208711 Statistical Theory 1 **Fall 2019**

Lecture 19: October 22

Lecturer: Donlapark Pornnopparath

The following definition, although seems quite unintuitive, is an important building block in the proofs of celebrated Basu's Theorem and Lehmann-Scheffé Theorem:

Definition 19.1. A statistic T is *complete* for a family of distribution with pdf $f(t|\theta)$ parametrized by $\theta \in \Theta$ if for any measurable function g

$$\mathbb{E}_\theta[g(T)] = 0 \quad \text{for all} \quad \theta \quad \text{implies} \quad g(T) = 0 \quad \text{a.e. for all } \theta. \tag{19.1}$$

The condition (19.1) can be written as

$$\int g(t)f(t|\theta)\ dt = 0 \quad \text{for all} \quad \theta \quad \text{implies} \quad g(T) = 0 \quad \text{a.e. for all } \theta..$$

Geometrically, this means that the function space $\mathcal{F} = \{f(x|\theta)\,|\theta \in \Theta\}$ spans the whole space of functions of T. Therefore, the term *complete* refers to the implication that the family of distribution parametrized by θ is complete for T (and its functions).

Example 19.2. Let $T \sim \text{Bin}(n, p)$ with n fixed. Then for any function g, we set

$$\begin{aligned} 0 = \mathbb{E}_p g(T) &= \sum t = 0^n g(t)\binom{n}{t}p^t(1-p)^{n-t} \\ &= (1-p)^n \sum_{t=0}^{n} g(t)\binom{n}{t}\left(\frac{p}{1-p}\right)^t \\ &= (1-p)^n \sum_{t=0}^{n} g(t)\binom{n}{t}r^t, \end{aligned}$$

where $r = p/(1-p)$. This is a polynomial in r which can be zero for all $0 < r\infty$ only if $g(t) = 0$ for all $t = 0, 1, \ldots, n$. Thus T is a complete statistic. ⋄

Example 19.3. Let $X_1, X_2, \ldots, X_n \sim \text{unif}(0, \theta)$ be iid. We know that the statistic $T(\boldsymbol{X}) = \max_i X_i$ is sufficient. To find the pdf $f(t|\theta)$ of T, we start with its cdf:

$$\begin{aligned} \Pr(T \le t|\theta) &= \Pr(X_1 \le t, X_2 \le t, \ldots, X_n \le t) \\ &= \Pr(X_1 \le t)\Pr(X_2 \le t)\ldots\Pr(X_n \le t) \\ &= \left(\frac{t}{\theta}\right)^n. \end{aligned}$$

Thus the pdf is the derivative nt^{n-1}/θ^n. We now set

$$0 = \mathbb{E}_\theta[g(T)] = n\theta^{-n}\int_0^\theta g(t)t^{n-1}\ dt..$$

Since this is a constant (which is zero) for all θ, we have that

$$\begin{aligned} 0 = \frac{d}{d\theta}\mathbb{E}_\theta[g(T)] &= n\frac{d}{d\theta}\left[\theta^{-n}\int_0^\theta g(t)t^{n-1}\,dt\right] \\ &= n\theta^{-n}\frac{d}{d\theta}\int_0^\theta g(t)t^{n-1}\,dt + n\left(\frac{d}{d\theta}\theta^{-n}\right)\int_0^\theta g(t)t^{n-1}\,dt \\ &= n\theta^{-n}g(\theta)\theta^{n-1} + 0 \\ &= n\theta^{-1}g(\theta). \end{aligned}$$

for all $\theta \in (0,\infty)$, implying that $g(T) = 0$ and so T is a complete statistic. ◇

The following theorem shows that completeness helps with the optimal data reduction.

Theorem 19.4 (Bahadur's theorem)**.** If a statistic T is complete and sufficient, then it is minimal sufficient.

To study more about complete statistics, we first mention another family of statistics that is opposite of being sufficient i.e. they contain no information about the parameter.

Definition 19.5. A statistic V is *ancillary* if its distribution does not depend on θ.

A trivial example is the constant statistic $V = c$. Here is another example:

Example 19.6. Let X_1, X_2 iid from $N(\theta, 1)$. Then $X_2 - X_1 \sim N(0, 2)$ is an ancillary statistic. ◇

The following theorem states that complete sufficient statistics contain no ancillary information.

Theorem 19.7 (Basu's theorem)**.** If T is a complete sufficient statistic and V is an ancillary statistic for a family of distribution parametrized by θ, then T and V are independent.

Proof: For any set A, let $q_A(T)$ be a function on T defined by $q_A(t) = \Pr_\theta(V \in A|T = t)$ and $p_A(V) = \Pr_\theta(V \in A)$. By sufficiency and ancillarity, both $q_A(T)$ and $p_A(V)$ are independent of θ. By smoothing, we have

$$\mathbb{E}_\theta[q_A(T)] = \mathbb{E}_\theta[\Pr_\theta(V \in A|T)] = \Pr_\theta[V \in A] = p_A(V).$$

for all θ. In other words, $\mathbb{E}_\theta[q_A(T) - p_A(V)] = 0$, and so completeness yields $q_A(T) = p_A(V)$ independent of T a.e. for all θ. Therefore, by smoothing again

$$\begin{aligned} \Pr_\theta(T \in A, V \in B) &= \mathbb{E}_\theta[1_A(T)1_B(V)] \\ &= \mathbb{E}_\theta[\mathbb{E}_\theta[1_A(T)1_B(V)|T]] \\ &= \mathbb{E}_\theta[1_A(T)\mathbb{E}_\theta[1_B(V)|T]] \\ &= \mathbb{E}_\theta[1_A(T)q_B(T)] \\ &= p_B(V)\mathbb{E}_\theta[1_A(T)] \\ &= p_A(T)p_B(V) \\ &= \Pr(T \in A)\Pr(V \in B). \end{aligned}$$

Thus T and V are independent. ■ Before giving an application of this theorem, we mention an important family of distributions that have a general form of complete statistics.

Example 19.8. Let $T_1, T_2, \ldots, T_s$ be functions. An *exponential family* parametrized by $\theta \in \Theta$ is defined by the pdf

$$f(\boldsymbol{x}|\theta) = h(\boldsymbol{x})g(\theta)\exp\left(\sum_{i=1}^{s}\eta_i(\theta)T_i(\boldsymbol{x})\right) \tag{19.2}$$

where $\eta_1, \eta_2, \ldots, \eta_s$ are functions of the parameter θ. Most of the distributions we have mentioned so far are exponential families. For example, the pdf of a Gaussian distribution can be written as

$$\begin{aligned} f(x|\mu,\sigma^2) &= \frac{1}{\sqrt{2\pi\sigma^2}}e^{-(x-\mu)^2/(2\sigma^2)} \\ &= \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{\mu^2}{2\sigma^2}\right)\exp\left(\frac{\mu}{\sigma^2}x - \frac{1}{2\sigma^2}x^2\right). \end{aligned}$$

In this case, $T_1(x) = x, T_2(x) = x^2, \eta_1(\mu,\sigma^2) = \mu/\sigma^2$ and $\eta_2(\mu,\sigma^2) = -1/(2\sigma^2)$. An important result regarding the exponential family (19.2) is that, given s random sample $X_1, X_2, \ldots, X_n$ iid from this distribution, the statistic

$$T(\boldsymbol{X}) = \left(\sum_{i=1}^{n}T_1(X_i), \sum_{i=1}^{n}T_2(X_i), \ldots, \sum_{i=1}^{n}T_s(X_i)\right)$$

is complete as long as the parameter space Θ has nonempty interior in its ambient Euclidean space's topology. ◇

Example 19.9. Let $X_1, X_2, \ldots, X_n$ be iid from $\exp(\lambda)$. In this case, $T_1(x) = x$ and the statistic

$$T(\boldsymbol{X}) = \sum_{i=1}^{n}X_i$$

is a complete statistic. Also, it is not hard to show using the Factorization theorem that $T(\boldsymbol{X})$ is a sufficient statistic. Now if we want to find the expected value of another statistic

$$g(\boldsymbol{X}) = \frac{X_n}{X_1 + X_2 + \ldots + X_n}$$

we just have to note that $X_i/\lambda \sim \exp(1)$ and so the distribution of X_i/X_n is independent of λ, implying that $g(\boldsymbol{X})$ is ancillary. Using Basu's theorem,

$$\theta = \mathbb{E}_\theta[X_n] = \mathbb{E}_\theta[T(\boldsymbol{X})g(\boldsymbol{X})] = \mathbb{E}_\theta[T(\boldsymbol{X})]\mathbb{E}_\theta[g(\boldsymbol{X})] = n\theta\mathbb{E}_\theta g(\boldsymbol{X}).$$

Therefore, $\mathbb{E}_\theta g(\boldsymbol{X}) = 1/n$. ◇

Example 19.10. Let $X_1, X_2, \ldots, X_n$ be iid from $N(\mu, \sigma^2)$. Here, we provide an alternative proof that $\overline{X}$ and S^2 are independent. First, we fix σ, so the pdf of $\overline{X} \sim N\left(\mu, \sigma^2/n\right)$ is given by

$$f(\bar{x}|\mu) = \frac{n^{1/2}}{(2\pi\sigma^2)^{1/2}}\exp\left(\frac{n\mu\bar{x}}{\sigma^2} - \frac{n(\bar{x})^2}{2\sigma^2} - \frac{n\mu^2}{2\sigma^2}\right)$$

which forms an exponential family with $T_1(\bar{x}) = \bar{x}$ and the rest of terms written as $h(\bar{x})g(\theta)$. Hence, $\overline{X}$ is complete, and we already know from previous lectures that it is sufficient. We also know that the distribution of S^2 only depends on σ^2, and so it is ancillary. Therefore, it follows from Basu's theorem that $\overline{X}$ and S^2 are independent. Since this is true for any σ, they are independent for any μ and σ. ◇