208711 Statistical Theory 1 **Fall 2019**

Lecture 18: October 18

Lecturer: Donlapark Pornnopparath

We give two more examples of sufficient statistics.

Example 18.1. Let $X_1, X_2, \ldots, X_n$ be a random sample from $u(\theta, \theta+1)$ and consider the statistic $T = (\min_i X_i, \max_i X_i)$. We will write the pdf of the sample in terms of indicator functions:

$$1_{(a,b)}(x) = \begin{cases} 1 & \text{if } x \in (a,b) \\ 0 & \text{otherwise} \end{cases}.$$

The pdf of X_i is $f(x_i|\theta) = 1_{(\theta,\theta+1)}(x_i)$. Therefore,

$$f(x_1, x_2, \ldots, x_n|\theta) = \prod_{i=1}^{n} 1_{(\theta,\theta+1)}(x_i) = 1_{(\theta,\theta+1)}(\min_i x_i) 1_{(\theta,\theta+1)}(\max_i x_i).$$

The sufficiency of T follows from the factorization theorem with $g(t,\theta) = 1_{(\theta,\theta+1)}(\min_i x_i)1_{(\theta,\theta+1)}(\max_i x_i)$ and $h(x) = 1$. ◇

Example 18.2. Let $X_1, X_2, \ldots, X_n$ be a random sample from $N(\mu, \sigma^2)$ and $T = (\overline{X}, S^2)$. Then, as in previous examples show, we have

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

which is a function of T, μ and σ^2. Thus, T is a sufficient statistic for μ, σ^2. ◇

18.1 Minimal sufficient statistics

We can see that a distribution can have (infinitely) many sufficient statistics. For example,

$$T(X_1, X_2, \ldots, X_n) = (X_1, X_2, \ldots, X_n) \quad \text{and} \quad T(X_1, X_2, \ldots, X_n) = \overline{X}$$

are both sufficient statistics for μ. Moreover, any statistic $\tilde{T}$ that has $T = F(\tilde{T})$ for any function F is also a sufficient statistic since

$$f(x|\theta) = g(t,\theta)h(x) = g(F(\tilde{t},\theta)h(x).$$

This means that both T and $\tilde{T}$ contain all information about the parameter θ, but the former has a better data reduction than the latter. From this observation, one might want to find a statistic that gets the most data reduction while retaining the information about the parameter.

Definition 18.3. A statistic T is *minimal sufficient* if for any sufficient statistic $\tilde{T}$ there exists a function F such that $T = F(\tilde{T})$.

Example 18.4. Note that a minimal sufficient statistic is not unique, since any one-to-one function of a minimal sufficient statistic is also minimal sufficient. In the case of Bernoulli or normal distributed sample $X_1, X_2, \ldots, X_n$, the sufficient statistic $X_1, X_2, \ldots, X_n$ is not minimal, because it is not a function of $X_1 + X_2 + \ldots + X_n$) while the latter is also a sufficient statistic. $\diamond$

The concept of sufficient statistics can be used to study the likelihood function $\mathcal{L}(\theta|x) = f(x|\theta)$. Specifically, a sufficient statistic T defines the shape of the likelihood function i.e. if $T(x) = T(y)$ then $\mathcal{L}(\theta|x)$ and $\mathcal{L}(\theta|y)$ differ by a constant scale with respect to θ:

$$\mathcal{L}(\theta|y) = g(T(y), \theta)h(y) = g(T(x), \theta)h(y) = \frac{h(y)}{h(x)}\mathcal{L}(\theta|x). \tag{18.1}$$

The following theorem shows that the statistic is minimal sufficient if there is a one-to-one correspondence between the realized value of the statistic and the shape of the likelihood.

Theorem 18.5. Let $f(x|\theta)$ be the pdf of a distribution. If there exists a function $T(x)$ such that $f(x|\theta) = c(x, y)f(y|\theta)$ for some function $c(x, y)$ independent of θ if and only if $T(x) = T(y)$, then T is a minimal sufficient statistics.

Proof: First, we prove that T is sufficient. Let $\mathcal{X}$ be the sample space and $S = T(\mathcal{X})$ be the image of $\mathcal{X}$ under T. From each $S_t = \{x : T(x) = t\} \subset S$ we pick a representative $y_t \in S_t$. We now compute

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

where $g(t, \theta) = f(y_t|\theta)$ and $h(x) = f(x|\theta)/f(y_{(T(x)}|\theta)$ is independent of θ from the assumptions on T. Thus T is sufficient from the factorization theorem.
We now prove that T is minimal. Let T' be another sufficient statistic. Then by the factorization theorem $f(x|\theta) = g'(T'(x), \theta)h'(x)$ for some functions g' and h'. For any x and y such that $T'(x) = T'(y)$ it follows from (18.1) that

$$f(y|\theta) = \frac{h(y)}{h(x)}f(x|\theta),$$

where $h(y)/h(x)$ does not depend on θ. Thus, the assumptions on T tells us that $T'(x) = T'(y)$ implies $T(x) = T(y)$. In other words, T is a function of T'. $\blacksquare$

Example 18.6. Let $X_i, Y_i, i = 1, 2, \ldots, n$ be two random samples from $N(\mu, \sigma^2)$ where μ and σ are unknown. We would like to know if $(\overline{X}, S^2)$ is a minimal sufficient statistic for this parameter. To find out, we compute the ratio

$$\begin{aligned}\frac{f(\boldsymbol{x}|\mu, \sigma^2)}{f(\boldsymbol{y}|\mu, \sigma^2)} &= \frac{(2\pi\sigma^2)^{(-n/2)} \exp\left(-(n-1)s_x^2/(2\sigma^2)\right) \exp\left(-n(\bar{x} - \mu)^2/(2\sigma^2)\right)}{(2\pi\sigma^2)^{(-n/2)} \exp\left(-(n-1)s_y^2/(2\sigma^2)\right) \exp\left(-n(\bar{y} - \mu)^2/(2\sigma^2)\right)} \\ &= \exp\left([-(n-1)(s_x^2 - s_y^2) - n(\bar{x}^2 - \bar{y}^2) + 2n\mu\,(\bar{x} - \bar{y})]/\left(2\sigma^2\right)\right).\end{aligned}$$

This is independent of μ and σ^2 if and only if $\bar{x} = \bar{y}$ and $s_x^2 = s_y^2$. Thus, by Theorem 18.5, $(\overline{X}, S^2)$ is a minimal sufficient statistic for (μ, σ^2). $\diamond$

Example 18.7. Let $X_i, Y_i, i = 1, 1, \ldots, n$ be two random samples from Pois(λ). We would like to know that the statistic

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

is a minimal complete statistic. So we compute the ratio

$$\begin{aligned}\frac{f(\boldsymbol{x}|\lambda)}{f(\boldsymbol{y}|\lambda)} &= \frac{\prod_i \lambda^{x_i} e^{-\lambda}/x_i!}{\prod_i \lambda^{y_i} e^{-\lambda}/y_i!} \\ &= \frac{\prod_i y_i}{\prod_i x_i}\lambda^{\sum_i x_i - \sum_i y_i}.\end{aligned}$$

which is a constant as a function of λ only if $\sum_i x_i = \sum_i y_i$. Thus $T(\boldsymbol{X}) = \sum_i X_i$ is a minimal sufficient statistic. ◇

Example 18.8. Let $X_i, Y_i, i = 1, 1, \ldots, n$ be two random samples from unif$(\theta, \theta + 1)$. We want to know whether

$$T(\boldsymbol{X}) = \left(\min_i X_i, \max_i X_i\right)$$

is a minimal sufficient statistic. Recall from Example 18.1 that

$$f(\boldsymbol{x}|\theta) = 1_{(\theta,\theta+1)}(\min_i x_i) 1_{(\theta,\theta+1)}(\max_i x_i).$$

To avoid division by zero in the following argument, we assume that the observations are contained within an interval of length $\frac{1}{2}$. For $f(\boldsymbol{x}|\theta)/f(\boldsymbol{y}|\theta)$ to not depend on θ, it must be the case that $x_m = \min_i x_i$ and $y_m = \min_i y_i$ are equal, otherwise the value of the ratio will be different between $\theta = (x_m + y_m)/2$ and $\theta = \min(x_m, y_m) - \frac{1}{2}$. Similarly, we must have $\max_i x_i = \max_i y_i$. Therefore, $T(\boldsymbol{X}) = (\min_i X_i, \max_i X_i)$ is a minimal sufficient statistic for θ. ◇