression is a polynomial in $\varphi$ of degree $n$.

For this polynomial to be 0 for all $\varphi > 0$, it must be true that the coefficient of $\varphi^t$, which is $g(t) \binom{n}{t}$, is 0 for every $t$.

This shows that $g(t) = 0$ for $t = 0, 1, \ldots, n$ and hence $P_\theta(g(T) = 0) = 1$ for all $\theta$, i.e., $T$ is complete.
Example 6.2.23.

Let $X_1, \ldots, X_n$ be iid from $\text{uniform}(0, \theta)$, $\theta > 0$, with pdf

$$f_\theta(x) = \begin{cases} 1 & 0 < x_i < \theta, i = 1, \ldots, n \\ 0 & \text{otherwise} \end{cases} = \begin{cases} 1 & 0 < x(n) < \theta \\ 0 & \text{otherwise} \end{cases}$$

By the factorization theorem, $X(n)$ is sufficient for $\theta$.

Using the result for the order statistics in Chapter 5, $X(n)$ has pdf

$$f_{X(n)}(t) = \begin{cases} nt^{n-1}\theta^{-n} & 0 < t < \theta \\ 0 & \text{otherwise} \end{cases}$$

For any $g$,

$$E_\theta[g(X(n))] = n\theta^{-n} \int_0^{\theta} g(t)t^{n-1} dt = 0 \quad \text{all } \theta > 0$$

implies that

$$0 = \int_0^{\theta} g(t)t^{n-1} dt \quad \text{and} \quad 0 = \frac{d}{d\theta} \int_0^{\theta} g(t)t^{n-1} dt = g(\theta)\theta^{n-1} \quad \text{all } \theta > 0$$

Thus, $g(t) = 0$ for all $t > 0$, which means that $P_\theta(g(T) = 0) = 1$ for all $\theta$. Hence, $X(n)$ is also complete.
The relationship between an ancillary statistic and a complete and sufficient statistic is characterized in the following result.

**Theorem 6.2.24 (Basu’s theorem)**

Let $V$ and $T$ be two statistics of $X$ from a population indexed by $\theta \in \Theta$. If $V$ is ancillary and $T$ is boundedly complete and sufficient for $\theta$, then $V$ and $T$ are independent with respect to $P_\theta$ for any $\theta \in \Theta$.

**Proof.**

Let $B$ be an event on the range of $V$ and $A$ an event on the range of $T$. From the 3rd definition of the independence of random variables, we only need to show that

$$P_\theta(T^{-1}(A) \cap V^{-1}(B)) = P_\theta(T^{-1}(A))P_\theta(V^{-1}(B)), \quad \theta \in \Theta$$

Since $V$ is ancillary, $P_\theta(V^{-1}(B)) = P_B$ does not depend on $\theta$. As $T$ is sufficient, $E_\theta[I_B(V)|T] = h_B(T)$ is a function of $T$ (not depending on $\theta$), where $I_B(V)$ is the indicator function of $\{V \in B\}$.

Since

$$E_\theta[h_B(T)] = E_\theta\{E[I_B(V)|T]\} = E_\theta\{I_B(V)\} = P_\theta(V^{-1}(B)) = P_B \quad \theta \in \Theta,$$
by the bounded completeness of $T$,

$$P_\theta(h_B(T) = P_B) = 1 \quad \theta \in \Theta$$

Then the result follows from

$$P_\theta(T^{-1}(A) \cap V^{-1}(B)) = E_\theta\{ E_\theta[I_A(T)I_B(V)|T] \}$$

$$= E_\theta\{ I_A(T)E_\theta[I_B(V)|T] \}$$

$$= E_\theta\{ I_A(T)P_B \} = P_BE_\theta\{ I_A(T) \}$$

$$= P_\theta(T^{-1}(A))P(V^{-1}(B)) \quad \theta \in \Theta$$

- If a minimal sufficient statistic $T$ is not complete, then there may be an ancillary statistic $V$ such that $V$ and $T$ are not independent. An example is in Example 6.2.15, $T = (X_1, X_n)$ is minimal sufficient but not complete, and $T$ and the ancillary statistic $V = X_n - X_1$ is not independent.

- Basu’s theorem is useful in proving the independence of two statistics.

We first state without proof the following useful result.
Theorem 6.2.25 (complete statistics in exponential families)

Let \( X = (X_1, \ldots, X_n) \) be a random sample with pdf

\[
f_{\theta}(x) = h(x) c(\theta) \exp \left( \sum_{j=1}^{k} w_j(\theta) t_j(x) \right) \quad \theta \in \Theta,
\]

Then the statistic

\[
T(X) = \left( \sum_{i=1}^{n} t_1(X_i), \ldots, \sum_{i=1}^{n} t_k(X_i) \right)
\]

is complete as long as \( \Theta \) contains an open set in \( \mathbb{R}^k \) (i.e., the family of distributions is of full rank).

- Note that \( T \) is also sufficient for \( \theta \) (without requiring any condition on \( \Theta \)).
- Compared with the result of minimal sufficient statistics in curved exponential families, the condition on \( \Theta \) in this theorem is stronger.

We illustrate the application of Basu’s theorem and Theorem 6.2.25 in the normal distribution family.
Example (the normal family).

Suppose that \( X_1, \ldots, X_n \) are iid from \( N(\mu, \sigma^2) \), \( \mu \in \mathbb{R}, \sigma > 0, \theta = (\mu, \sigma^2) \). The joint pdf of \( X_1, \ldots, X_n \) is

\[
(2\pi)^{-n/2} \exp \left\{ \eta_1 T_1 + \eta_2 T_2 - n \zeta(\eta) \right\},
\]

where \( T_1 = \sum_{i=1}^n X_i, T_2 = -\sum_{i=1}^n X_i^2, \eta = (\eta_1, \eta_2) = \left( \frac{\mu}{\sigma^2}, \frac{1}{2\sigma^2} \right), \) and \( \zeta(\eta) \) is a function of \( \eta \).

This is an exponential family of full rank, since \( \mathbb{R} \times (0, \infty) \) is open.

By Theorem 6.2.25, \( T(X) = (T_1, T_2) \) is complete and sufficient for \( \theta \).

It can be shown that any one-to-one function of a complete and sufficient statistic is also complete and sufficient.

Thus, \( (\bar{X}, S^2) \) is complete and sufficient for \( \theta \).

We now apply Basu’s theorem to show that \( \bar{X} \) and \( S^2 \) are independent for any \( \theta \).

For this purpose, we consider a sub-family with unknown \( \mu \in \mathbb{R} \) and a known (fixed) \( \sigma^2 > 0 \).

Note that we only need to show that \( \bar{X} \) and \( S^2 \) are independent for every fixed \( (\mu, \sigma^2) \).
If $X_1, \ldots, X_n$ are iid from $N(\mu, \sigma^2)$ with $\mu \in \mathbb{R}$ and a known $\sigma > 0$, then it can be easily shown that the family is an exponential family of full rank with parameter $\mu \in \mathbb{R}$.

By Theorem 6.2.25, $\bar{X}$ is complete and sufficient for $\mu$.

Since $S^2 = (n-1)^{-1} \sum_{i=1}^{n} (Z_i - \bar{Z})^2$, where $Z_i = X_i - \mu$ is $N(0, \sigma^2)$ and $\bar{Z} = n^{-1} \sum_{i=1}^{n} Z_i$, $S^2$ is an ancillary statistic ($\sigma^2$ is known).

By Basu’s theorem, $\bar{X}$ and $S^2$ are independent with respect to $N(\mu, \sigma^2)$ for any $\mu \in \mathbb{R}$ and $\sigma^2 > 0$.

Note that this proof is simpler than the proof we gave in Chapter 5.

**Example**

Let $X_1, \ldots, X_n$ be a random sample from the family of pdf’s of the form

$$f_\theta(x) = C(\theta_1, \ldots, \theta_n) \exp\{-x^{2n} + \theta_1 x + \theta_2 x^2 + \cdots + \theta_n x^n\},$$

where $\theta_j \in \mathbb{R}$, $\theta = (\theta_1, \ldots, \theta_n)$, and $C(\theta_1, \ldots, \theta_n)$ is a normalizing constant such that $\int f(x)dx = 1$.

This family of pdf’s is an exponential family of full rank.

By Theorem 6.2.25, $U = (U_1, \ldots, U_n)$ is a complete and sufficient
Statistic for $\theta$, where $U_j = \sum_{i=1}^n X_i^j$, $j = 1, \ldots, n$.

We want to show that $T(X) = (X_{(1)}, \ldots, X_{(n)})$, the vector of order statistics, is also complete and sufficient for $\theta$.

The result follows if we can show that there is a one-to-one correspondence between $T(X)$ and $U(X)$.

Let $V_1 = \sum_{i=1}^n X_i$, $V_2 = \sum_{i<j} X_i X_j$, $V_3 = \sum_{i<j<k} X_i X_j X_k$, $\ldots$, $V_n = X_1 \cdots X_n$.

From the identities

$$U_k - V_1 U_{k-1} + V_2 U_{k-2} - \cdots + (-1)^{k-1} V_{k-1} U_1 + (-1)^k k V_k = 0,$$

$k = 1, \ldots, n$, there is a one-to-one correspondence between $U(X)$ and $V(X) = (V_1, \ldots, V_n)$.

From the identity

$$(t - X_1) \cdots (t - X_n) = t^n - V_1 t^{n-1} + V_2 t^{n-2} - \cdots + (-1)^n V_n,$$

there is a one-to-one correspondence between $V(X)$ and $T(X)$.

Hence, $T(X)$ is sufficient and complete for $\theta$.

In fact, $V(X)$ is also sufficient and complete for $\theta$. 
The relationship between minimal sufficiency and sufficiency with completeness is given by the following theorem.

### Theorem 6.2.28 (modified).

Suppose that $S$ is a sufficient statistic and $T$ is a complete and sufficient statistic. Then $T$ must be minimal sufficient and $S$ must be complete.

### Proof.

Since $S$ is minimal sufficient and $T$ is sufficient, there exists a Borel function $h$ such that $S = h(T)$.

Since $h$ cannot be a constant function and $T$ is complete, we conclude that $S$ is complete.

Consider $T - E(T|S) = T - E[T|h(T)]$, which is a Borel function of $T$ and hence can be denoted as $g(T)$.

Note that $E[g(T)] = 0$.

By the completeness of $T$, $g(T) = 0$ a.s., that is, $T = E(T|S)$ a.s. This means that $T$ is also a function of $S$ and, therefore, $T$ is minimal sufficient.
Example 6.2.20 (ancillary precision)

Let $X_1$ and $X_2$ be iid from the discrete uniform distribution on three points $\{\theta, \theta + 1, \theta + 2\}$, where $\theta \in \Theta = \{0, \pm 1, \pm 2, \ldots\}$.

Using the same argument as in Example 6.2.15, we can show that the order statistics $(X_{(1)}, X_{(2)})$ is minimal sufficient for $\theta$.

Let $M = (X_{(1)} + X_{(2)})/2$ and $R = X_{(2)} - X_{(1)}$ (mid-range and range).

Since $(M, R)$ is a one-to-one function of $(X_{(1)}, X_{(2)})$, it is also minimal sufficient for $\theta$.

Consider the estimation of $\theta$ using $(M, R)$.

Note that $R = (X_{(2)} - \theta) - (X_{(1)} - \theta)$ is the range of the two order statistics from the uniform distribution on $\{0, 1, 2\}$ and, hence the distribution of $R$ does not depend on $\theta$, i.e., $R$ is ancillary.

One may think $R$ is useless in the estimation of $\theta$ and only $M$ is useful.

Suppose we observe $(M, R) = (m, r)$ and $m$ is an integer.

From the observation $m$, we know that $\theta$ can only be one of the 3 values $m, m - 1, \text{and } m - 2$; however, we are not certain which of the 3 values is $\theta$. 
We can know more if \( r = 2 \), which must be the case that \( X(1) = m - 1 \) and \( X(2) = m + 1 \).

With this additional information, the only possible value for \( \theta \) is \( m - 1 \).

When \( m \) is an integer, \( r \) cannot be 1. If \( r = 0 \), then we know that \( X_1 = X_2 \) and we are not certain which of the 3 values is \( \theta \).

The knowledge of the value of the ancillary statistic \( R \) increases our knowledge about \( \theta \), although \( R \) alone gives us no information about \( \theta \).

**What we learn from the previous example?**

- An ancillary statistic that is a function of a minimal sufficient statistic \( T \) may still be useful for our knowledge about \( \theta \). (Note that the ancillary statistic is still a function of \( T \).)
- This cannot occur to a sufficient and complete statistic \( T \), since, if \( V(T) \) is ancillary, then by the completeness of \( T \), \( V \) must be a constant and is useless.
- Therefore, the sufficiency and completeness together is a much desirable (and strong) property.