Lec 4. Sufficient and complete Statistics

Zhang W.P.

Contents

4.1 Sufficient Statistics and Data Reduction 1
4.2 Completeness 12

Learning objectives

After this section, you should be able to ...

- understand the definition of sufficient and complete statistics
- can find the sufficient and complete statistic if existed

4.1 Sufficient Statistics and Data Reduction

Now we will talk about a special class of statistics called sufficient statistics. There are two ways to motivate sufficient statistics:

- **the data reduction viewpoint** where we would like to discard non-informative pieces of the dataset (for storage or other benefits), and
- **the risk reduction viewpoint** where we want to construct estimators that only depend on meaningful variation in the data. We will discuss shortly.

The goal in data reduction roughly is to find ways to reduce the size of a dataset without throwing away important information. We need to fix ideas more concretely to make progress on this abstract question

We focus on parametric models and suppose we observe samples

$$X_1, \ldots, X_n \sim f(x; \theta)$$

The typical statistical estimation problem is that we observe the samples as want to understand something about the unknown parameter θ. The goal of data reduction is to find statistics $T(X_1, \ldots, X_n)$ that contain all the information about the unknown parameter θ

Sufficient Statistic: A statistic $T(X_1, \ldots, X_n)$ is said to be sufficient for the parameter θ if

$$f(x_1, \ldots, x_n \mid T(X_1, \ldots, X_n) = t; \theta)$$

Def.

does not depend on θ.

The Factorization Theorem

A much simpler characterization of sufficiency comes from what is called the **Neyman-Fisher factorization criterion**

Theorem 1. *$T(X_1, \ldots, X_n)$ is sufficient for θ if and only if the joint pdf/pmf of $(X_1, \ldots, X_n)$ can be factored as*

$$p(x_1, \ldots, x_n; \theta) = h(x_1, \ldots, x_n) \times g(T(x_1, \ldots, x_n); \theta)$$

The factorization theorem is relatively straightforward to prove, at least in the discrete case.

Proof. Factorization $\Longrightarrow$ sufficiency:

$$\begin{aligned} p\left(x_1, \ldots, x_n \mid T = t; \theta\right) &= \frac{p\left(x_1, \ldots, x_n, T = t; \theta\right)}{p(T = t; \theta)} \\ &= \frac{I\left(T\left(x_1, \ldots, x_n\right) = t\right) h\left(x_1, \ldots, x_n\right) \times g(t; \theta)}{\sum_{x_1, \ldots, x_n : T(x_1, \ldots, x_n) = t} h\left(x_1, \ldots, x_n\right) \times g(t; \theta)} \\ &= \frac{I\left(T\left(x_1, \ldots, x_n\right) = t\right) h\left(x_1, \ldots, x_n\right)}{\sum_{x_1, \ldots, x_n : T(x_1, \ldots, x_n) = t} h\left(x_1, \ldots, x_n\right)} \end{aligned}$$

which does not depend on θ.

Sufficiency $\Longrightarrow$ factorization: We simply define $g(t; \theta) = p(T = t; \theta)$, and $h\left(x_1, \ldots, x_n\right) = p\left(x_1, \ldots, x_n \mid T = t; \theta\right)$, where by sufficiency we note that the latter function does not depend on θ. Now, it is straightforward to verify that factorization theorem holds. □

Minimal sufficiency

The sufficient statistics are not unique. Furthermore, it seems, at least intuitively, that some sufficient statistics present much more reduction than others.

Minimal Sufficiency: A statistic $T(x_1, \ldots, x_n)$ is minimal sufficient if it is sufficient, and furthermore for any other sufficient statistic $S(x_1, \ldots, x_n)$, we can write $T(x_1, \ldots, x_n) = g(S(x_1, \ldots, x_n))$, i.e. T is a function of S. Def.

Theorem 2. *Suppose* $(X_1, \ldots, X_n) \sim f(x_1, \ldots, x_n; \theta)$, $\theta \in \Theta$. *Define*

$$R(x_1, \ldots, x_n, y_1, \ldots, y_n; \theta) = \frac{f(y_1, \ldots, y_n; \theta)}{f(x_1, \ldots, x_n; \theta)}$$

Suppose that a statistic T *has the following property:*

- $R(x_1, \ldots, x_n, y_1, \ldots, y_n; \theta)$ *does not depend on* θ *if and only if*

$$T(y_1, \ldots, y_n) = T(x_1, \ldots, x_n)$$

Then T *is a minimal sufficient statistic.*

Proof. We prove this in two steps. We first show that T is a sufficient statistic and then we check that it is minimal. We define the partition induced by T, as $\{A_t : t \in \text{Range}(T)\}$ and for each set in the partition A_t we associate a representative $(x_{t1}, \ldots, x_{tn}) \in A_t$.

T is sufficient: We look at the joint distribution at any $(x_1, \ldots, x_n)$. Suppose that $T(x_1, \ldots, x_n) = u$, then consider $(y_1, \ldots, y_n) := (x_{u1}, \ldots, x_{un})$. Observe that, $(y_1, \ldots, y_n)$ depends only on $T(x_1, \ldots, x_n)$, i.e. the point y is a function of the statistic T only. Now we have that,

$$p(x_1, \ldots, x_n; \theta) = p(y_1, \ldots, y_n; \theta) R(y_1, \ldots, y_n, x_1, \ldots, x_n; \theta)$$

and since $T(x_1, \ldots, x_n) = T(y_1, \ldots, y_n)$, R does not depend on θ. Recalling that $(y_1, \ldots, y_n)$ is only a function of $T(x_1, \ldots, x_n)$, we have that,

$$p(x_1, \ldots, x_n; \theta) = g(T(x_1, \ldots, x_n); \theta) h(x_1, \ldots, x_n)$$

where g corresponds to the first term and h corresponds to the R term. We conclude that T is sufficient.

T is minimal: As a preliminary we note that the definition of a minimal sufficient statistic could be equivalently written as: T is a MSS if for any other sufficient statistic S, if we have that $S(x_1, \ldots, x_n) = S(y_1, \ldots, y_n)$

then we also have that $T(x_1, \ldots, x_n) = T(y_1, \ldots, y_n)$ This is equivalent to the statement that T is a function of S. Consider, any other sufficient statistic S. Suppose that, $S(x_1, \ldots, x_n) = S(y_1, \ldots, y_n)$, then by the factorization theorem we have that,

$$\begin{aligned} p(x_1, \ldots, x_n; \theta) &= g(S(x_1, \ldots, x_n); \theta) h(x_1, \ldots, x_n) \\ &= g(S(y_1, \ldots, y_n); \theta) h(y_1, \ldots, y_n) \frac{h(x_1, \ldots, x_n)}{h(y_1, \ldots, y_n)} \\ &= p(y_1, \ldots, y_n; \theta) \frac{h(x_1, \ldots, x_n)}{h(y_1, \ldots, y_n)} \end{aligned}$$

so we have that $R(x_1, \ldots, x_n, y_1, \ldots, y_n; \theta)$ does not depend on θ. So we conclude that $T(x_1, \ldots, x_n) = T(y_1, \ldots, y_n)$ and so T is minimal. □

Lemma 1. *If T and S are minimal sufficient statistics, then there exist injective functions g_1 and g_2 such that $T = g_1(S)$ and $S = g_2(T)$.*

Proof. The definition of minimal sufficiency implies that there must exist some functions g_1 and g_2 such that $T = g_1(S), S = g_2(T)$. The task is to prove that g_1 and g_2 are injective on the ranges of S and T respectively.

Suppose x and x' are such that $g_1(S(x)) = g_1\left(S\left(x'\right)\right)$. Then $T(x) = T\left(x'\right)$ so $S(x) = g_2(T(x)) = g_2\left(T\left(x'\right)\right) = S\left(x'\right)$. This proves that g_1 is injective, and the same argument works for g_2. This completes the proof. □

Suppose that $Y_1, \ldots, Y_n$ are i.i.d Poisson (θ).

$$p\left(y_1, \ldots, y_n; \theta\right) = \frac{e^{-n\theta}\theta \sum y_i}{\prod y_i}, \qquad \frac{p\left(y_1, \ldots, y_n; \theta\right)}{p\left(x_1, \ldots, x_n; \theta\right)} = \frac{\theta[\sum y_i - \sum x_i]}{\prod y_i! / \prod x_i!}$$

which is independent of θ iff $\sum y_i = \sum x_i$. This implies that $T\left(X_1, \ldots, X_n\right) = \sum_{i=1}^{n} X_i$ is a minimal sufficient statistic for θ.

↑Exam

↓Exam

Ancillarity

Although a minimal sufficient statistic provides in some sense an optimal degree of data compression, it is still possible for it to contain much "extra" or ancillary material that does not provide by itself any information about θ.

(Ancillarity). A statistic V is ancillary if its distribution does not depend on θ.

Def.

4.2 Completeness

(Completeness). A statistic T is complete for θ if for any θ, $E_\theta[h(T)] = c$ implies $h(T) = c$ with probability one. By subtracting c this condition is equivalent to

$$E_\theta[h(T)] = 0 \quad \forall\ \theta \Rightarrow \quad h(T) = 0\ \ wp.1$$

Def.

T is boundedly complete if the above holds for all bounded funtion h.

Completeness implies that two different functions of the statistics T cannot have the same expectation.

Minimal sufficiency of complete sufficient statistic

Theorem 3. *(The Lehmann-Scheffé Theorem) Suppose X has density $f(x;\theta)$ and $T(X)$ is sufficient and complete for θ. Then T is minimal sufficient.*

A general result was proved by Bahadur.

Theorem 4. *(Bahadur's theorem). If T is a finite-dimensional boundedly complete sufficient statistic, then it is minimal sufficient.*

Basu's Theorem

Theorem 5. *If T is a boundedly complete sufficient statistic for the family of distributions $\mathcal{P}_\theta$ then any ancillary statistic V is independent of T.*

Proof. By ancillarity, $p_C = P(V \in C)$ does not depend on θ for any C. Let $q_C(t) = P(V \in C \mid T = t)$. Then $E_\theta [q_C(T)] = p_C$, and so by completeness, $q_C(T) = p_C$ wp. 1. This proves the independence of V and T. □

Remark 1. *If T is minimal sufficient, R ancillary and R is a function of T (so $R = h(T)$ is not independent of T), then T is not complete.*