208711 Statistical Theory 1 **Fall 2019**

Lecture 17: October 15

Lecturer: Donlapark Pornnopparath

17.1 Sufficient statistics

For the majority of the course, we will focus on distributions parametrized by $\theta \in \Theta$. Here, θ can be a scalar or a vector and Θ is the *parameter space* of all possible parameters.

From a random sample $X_1, X_2, \ldots, X_n$, we want to make an inference i.e.estimate the correct value of θ. However, gaining some insight from high dimensional data $x_1, x_2, \ldots, x_n$ might be troublesome, and so the notion of *data reduction*, where the data is transformed into a lower dimensional object, was introduced. For a particular value of θ, we want to find a method of data reduction that captures most of the information about θ.

Definition 17.1. Let $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ be a random sample from a distribution with pdf $f(x|\theta)$. Then we say that $T(\boldsymbol{X})$ is a sufficient statistic for θ if the distribution of $\boldsymbol{X}$ given that $T(\boldsymbol{X}) = t$ is independent of θ i.e. $f(\boldsymbol{x}|t, \theta) = f(\boldsymbol{x}|t)$.

Example 17.2. Let $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ be a random sample from a discrete uniform distribution unif$\{1, n\}$. Then $T = T(\boldsymbol{X}) = \max_i X_i$ is a sufficient statistic because $\Pr\left(X_i = a | T = b, n\right) = \frac{1}{b}$ if $a \leq b$ and 0 otherwise. Both of these values do not depend on n. ◇

Example 17.3. Let $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ be a random sample from Ber(p). Suppose that we have observed these variables as $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ and recorded the value of $k = T(\boldsymbol{x}) = x_1 + x_2 + \ldots + x_n$. Then $T(\boldsymbol{X}) \sim \text{Bin}(n, p)$ and

$$\Pr\left(x_1, x_2, \ldots, x_n | k, p\right) = \frac{\Pr(x_1, x_2, \ldots, x_n, k|p)}{\Pr(k|p)}.$$

We know that

$$\Pr\left(x_1, x_2, \ldots, x_n, k|p\right) = \begin{cases} \Pr(x_1, x_2, \ldots, x_n|p) & \text{if } \sum x_i = k \\ 0 & \text{otherwise.} \end{cases}$$

Therefore, if $\sum_i x_i = k$,

$$\Pr\left(x_1, x_2, \ldots, x_n | k, p\right) = \frac{\prod_i \Pr(x_i)}{\binom{n}{k}\theta^k(1-\theta)^{n-k}} = \frac{\theta^k(1-\theta)^{n-k}}{\binom{n}{k}\theta^k(1-\theta)^{n-k}} = \frac{1}{\binom{n}{k}}.$$

Since this does not depend on p, $\sum X_i$ is a sufficient statistic for p. ◇

Example 17.4. Let $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ be iid sample from $N(\mu, \sigma^2)$ where σ is fixed. Let $f_{\boldsymbol{X}}$ be its pdf and $T = \overline{X} = \frac{X_1, X_2, \ldots, X_n}{n}$. From the same computations as in the previous example, we have

$$\begin{aligned}
f_{\boldsymbol{X}}(x_1, x_2, \ldots, x_n | \bar{x}, \mu) &= \frac{f_{\boldsymbol{X},T}(x_1, x_2, \ldots, x_n, \bar{x}|\mu)}{f_T(\bar{x}|\mu)} \\
&= \frac{(2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}\sum_i (x_i - \mu)^2\right)}{(2\pi\sigma^2/n)^{-1/2} \exp\left(-\frac{n}{2\sigma^2}\sum_i (\bar{x} - \mu)^2\right)} \\
&= \frac{n^{-1/2}(2\pi\sigma^2)^{-(n-1)/2} \exp\left(-\frac{1}{2\sigma^2}\sum_i (x_i - \bar{x})^2\right) \exp\left(-\frac{n}{2\sigma^2}\sum_i (\bar{x} - \mu)^2\right)}{\exp\left(-\frac{n}{2\sigma^2}\sum_i (\bar{x} - \mu)^2\right)} \\
&= n^{-1/2}(2\pi\sigma^2)^{-(n-1)/2} \exp\left(-\frac{1}{2\sigma^2}\sum_i (x_i - \bar{x})^2\right).
\end{aligned}$$

Thus $\overline{X}$ is a sufficient statistics for θ. ◇

From previous examples, we see that a statistic $T = T(\boldsymbol{X})$ is sufficient for a parameter θ if the quantity

$$\frac{f_{\boldsymbol{X}}(\boldsymbol{x}|\theta)}{f_T(T(\boldsymbol{x})|\theta)} \tag{17.1}$$

does not depend on θ. The following theorem states that there is a stronger connection between the pdf $f(x|\theta)$ and a sufficient statistic $T(\boldsymbol{X})$:

Theorem 17.5 (Fisher-Neyman)**.** Let $\boldsymbol{X}$ be a random variable with pdf $f(\boldsymbol{x}|\theta)$ The statistic $T = T(\boldsymbol{X})$ is sufficient for θ if and only if f can be factored as

$$f(\boldsymbol{x}|\theta) = g(T(\boldsymbol{x}), \theta)h(\boldsymbol{x}),$$

where the function h does not depend on θ.

Proof: $\Leftarrow$ Let $T(\boldsymbol{x}) = t$ .Assuming such factorization exist, it follows that

$$\begin{aligned}
\frac{f_{\boldsymbol{X}}(\boldsymbol{x}|\theta)}{f_{T(\boldsymbol{X})}(t|\theta)} &= \frac{f_{\boldsymbol{X}}(\boldsymbol{x}|\theta)}{\int f_{\boldsymbol{X},T}(\boldsymbol{y}, t)|\theta) d\boldsymbol{y}} \\
&= \frac{f_{\boldsymbol{X}}(\boldsymbol{x}|\theta)}{\int_{T(\boldsymbol{y})=t} f_{\boldsymbol{X}}(\boldsymbol{y}|\theta) d\boldsymbol{y}} \\
&= \frac{f_{\boldsymbol{X}}(\boldsymbol{x}|\theta)}{\int_{T(\boldsymbol{y})=t} g(T(\boldsymbol{y}), \theta)h(y) d\boldsymbol{y}} \\
&= \frac{g(t,\theta)h(x)}{g(t,\theta)\int_{T(\boldsymbol{y})=t} h(y) d\boldsymbol{y}} \\
&= \frac{h(x)}{\int_{T(\boldsymbol{y})=t} h(y) d\boldsymbol{y}},
\end{aligned}$$

which does not depend on θ, so it follows from the discussion above that $T(\boldsymbol{X})$ is a sufficient statistic for θ. $\Rightarrow$ By sufficiency of $T(\boldsymbol{X})$, $f(\boldsymbol{x}|t,\theta) = f(\boldsymbol{x}|t)$. Then

$$f_{\boldsymbol{X}}(\boldsymbol{x}|\theta) = f_{\boldsymbol{X},T}(\boldsymbol{x}|t,\theta) f_T(t|\theta) . = f_{\boldsymbol{X},T}(\boldsymbol{x}|t) f_T(t|\theta).$$

We finish the proof by letting $g(t,\theta) = f_T(t|\theta)$ and $h(x) = f_{\boldsymbol{X},T}(\boldsymbol{x}|t)$, which is independent of θ. ■

Example 17.6. Observing $x_1, x_2, \ldots, x_n$, the pmf of Ber(p) is

$$f(x_1, x_2, \ldots, x_n|p) = \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i} = \theta^{\sum_i x_i}(1-\theta)^{n-\sum_i x_i}.$$

Applying Theorem 17.5 with $g(t, \theta) = \theta^t(1-\theta)^{1-t}$ and $h(x) = 1$, we have that $\sum_i X_i$ is a sufficient statistic for p. $\diamond$

Example 17.7. From Example 17.4, we have the following factorization for a random sample $X_1, X_2, \ldots, X_n$:

$$\begin{aligned} f(x_1, x_2, \ldots, x_n|\mu) &= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2}\sum_i (x_i-\mu)^2\right) \\ &= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2}\sum_i (x_i-\overline{x})^2\right) \exp\left(-\frac{n}{2\sigma^2}\sum_i (\overline{x}-\mu)^2\right) \\ &= h(x)g(\bar{x}, \mu), \end{aligned}$$

where $h(x) = (2\pi\sigma)^{-n/2}e^{-\sum_i (x_i-\overline{x})/2\sigma^2}$ and $h(t, \mu) = e^{-n(t-\mu)^2/2\sigma^2}$. Thus $\overline{x} = \sum_i x_i$ is a sufficient statistics for μ. $\diamond$