February 4, 2016

1 Complete Statistics

Suppose $X \sim P_{\theta}, \theta \in \Theta$. Let $(X_{(1)}, \ldots, X_{(n)})$ denote the order statistics.

Definition 1. A statistic T = T(X) is complete if

$$E_{\theta}g(T) = 0$$
 for all θ

implies

$$P_{\theta}(g(T)=0) = 1$$
 for all θ .

(Note: E_{θ} denotes expectation computed with respect to P_{θ}).

Example: $X = (X_1, \ldots, X_n)$ iid $N(\theta, 1)$. $T(X) = (X_1, X_2)$ is a statistic which is <u>not</u> complete because

$$E(X_1 - X_2) = 0 \quad \text{for all } \theta.$$

and $X_1 - X_2$ here is a function of T where $g : \mathbb{R}^2 \to \mathbb{R}$ given by $(x_1, x_2) \mapsto x_1 - x_2$ but $P(X_1 - X_2 = 0) \neq 1$ for all θ .

More formally: T is <u>not</u> complete because the function g satisfies

$$E_{\theta}g(T) = 0 \quad \text{for all} \quad \theta \quad \text{but}$$
$$P_{\theta}(g(T) = 0) \neq 1 \quad \text{for all} \quad \theta.$$

Example: $X = (X_1, \ldots, X_n)$ iid $\text{Unif}(\theta, \theta + 1)$. $T = T(X) = (X_{(1)}, X_{(n)})$ is MSS. But T is <u>not</u> complete. We know that $S(X) = X_{(n)} - X_{(1)}$ is ancillary. Thus

$$E(X_{(n)} - X_{(1)}) = c$$

where c does not depend on θ and therefore,

$$E(X_{(n)} - X_{(1)} - c) = 0 \quad \text{for all} \quad \theta$$

but clearly,

$$P(X_{(n)} - X_{(1)} - c = 0) \neq 1$$
 for all θ .

Hence taking $g(T) = X_{(n)} - X_{(1)} - c$, we can see that T is not complete. Example: $X = X_1, \ldots, X_n$ iid Unif $(0, \theta)$ $T = T(X) = X_{(n)}$ is MSS. Fact: T is also complete. *Proof.* Assume there exists g such that

$$E_{\theta}g(T) = 0$$
 for all θ .

T has cdf $H(t)=\big(t/\theta)^n, 0\leq t\leq \theta$ and pdf $nt^{n-1}/\theta^n, 0\leq t\leq \theta.$ Then

$$Eg(T) = \int_0^\theta g(t) \frac{nt^{n-1}}{\theta^n} dt = 0 \quad \text{for all} \quad \theta > 0.$$

implies

$$\int_0^\theta g(t)nt^{n-1}dt = 0 \quad \text{for all} \quad \theta > 0.$$

implies (by differentiating both sides and using the Fundamental Theorem of Calculus)

$$g(\theta)n\theta^{n-1}dt = 0$$
 for all $\theta > 0$.

implies

$$g(t) = 0$$
 for all $t > 0$.

implies

$$P_{\theta}(g(T) = 0) = 1$$
 for all $\theta > 0$.

Theorem 1. Suppose X_1, \ldots, X_n iid with pdf (pmf)

$$f(x \mid \theta) = c(\theta)h(x) \exp\left\{\sum_{j=1}^{k} w_j(\theta)t_j(x)\right\}$$

for $\theta = (\theta_1, \dots, \theta_k) \in \Theta$. Let $X = (X_1, \dots, X_n)$. Define

$$T(X) = \left(\sum_{i=1}^{n} t_1(X_i), \sum_{i=1}^{n} t_2(X_i), \dots, \sum_{i=1}^{n} t_k(X_i)\right)$$

Then

- 1. T(X) is sufficient statistic for θ .
- 2. If Θ contains an open set in \mathbb{R}^k , then T(X) is complete. (More precisely, if

$$\{(w_1(\theta), w_2(\theta), \dots, w_k(\theta)) : \theta \in \Theta\}$$

contains an open set in \mathbb{R}^k , then T(X) is complete.)

Remarks:

- 1. The statistic T(X) in the Theorem is called the natural sufficient statistic.
- 2. $\eta = (\eta_1, \ldots, \eta_k) \equiv (w_1(\theta), \ldots, w_k(\theta))$ is called the natural parameter of the exponential family.
- 3. Condition (2) is the "open set condition" (OSC). The OSC is easily verified by inspection. Let $A \subset \mathbb{R}^k$. A contains an open set in \mathbb{R}^k iff A contains a k-dimensional ball. That is, there exists $x \in \mathbb{R}^k$ and r > 0 such that $B(x,r) \subset A$. Here B(x,r)denotes the ball of radius r about x. Let $A \subset \mathbb{R}$ (take k = 1). A contains an open set in \mathbb{R} if and only if A contains an interval. That is, there exists c < d such that $(c,d) \subset A$.

Facts:

- 1. Under weak conditions (which are almost always true, a complete sufficient statistic is also minimal. Abbreviation: $CSS \Rightarrow MSS$. (but MSS does not imply CSS as we saw earlier).
- 2. A one-to-one function of a CSS is also a CSS (See later remarks). Reminder: A 1-1 function of an MSS is also an MSS.

Example: The Bernoulli pmf is an exponential family (1pef):

$$p(x \mid \theta) = \theta^{x} (1 - \theta)^{1 - x}, x \in \{0, 1\} = (1 - \theta) I(x \in \{0, 1\}) \exp\left[x \log\left(\frac{\theta}{1 - \theta}\right)\right].$$

If X_1, \ldots, X_n are iid $p(x \mid \theta)$, then $T = \sum_i X_i$ is a SS. It is also complete if Θ contains an interval. Here is a direct proof of completeness (not relying on our general theorem on exponential families).

Proof. We know $T \sim \text{Binomial}(n, \theta)$. If Eg(T) = 0 for all $\theta \in \Theta \subset (0, 1)$, then

$$\sum_{k=0}^{n} g(k) \binom{n}{k} \theta^{k} (1-\theta)^{n-k} = 0$$

for all $\theta \in \Theta$. Then

$$(1-\theta)^n \sum_{k=0}^n g(k) \binom{n}{k} \left(\frac{\theta}{1-\theta}\right)^k = 0$$

for all $\theta \in \Theta$. Then

$$\sum_{k=0}^{n} a_k u^k = 0$$

for all $\theta \in \Theta$ where $a_k = g(k) \binom{n}{k}$ and $u = \theta/(1-\theta)$. This if Θ contains an interval, then the above implies that the polynomial $\psi(u) = \sum_{k=1}^n a_k u^k$ is identically zero in some interval. This implies all the coefficients a_k must be zero, which further implies g(k) = 0for $k = 0, 1, \ldots, n$ so that $P_{\theta}(g(T) = 0) = 1$ for all $\theta \in \Theta$.

In a homework exercise, you show that $T = \sum_{i=1}^{n} X_i$ is complete when X_1, \ldots, X_n are iid Poisson(λ) in a similar way, using the fact that an infinite power series (an analytic function) is identically zero in some interval if and only if all the coefficients are zero. Example: The N(θ , 1) family is a 1pef with $w(\theta) = \theta$, t(x) = x. Let $X = (X_1, \ldots, X_n)$ iid N(θ , 1). $T(X) = \sum_{i=1}^{n} X_i$ is the natural SS (It is sufficient for any Θ). Is T complete? This depends on Θ .

- 1. $\Theta = \mathbb{R}$. Yes. (OSC holds)
- 2. $\Theta = [0.01, 0.02]$. Yes. (OSC holds)
- 3. $\Theta = (1, 2) \cup \{4, 7\}$. Yes. (OSC holds)
- 4. $\Theta = \mathbb{Z}$ (the integers). OSC fails so Theorem says nothing. But can show that it is <u>not</u> complete.
- 5. $\Theta = \mathbb{Q}$ (the rationals). OSC fails so Theorem says nothing. Yes or no? Don't know.
- 6. Θ = Cantor set. OSC fails so Theorem says nothing. Yes or no? Don't know.
- 7. Θ = finite set. OSC fails so Theorem says nothing. But can show that it is <u>not</u> complete.

<u>Remark:</u> In general, it is typically true that if Θ is finite and the support of T = T(X) is infinite, then T is <u>not</u> complete.

Example: The N(μ, σ^2) family with $\theta = (\mu, \sigma^2)$ is a 2pef with

$$w(\theta) = \left(\frac{\mu}{\sigma^2}, \frac{-1}{2\sigma^2}\right), \quad t(x) = (x, x^2).$$

Let $X = (X_1, \ldots, X_n)$ iid $N(\mu, \sigma^2)$. Then $T(X) = (\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2)$ is the natural SS. (It is a SS for any Θ). T(X) is a one-to-one function of $U(X) = (\bar{x}, s^2)$. So T is CSS iff U is CSS. Is T (or U) complete? That depends on Θ .

- 1. $\Theta_1 = \{(\mu, \sigma^2) : \sigma^2 > 0\}$. OSC holds. Yes, complete.
- 2. $\Theta_2 = \{(\mu, \sigma^2) : \sigma^2 = \sigma_0^2\}$. OSC fails. Theorem says nothing. But we can prove that U is not complete.

Proof. Let $g(x_1, x_2) = x_2 - \sigma_0^2$. Then $Eg(U) = E(s^2 - \sigma_0^2) = \sigma^2 - \sigma_0^2 = 0$ for all $\theta \in \Theta_2$.

3. $\Theta_3 = \{(\mu, \sigma^2) : \mu = \mu_0, \sigma^2 > 0\}$. OSC fails. Theorem says nothing. But we can prove that U is not complete.

Proof. Let $g(x_1, x_2) = x_1 - \mu_0$. Then $Eg(U) = E(\bar{x} - \mu_0) = \mu - \mu_0 = 0$ for all $\theta \in \Theta_3$.

4. $\Theta_4 = \{(\mu, \sigma^2) : \mu = \sigma^2, \sigma^2 > 0\}$. OSC fails. Theorem says nothing. But we can prove that U is not complete.

Proof. Let $g(x_1, x_2) = x_1 - x_2$. Then $Eg(U) = E(\bar{x} - s^2) = \mu - \sigma^2 = 0$ for all $\theta \in \Theta_4$.

(<u>Note</u>: It is more natural to describe the families $\Theta_2, \Theta_3, \Theta_4$ as 1pef's. If you do this, you get different natural sufficient statistics, which turn out to be complete.)

5. $\Theta_5 = \{(\mu, \sigma^2) : \mu^2 = \sigma^2, \sigma^2 > 0\}$. OSC fails. Theorem says nothing. But we can prove that U is not complete.

Proof. Homework

- 6. $\Theta_6 = [1,3] \times [4,6]$. OSC holds. Yes, complete.
- 7. $\Theta_7 = \Theta_6 \cup \{(5,1), (4,2)\}$. OSC holds. Yes, complete.
- 8. Θ_8 = complicated wavy curve. OSC fails. Theorem says nothing. But hard to conclude anything.

Corollary 1. Suppose $X \in \mathbb{R}^m$ has joint pdf (pmf)

$$f(x \mid \theta) = c(\theta)h(x) \exp\left\{\sum_{j=1}^{k} w_j(\theta)t_j(x)\right\}$$

for all $x \in \mathbb{R}^m$ where $\theta = (\theta_1, \dots, \theta_k) \in \Theta$. Define $T(X) = (t_1(X), t_2(X), \dots, t_k(X))$. Then

- 1. T(X) is sufficient statistic for θ .
- 2. If Θ contains an open set in \mathbb{R}^k , then T(X) is complete.

(More precisely, if

$$\{(w_1(\theta), w_2(\theta), \dots, w_k(\theta)) : \theta \in \Theta\}$$

contains an open set in \mathbb{R}^k , then T(X) is complete.)

Example: Return to Simple Linear Regression: X_1, \ldots, X_n independent with $X_i \sim N(\beta_0 + \beta_1 z_i, \sigma^2)$. $\theta = (\beta_0, \beta_1, \sigma^2)$ and $\Theta = \mathbb{R}^2 \times (0, \infty)$. Recall that the joint density of $X = (X_1, \ldots, X_n)$ is

$$f(x \mid \theta) = (2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2} \left(\sum_i x_i^2 - 2\beta_0 \sum_{i=1}^n x_i - 2\beta_1 \sum_{i=1}^n z_i x_i + \sum_{i=1}^n (\beta_0 + \beta_1 z_i)^2\right)\right\}$$

= $\left[(2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2} \sum_i (\beta_0 + \beta_1 z_i)^2\right\}\right] \times 1$
 $\times \exp\left\{\frac{-1}{2\sigma^2} \sum_i x_i^2 + \frac{\beta_0}{\sigma^2} \sum_i x_i + \frac{\beta_1}{\sigma^2} \sum_i z_i x_i\right\}$
= $c(\theta)h(x) \exp\left\{\sum_{j=1}^3 w_j(\theta)t_j(x)\right\}$

where

$$w(\theta) = (w_1(\theta), w_2(\theta), w_3(\theta)) = \left(\frac{-1}{2\sigma^2}, \frac{\beta_0}{\sigma^2}, \frac{\beta_1}{\sigma^2}\right),$$

$$t(x) = (t_1(x), t_2(x), t_3(x)) = \left(\sum_{i=1}^n x_i^2, \sum_{i=1}^n x_i, \sum_{i=1}^n z_i x_i\right).$$

The data vector X may be regarded as a single observation from an n-dimensional 3pef. Since $\Theta \subset \mathbb{R}^3$ satisfied the OSC, the statistic $T(X) = t(X) = (\sum_{i=1}^n X_i^2, \sum_{i=1}^n X_i, \sum_{i=1}^n z_i X_i)$ is a CSS.

<u>Notation</u>: $X \sim P_{\theta}, \theta \in \Theta.S(X) = \psi(T(X))$ for some ψ . $\Theta_1 \subset \Theta_2 \subset \Theta$. Sufficiency:

- 1. If S(X) is sufficient, then T(X) is sufficient.
- 2. If T(X) is sufficient for Θ_2 , then T(X) is sufficient for Θ_1 .

Completeness:

- 1. If T(X) is complete, then S(X) is complete.
- 2. If T(X) is complete for Θ_1 , then T(X) is complete for Θ_2 (under mild regularity conditions).

Proof. Proof of 1. $E_{\theta}g(S(X)) = 0$ for all $\theta \in \Theta \implies E_{\theta}g(\psi(T(X))) = 0$ for all $\theta \in \Theta \implies P_{\theta}\{g(\psi(T(X))) = 0\} = 1$ for all $\theta \in \Theta$ (by completeness of T(X)) which implies $P_{\theta}\{g(S(X)) = 0\} = 1$ for all θ . Proof of 2. $E_{\theta}g(T(X)) = 0$ for all $\theta \in \Theta_2 \implies E_{\theta}g(T(X)) = 0$ for all $\theta \in \Theta_1 \implies P_{\theta}\{g(T(X)) = 0\} = 1$ for all $\theta \in \Theta_1$ (by completeness of T(X) for Θ_1) which implies $P_{\theta}\{g(T(X)) = 0\} = 1$ for all $\theta \in \Theta_2$ (under mild assumptions).

Ancillarity:

- 1. If T(X) is ancillary, then S(X) is ancillary.
- 2. If T(X) is ancillary for Θ_2 , then T(X) is ancillary for Θ_1 .

Proof. <u>Proof of 1.</u> Uses $Y \stackrel{d}{=} Z \implies \psi(Y) \stackrel{d}{=} \psi(Z)$. <u>Proof of 2.</u> Trivial

2 Basu's results

Suppose $X \sim P_{\theta}, \theta \in \Theta$.

Lemma 1. (Basu's Lemma) If T(X) is complete and sufficient (for $\theta \in \Theta$), and S(X) is ancillary, then S(X) and T(X) are independent for all $\theta \in \Theta$.

In other words, a complete sufficient statistic is independent of any ancillary statistic.

2.1 Remarks:

Let S = S(X), T = T(X). Let E_{θ} denote expectation w.r.t. P_{θ} .

1. The joint distribution of (S, T) depends on θ , so in general it is possible for S and T to be independent for some values of θ , but not for others. (Basu's Lemma says this does not happen in this case.)

- 2. For any rv's Y and Z, we know that E(Y | Z) = g(Z), i.e., the conditional expectation is a function of Z. If the joint distribution of (Y, Z) depends on a parameter θ , then $E_{\theta}(Y | Z) = g(Z, \theta)$, i.e., the conditional expectation is a function of both Z and θ . (However, this function may turn out to be constant in one or both variables.)
- 3. In general, $E(Y) = E\{E(Y \mid Z)\}$ and $E_{\theta}(Y) = E_{\theta}E_{\theta}(Y \mid Z)$.
- 4. To show that Y and Z are independent, it suffices to show that $\mathcal{L}(Y \mid Z) = \mathcal{L}(Y)$ which means that $P(Y \in A \mid Z) = P(Y \in A)$ for all (Borel) sets A. Let $w(Y) = I(Y \in A)$. Then $P(Y \in A) = Ew(Y)$ and $P(Y \in A \mid Z) = E(w(Y) \mid Z)$. The indicator function w(Y) is a bounded (Borel measurable) function. Therefore we have:

To show that Y and Z are independent, it suffices to show E(w(Y) | Z) = Ew(Y) for all bounded (Borel measurable) functions.

5. Thus, to show that S and T are independent for all θ , it suffices to show that $E_{\theta}(w(S) \mid T) = E_{\theta}w(S)$ for all θ and all bounded measurable functions w(S).

2.2 Proof of Basu's Lemma

Proof. Let w(S) be a given bounded function of S. Consider both sides of the identity:

$$E_{\theta}w(S) = E_{\theta}[E_{\theta}(w(S) \mid T)]$$

for all θ . Consider the LHS. Since S is ancillary, the distribution of w(S) is the same for all θ so that the LHS is constant in θ . Call this constant c.

Now consider the RHS. We know that $E_{\theta}(w(S) \mid T)$ will be some function of θ and T. However, since T is a sufficient statistic, $\mathcal{L}(X \mid T)$ does not depend on θ . Since S = S(X), this implies $\mathcal{L}(S \mid T)$ does not depend on θ so that in turn $\mathcal{L}(w(S) \mid T)$ does not depend on θ . Thus, by sufficiency, $E_{\theta}(w(S) \mid T)$ is constant in θ and must be a function of T only. Call this function $\psi(T)$. That is,

$$\psi(T) = E_{\theta}(w(S) \mid T).$$

The original identity can now be written as

$$c = E_{\theta}\psi(T) \text{ for all } \theta \text{ or equivalently} \\ 0 = E_{\theta}(\psi(T) - c) \text{ for all } \theta.$$

Since T is complete, this implies

$$P(\psi(T) - c = 0) = 1 \quad \text{for all} \quad \theta \quad \text{or equivalently}$$

$$\psi(T) = c \quad \text{with probability one for all} \quad \theta.$$

This means

$$E_{\theta}(w(S) \mid T) = E_{\theta}w(S)$$
 with probability one for all θ .

Since w(S) is an arbitrary bounded function, by the earlier discussion this implies S and T are independent for all θ .

2.3 Applications of Basu's theorem

1. <u>Example</u>: Let $X = (X_1, ..., X_n)$ iid Unif $(0, \theta)$. Recall: $T(X) = X_{(n)} = \max\{X_i\}$ is a CSS. Unif $(0, \theta)$ is a scale family, so any scale invariant statistic S(X) is ancillary. Thus, by Basus lemma, all of the following are independent of $X_{(n)}$ for all θ :

$$\frac{\bar{x}}{s}, \quad \frac{X_{(1)}}{X_{(n)}}, \quad \left(\frac{X_{(1)}}{X_{(n)}}, \dots, \frac{X_{(n-1)}}{X_{(n)}}\right)$$

- 2. Example: (Using Basu's Lemma to obtain an indirect proof of completeness.) Let $\overline{X} = (X_1, \ldots, X_n)$ iid Unif $(\theta, \theta + 1)$. Recall: $T(X) = (X_{(1)}, X_{(n)})$ is a MSS. $S(X) = X_{(n)} X_{(1)}$ is ancillary. Since S is a function of T, the rvs S and T cannot be independent. Thus T cannot be complete (for then we would get a contradiction with Basus Lemma).
- 3. <u>Example</u>: Let $X = (X_1, \ldots, X_n)$ be iid $N(\mu, \sigma^2)$. Let \bar{x}, s^2, Z be the sample mean, sample variance, and standardized residuals (z-scores) of the data X (Recall: $Z = (Z_1, \ldots, Z_n)$ with $Z_i = (X_i - \bar{x})/s$. Fact: \bar{x}, s^2, Z are mutually independent.

Proof. We first show that the pair (\bar{x}, s^2) is independent of Z, and then show that \bar{x} and s^2 are independent. Each stage uses Basu's Lemma.

<u>Stage 1:</u> (\bar{x}, s^2) is independent of Z: Consider the family of all $N(\mu, \sigma^2)$ distributions (with both parameters allowed to vary). Recall: (\bar{x}, s^2) is a CSS. This is a location-scale family so that any location-scale invariant statistic is ancillary. Z is location-scale invariant. Thus, Z is ancillary so that (by Basu's Lemma) it must be independent of (\bar{x}, s^2) for all (μ, σ^2) .

Stage 2: \bar{x} and s^2 are independent: Fix σ^2 at an arbitrary value σ_0^2 and consider the family of $N(\mu, \sigma_0^2)$ distributions, μ unknown. Recall: This is a 1pef and the natural SS $\sum_i X_i$ is a CSS. \bar{x} is a 1-1 function of this and so also a CSS. This is a location family so that any location invariant statistic is ancillary. s^2 is location invariant. Thus, s^2 is ancillary so (by Basu's Lemma) it must be independent of \bar{x} for all μ (and also for all σ^2 since σ_0^2 is arbitrary).

4. <u>Example</u>: (an amusing calculation via Basu) Let $X = (X_1, \ldots, X_n)$ iid $N(0, \sigma^2), \sigma^2 > 0$.

<u>Goal:</u> Compute ES where

$$S = \frac{(\sum_i X_i)^2}{\sum_i X_i^2}$$

This is a 1pef with $\theta = \sigma^2$, $t(x) = x^2$ and $w(\theta) = -1/(2\sigma^2)$. Therefore $T(X) = \sum_i X_i^2$ is CSS. This is also a scale family so that scale invariant statistics are ancillary. S is scale invariant implies S ancillary which implies (by Basu) S independent of T. Thus E(ST) = (ES)(ET) which implies ES = E(ST)/ET which becomes

$$ES = \frac{E(\sum_{i} X_{i})^{2}}{E\sum_{i} X_{i}^{2}} = \frac{n\sigma^{2}}{n\sigma^{2}} = 1.$$