# Lecture 11

## 11.1 Sufficient statistic.

(Textbook, Section 6.7)

We consider an i.i.d. sample $X_1, \ldots, X_n$ with distribution $\mathbb{P}_\theta$ from the family $\{\mathbb{P}_\theta : \theta \in \Theta\}$. Imagine that there are two people A and B, and that

1. A observes the entire sample $X_1, \ldots, X_n$,

2. B observes only one number $T = T(X_1, \ldots, X_n)$ which is a function of the sample.

Clearly, A has more information about the distribution of the data and, in particular, about the unknown parameter $\theta$. However, in some cases, for some choices of function $T$ (when $T$ is so called sufficient statistics) B will have as much information about $\theta$ as A has.

**Definition.** $T = T(X_1, \cdots, X_n)$ is called *sufficient statistics* if

$$\mathbb{P}_\theta(X_1, \ldots, X_n | T) = \mathbb{P}'(X_1, \ldots, X_n | T), \tag{11.1}$$

i.e. the conditional distribution of the vector $(X_1, \ldots, X_n)$ given $T$ does not depend on the parameter $\theta$ and is equal to $\mathbb{P}'$.

If this happens then we can say that $T$ contains all information about the parameter $\theta$ of the disribution of the sample, since given $T$ the distribution of the sample is always the same no matter what $\theta$ is. Another way to think about this is: why the second observer $B$ has as much information about $\theta$ as observer A? Simply, given $T$, the second observer $B$ can generate another sample $X'_1, \ldots, X'_n$ by drawing it according to the distribution $\mathbb{P}'(X_1, \cdots, X_n | T)$. He can do this because it does not require the knowledge of $\theta$. But by (11.1) this new sample $X'_1, \ldots, X'_n$ will have the same distribution as $X_1, \ldots, X_n$, so B will have at his/her disposal as much data as the first observer A.

The next result tells us how to find sufficient statistics, if possible.

**Theorem.** (Neyman-Fisher factorization criterion.) $T = T(X_1, \ldots, X_n)$ *is sufficient statistics if and only if the joint p.d.f. or p.f. of* $(X_1, \ldots, X_n)$ *can be represented* *as*

$$f(x_1,\ldots,x_n|\theta) \equiv f(x_1|\theta)\ldots f(x_n|\theta) = u(x_1,\ldots,x_n)v(T(x_1,\ldots,x_n),\theta) \qquad (11.2)$$

*for some function $u$ and $v$. ($u$ does not depend on the parameter $\theta$ and $v$ depends on the data only through $T$.)*

**Proof.** We will only consider a simpler case when the distribution of the sample is discrete.

**1.** First let us assume that $T = T(X_1,\ldots,X_n)$ is sufficient statistics. Since the distribution is discrete, we have,

$$f(x_1,\ldots,x_n|\theta) = \mathbb{P}_\theta(X_1 = x_1,\ldots,X_n = x_n),$$

i.e. the joint p.f. is just the probability that the sample takes values $x_1,\ldots,x_n$. If $X_1 = x_1,\ldots,X_n = x_n$ then $T = T(x_1,\ldots,x_n)$ and, therefore,

$$\mathbb{P}_\theta(X_1 = x_1,\ldots,X_n = x_n) = \mathbb{P}_\theta(X_1 = x_1,\ldots,X_n = x_n, T = T(x_1,\ldots,x_n)).$$

We can write this last probability via a conditional probability

$$\begin{aligned}&\mathbb{P}_\theta(X_1 = x_1,\ldots,X_n = x_n, T = T(x_1,\ldots,x_n))\\ &= \mathbb{P}_\theta(X_1 = x_1,\ldots,X_n = x_n|T = T(x_1,\ldots,x_n))\mathbb{P}_\theta(T = T(x_1,\ldots,x_n)).\end{aligned}$$

All together we get,

$$f(x_1,\ldots,x_n|\theta) = \mathbb{P}_\theta(X_1 = x_1,\ldots,X_n = x_n|T = T(x_1,\ldots,x_n))\mathbb{P}_\theta(T = T(x_1,\ldots,x_n)).$$

Since $T$ is sufficient, by definition, this means that the first conditional probability

$$\mathbb{P}_\theta(X_1 = x_1,\ldots,X_n = x_n|T = T(x_1,\ldots,x_n)) = u(x_1,\ldots,x_n)$$

for some function $u$ independent of $\theta$, since this conditional probability does not depend on $\theta$. Also,

$$\mathbb{P}_\theta(T = T(x_1,\ldots,x_n)) = v(T(x_1,\ldots,x_n),\theta)$$

depends on $x_1,\ldots,x_n$ only through $T(x_1,\ldots,x_n)$. So, we proved that if $T$ is sufficient then (11.2) holds.

**2.** Let us now show the opposite, that if (11.2) holds then $T$ is sufficient. By definition of conditional probability, we can write,

$$\begin{aligned}&\mathbb{P}_\theta(X_1 = x_1,\ldots,X_n = x_n|T(X_1,\ldots,X_n) = t)\\ &= \frac{\mathbb{P}_\theta(X_1 = x_1,\ldots,X_n = x_n, T(X_1,\ldots,X_n) = t)}{\mathbb{P}_\theta(T(X_1,\ldots,X_n) = t)}.\end{aligned} \qquad (11.3)$$

First of all, both side are equal to zero unless

$$t = T(x_1, \ldots, x_n), \tag{11.4}$$

because when $X_1 = x_1, \ldots, X_n = x_n$, $T(X_1, \ldots, X_n)$ must be equal to $T(x_1, \ldots, x_n)$. For this $t$, the numerator in (11.3)

$$\mathbb{P}_\theta(X_1 = x_1, \ldots, X_n = x_n, T(X_1, \ldots, X_n) = t) = \mathbb{P}_\theta(X_1 = x_1, \ldots, X_n = x_n),$$

since we just drop the condition that holds anyway. By (11.2), this can be written as

$$u(x_1, \ldots, x_n)v(T(x_1, \ldots, x_n), \theta) = u(x_1, \ldots, x_n)v(t, \theta).$$

As for the denominator in (11.3), let us consider the set

$$A(t) = \{(x_1, \ldots, x_n) : T(x_1, \ldots, x_n) = t\}$$

of all possible combinations of the $x$'s such that $T(x_1, \ldots, x_n) = t$. Then, obviously, the denominator in (11.3) can be written as,

$$\begin{aligned} &\mathbb{P}_\theta(T(X_1, \ldots, X_n) = t) = \mathbb{P}_\theta((X_1, \ldots, X_n) \in A(t)) \\ &= \sum_{(x_1, \cdots, x_n) \in A(t)} \mathbb{P}_\theta(X_1 = x_1, \ldots, X_n = x_n) = \sum_{(x_1, \cdots, x_n) \in A(t)} u(x_1, \ldots, x_n)v(t, \theta) \end{aligned}$$

where in the last step we used (11.2) and (11.4). Therefore, (11.3) can be written as

$$\frac{u(x_1, \ldots, x_n)v(t, \theta)}{\sum_{A(t)} u(x_1, \ldots, x_n)v(t, \theta)} = \frac{u(x_1, \ldots, x_n)}{\sum_{A(t)} u(x_1, \ldots, x_n)}$$

and since this does not depend on $\theta$ anymore, it proves that $T$ is sufficient.

□

**Example.** Bernoulli Distribution $B(p)$ has p.f. $f(x|p) = p^x(1-p)^{1-x}$ for $x \in \{0, 1\}$. The joint p.f. is

$$f(x_1, \cdots, x_n|p) = p^{\sum x_i}(1-p)^{n - \sum x_i} = v(\sum X_i, p),$$

i.e. it depends on $x$'s only through the sum $\sum x_i$. Therefore, by Neyman-Fisher factorization criterion $T = \sum X_i$ is a sufficient statistic. Here we set

$$v(T, p) = p^T(1-p)^{n-T} \text{ and } u(x_1, \ldots, x_n) = 1.$$