Math 502AB - Lecture 20

Dr. Jamshidian

November 1, 2017

1 Lecture - Part 1

1.1 Chapter 6 - Section 6.2.1 - Sufficient Statistics

Suppose that we have a sample $X_1, ..., X_n$ from a density (or *pmf*) $f(x|\theta)$ and we are interested in estimating θ based on our sample. We would usually use a statistic $T(X_1, ..., X_n)$ to estimate θ.

If $X_1, ..., X_n | T(X_1, ..., X_n)$ is **not** a function of θ, then $T(X_1, ..., X_n)$ is a *sufficient statistic*.

Example

1. Let $X_1, .., X_n$ be a random sample from a distribution with *pmf*

$$f(X|\theta) = \begin{cases} \theta^x(1-\theta)^{1-x} & x = 0, 1 \quad 0 \leq \theta \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Consider the statistic,

$$Y = T(X) = \sum_{i=1}^{n} x_i \sim Binomial(n, p)$$

$$f_{T(X)}(y) = \binom{n}{y} \theta^y (1-\theta)^{n-y} \quad y = 0, 1, ..., n$$

Now consider the conditional probability

$$P(X_1 = x_1, X_2 = x_2, ..., X_n = x_n | T(X) = y)$$

There are two cases here:

(a) If $\sum x_i \neq y$,

$$\Rightarrow P(X_1 = x_1, ..., X_n = x_n | T(X) = y) = 0$$

(b) If $\sum x_i = y$. We then have that $X_1 = x_1, ..., X_n = x_n$ is a *subset* of $T(X) = y$, since it is only one way we can get $\sum X_i = y$. Then we have:

$$\begin{aligned}
P(A|B) &= \frac{P(A \cap B)}{P(B)} = \frac{P(A)}{P(B)} = \frac{P(X_1 = x_1, ..., X_n = x_n)}{P(T(X) = y)} \\
&= \frac{\prod_{i=1}^{n} \theta^{x_i}(1-\theta)1 - x_i}{\binom{n}{y}\theta^y(1-\theta)^{n-y}} \\
&= \frac{\theta^{\sum x_i}(1-\theta)^{n-\sum x_i}}{\binom{n}{y}\theta^y(1-\theta)^{n-y}} \\
&= \frac{\theta^{y}(1-\theta)^{n-y}}{\binom{n}{y}\theta^y(1-\theta)^{n-y}} \\
&= \frac{1}{\binom{n}{y}}
\end{aligned}$$

Thus, this isn't a function of θ, and we have a sufficient statisitc.

1.1.1 Sufficient Statistic (Discrete Case)

Let $f_Y(y|\theta)$ be the *pmf* of the statistic $Y = T(X_1, ..., X_n)$, where $X_1, ..., X_n$ is a random sample from a discrete distribution with *pmf* $f(x|\theta)$. Then,

$$P(X_1 = x_1, ..., X_n = x_n | Y = y) = \frac{\prod_{i=1}^{n} f(x_i|\theta)}{f_Y(T(X_1, ..., X_n|\theta)}$$

We say that $Y = T(X_1, ..., X_n)$ is a sufficient statistic for θ if this ratio **does not** depend on θ.

1.1.2 Sufficient Statistic (Continuous Case)

Let $X_1, ..., X_n$ be a sample of size n from a distribution with *pdf* $f(x|\theta)$. Let $Y = T(X_1, ..., X_n)$ be a statistic with *pdf* $f_Y(y|\theta)$. Then Y is a sufficient statistic for θ **if and only if**

$$\frac{\prod_{i=1}^{n} f(x_i|\theta)}{f_Y(T(X_1, ..., X_n|\theta)} = H(X_1, ..., X_n)$$

where $H(X_1, ..., X_n)$ does not depend on θ.

Examples:

1. Let $X_1, ..., X_n \sim gamma(\alpha = 2, \beta = \theta)$ (**iid**). Consider $Y = \sum X_i$. Is this a sufficient statistic to estimate θ.

$$Y = \sum_{i=1}^{n} X_i \sim gamma(\alpha = 2n, \beta = \theta)$$

Then, we look at,

$$\begin{aligned}
\frac{\prod_{i=1}^{n} f(x_i|\theta)}{f_Y(\sum X_i|\theta)} &= \frac{\prod_{i=1}^{n}\left[\frac{1}{\Gamma(2)\theta^2}x_i^{2-1}e^{-x_i/\theta}\right]}{\frac{1}{\Gamma(2n)\theta^{2n}}\left(\sum X_i\right)^{2n-1}e^{-\sum X_i/\theta}} \\
&= \frac{\frac{1}{\theta^{2n}}e^{-\sum X_i/\theta}\prod_{i=1}^{n} X_i}{\frac{1}{\Gamma(2n)\theta^{2n}}\left(\sum X_i\right)^{2n-1}e^{-\sum X_i/\theta}} \\
&= \frac{\Gamma(2n)\prod_{i=1}^{n} X_i}{\left(\sum X_i\right)^{2n-1}}
\end{aligned}$$

Since this is not a function of θ, we have a sufficient statistic.

2. Let $X_{(1)}, ..., X_{(n)}$ denote the order statistics of a random sample of size n from a distribution with *pdf*

$$f(x|\theta) = e^{-(x-\theta)}\mathcal{I}_{(\theta,\infty)}(x)$$

Show that $Y = X_{(1)} = min(X_1, ..., X_n)$ is a sufficient statistic to estimate θ.

First, we must find the distribution for Y:

$$\begin{aligned}
f_Y(y) &= n(1 - F_{X_1}(y))^{n-1}f_{X_1}(y) \\
F_{X_1}(y) &= \int_{\theta}^{y} e^{-(x_1-\theta)}dx = 1 - e^{-(y-\theta)} \\
f_Y(y) &= ne^{-n(y-\theta)}\mathcal{I}_{(\theta,\infty)}(y)
\end{aligned}$$

Now, we consider the ratio:

$$\begin{aligned}
\frac{\prod_{i=1}^{n} f_{X_1}(X_i|\theta)}{f_Y(T(X_i|\theta))} &= \frac{\prod_{i=1}^{n} e^{-(x_i-\theta)}\mathcal{I}_{\theta,\infty}(x_i)}{ne^{-n(min(X_1,...,X_n))}\mathcal{I}_{(\theta,\infty)}} \\
&= \frac{e^{-\sum x_i}e^{n\theta}\prod_{i=1}^{n}\mathcal{I}_{(\theta,\infty)(x_i)}}{ne^{n\theta}e^{-min(x_1,...,x_n)}\mathcal{I}_{(\theta,\infty)}(min\,(x_1,...,x_n))}
\end{aligned}$$

We must consider:

$$\begin{aligned}
\prod_{i=1}^{n}\mathcal{I}_{(\theta,\infty)}(x_i) = 1 &\iff X_1 > \theta, ..., X_n > \theta \\
&\iff min(x_1, ..., x_n) > \theta \\
&\iff \mathcal{I}_{(\theta,\infty)}(min(x_1, ..., x_n)) = 1
\end{aligned}$$

This is important since if the minimum is below θ, the probability is zero anyway so these indicator functions cancel out. This leaves us with terms that are *not* dependent upon θ, and thus it is a sufficient statistic.

2 Lecture - Part 2

So, now we have looked at how to determine if something is a sufficient statistic. How can we figure out what a sufficient statistic might be?

2.1 Factorization Theorem

Let $X_1, ..., X_n$ be **iid** from a distribution with *pdf* (or *pmf*) $f(X|\theta)$, where $\theta \in \Theta$ is an unknown parameter. A statistic $T(X_1, ..., X_n)$ is a sufficient statistic for θ **if and only if** the joint *pdf*, or the joint *pmf* function $f(x_1, ..., x_n|\theta)$ of $X_1, ..., X_n$ can be factorized as follows for all $x = (x_1, ..., x_n) \in \mathbb{R}^n$ and values of $\theta \in \Theta$

$$f(x_1, ..., x_n|\theta) = U(x_1, ..., x_n)V\left[T(x_1, ..., x_1|\theta)\right]$$

Here U and V are non-negative functions where U does not involve θ, and V depends on $x_1, ..., x_n$ *only* through $T(X_1, ..., X_n)$.

Proof (Discrete Case):

($\Rightarrow$) Suppose that the joint density can be factored as above. Let $T(X) = t$.

$$\begin{aligned} P(X_1 = x_1, ..., X_n = x_n|T(X_1, ..., X_n) = t) &= \frac{P(X_1 = x_1, ..., X_n = x_n \cap T(X_1, ..., X_n) = t)}{P(T(X_1, ..., X_n) = t)} \\ &= \frac{P(X_1 = x_1, ..., X_n = x_n)}{P(T(X_1, ..., X_n) = t)} \end{aligned}$$

We already know that the top part of this equation is the joint density (U). So, let's look at the bottom term:

$$\begin{aligned} P(T(X_1, ..., X_n) = t) &= \sum_{x:T(x)=t} f(x) \\ &= \sum_{x:T(x)=t} U(X)V(T(X|\theta)), \quad \text{by assumption} \\ &= \sum_{x:T(x)=t} U(x)V(t|\theta) = V(t|\theta) = \sum_{x:T(x)=t} U(x) \end{aligned}$$

So, now we have:

$$\begin{aligned} P(X = x) = f(x|\theta) &= u(x)v(T(x)|\theta) \\ &= u(x)v(t|\theta) \end{aligned}$$

Putting it back together, we have:

$$\frac{P(X = x)}{P(T(x) = t)} = \frac{u(x)v(t|\theta)}{v(t|\theta)\sum_{x:T(x)=t} u(x)}$$

The result is independent of θ, and thus $T(X)$ must be a sufficient statisitic.

($\Leftarrow$): Assume that $T(X)$ is a sufficient statistic for θ. This means that

$$P(X = x|T = t) \text{ is independent of } \theta$$

But,

$$\begin{aligned} P(X = x) &= P(X = x \cap T(x) = t) \\ &= P(X = x|T(x) = t)P(T(x) = t) \\ &= U(x)V(T(x)|\theta), \text{ by sufficiency} \end{aligned}$$

2.1.1 Examples

1. Let $X_1, ..., X_n$ be a sample from a *Poisson* distribution with mean θ (θ ¿ 0). Obtain a sufficient statistic for θ.

$$f(x_1, ..., x_n|\theta) = \frac{\prod_{i=1}^n e^{-\theta}\theta^{x_i}}{x_i!} = e^{-n\theta}\theta^{\sum x_i}\left(\frac{1}{\prod_{i=1}^n x_i!}\right)$$

   $\Rightarrow$ that $\sum x_i$ is a sufficient statistic

2. Suppose that $X_1, ..., X_n$ are a sample from a continuous distribution with *pdf*

$$\begin{cases} \theta x^{\theta-1} & \text{for } 0 < X < 1 \\ 0 & \text{otherwise} \end{cases}$$

   Obtain a sufficient statistic for θ

$$f(x_1, ..., x_n|\theta) = \prod_{i=1}^n \theta x_i^{\theta-1} = \theta^n \left(\prod_{i=1}^n x_i\right)^{\theta-1}$$

   Well, we can see that $U(X) = 1$ so this trivially implies to us that $\prod x_i$ is a sufficient statistic, and we are done.

3. Let $X_1, ..., X_n \sim N(\theta, \sigma^2)$ (**iid**) where σ^2 is known.

$$\begin{aligned} f(x_1, ..., x_n|\theta) &= \prod_{i=1}^n f(x_i|\theta) \\ &= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i - \theta)^2\right\} \end{aligned}$$

   The left term is fine, but the right hand term is not clearly separable. So, let's take a look at:

$$\begin{aligned} \sum_{i=1}^n (x_i - \theta)^2 &= \sum_{i=1}^n (x_i - \overline{x} + \overline{x} - \theta)^2 \\ &= \sum (x_i - \overline{x})^2 + n(\overline{x} - \theta) \end{aligned}$$

Now, we can rewrite our function as

$$\begin{aligned}
f(x_1, ..., x_n|\theta) &= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left\{-\frac{1}{2\sigma^2}\left[\sum(x_i - \overline{x})^2 + n(\overline{x} - \theta)^2\right]\right\} \\
&= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left\{-\frac{1}{2\sigma^2}\sum(x_i - \overline{x})^2\right\} \exp\left\{-\frac{n}{2\sigma^2}(\overline{x} - \theta)^2\right\}
\end{aligned}$$

The left two terms are our $U(X)$ since it does not depend on θ (σ^2 is known). The statistic on the right hand side is $\overline{X}$, so that is our sufficient statistic.

4. Let $f(x|\theta) = e^{-(x-\theta)}\mathcal{I}_{\theta,\infty}$. You recall that the sufficient statistic from earlier was $min(X_1, ..., X_n)$. Consider a sample of size $n = 3$. We then have,

$$e^{-(x_1-\theta)}e^{-(x_2-\theta)}e^{-(x_3-\theta)} = e^{-3X_{(3)}+3\theta}e^{-(x_1+x_2+x_3)+3X_{(3)}}$$

However, it is important to be mindful of the **domain** of X. Since the domain itself is a function of θ, we must include it in these terms. Thus, the above is **incorrect**.

We should have written the following

$$\begin{aligned}
\prod_{i=1}^{3}\left[e^{-(x_i-\theta)}\mathcal{I}_{(\theta,\infty)}(x_i)\right] &= e^{3\theta}\prod_{i=1}^{3}\mathcal{I}_{(\theta,\infty)}(x_i)e^{-\sum x_i} \\
&= e^{3\theta}\mathcal{I}_{\theta,\infty}(min(x_i))e^{-\sum x_i}
\end{aligned}$$

Since the left two terms are functions of the sample and θ, we see that it is V. The right term is only in terms of the sample, so it composes U. Thus, $min(X_i)$ is our sufficient statistic.

2.2 Properties of Sufficient Statistics

Let $X_1, ..., X_n$ be a sample from a *pdf* (or *pmf*). Suppose that $T(X_1, ..., X_n)$ and $T'(X_1, ..., X_n)$ are two statistics, and there exists a *bijective* (1-1 and onto) function g such that $T' = g(T)$. In particular, T' can be determined from T, without knowing $X_1, ..., X_n$. Then T' is a *sufficient statistic* for θ **if and only if** T is a sufficient statistic for θ.

Example

1. Suppose that $T_1 = \sum X_i$ and $T_2 = \frac{1}{n}\sum X_i = \overline{X}$. If one is a sufficient statistic than the other is as well, since there is a *bijective* function between them. In this case $g(T) = \frac{1}{n}T$

Proof:

Let $T' = g(T) \Rightarrow T = g^{-1}(T')$. This is a sufficient statistic **if and only if**

$$\begin{aligned} f_n(x_1, ..., x_n | T = t) &= U(X_1, ..., X_n) V(T(X_1, ..., X_n | \theta) \\ &= U(X_1, ..., X_n) V(g^{-1}(T'(X_1, ..., X_n)), \theta) \end{aligned}$$

This says that T' must be also a sufficient statistic, and we are done.

2.2.1 Examples

1. Let $X_1, ..., X_n \sim Beta(\alpha, \beta)$ (**iid**) where α is known, but β is unknown. Show that the following statistic is sufficient for β

$$T = \frac{1}{n}\left(\sum_{i=1}^{n} \log \frac{1}{1 - X_i}\right)^3$$

$$\begin{aligned} f(x|\beta) &= \prod_{i=1}^{n} \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x_i^{\alpha - 1}(1 - x_i)^{\beta - 1} \\ &= (\Gamma(\alpha))^{-n} \left(\prod_{i=1}^{n} x_i\right)^{\alpha - 1} \left[\left(\frac{\Gamma(\alpha + \beta)}{\Gamma(\beta)}\right)^n \left(\prod_{i=1}^{n}(1 - x_i)\right)^{\beta - 1}\right] \end{aligned}$$

We know $\prod(1 - X_i)$ is the sufficient statistic, but it doesn't look like the T given. However, if we apply this function, we get our T:

$$g(T') = \frac{-\log(T')^3}{n}$$