

6. Sufficiency and completeness

GULP Winter 2015

February 19, 2015

1 Sufficient statistics

Here we first give the definition of the sufficient statistics.

Definition Given a model parameterized with θ and a sample X , a statistic T is sufficient if the conditional distribution $X|T$ is free of θ .

It is trivial to see that the whole data is sufficient. And in most cases, the order statistics are sufficient, since once given the order statistics, the probability of observing the data is $1/n!$.

Here is another example.

Example Suppose the data is from Poisson distribution with parameter λ . Then the sample sum $T = \sum_i X_i$ is sufficient.

Proof. Suppose we are given $T = t$ and we calculate $\mathbb{P}(X_1 = x_1, \dots, X_n = x_n | T = t)$. We have

$$\begin{aligned}\mathbb{P}(X_1 = x_1, \dots, X_n = x_n | T = t) &= \frac{\mathbb{P}(X_1 = x_1, \dots, X_n = x_n, T = t)}{\mathbb{P}(T = t)} \\ &= \frac{\mathbb{P}(X_1 = x_1, \dots, X_{n-1} = x_{n-1}, X_n = t - \sum_{i=1}^{n-1} x_i)}{\mathbb{P}(T = t)} \\ &= \frac{\frac{\lambda^{x_1}}{x_1!} e^{-\lambda} \times \dots \times \frac{\lambda^{t - \sum_{i=1}^{n-1} x_i}}{(t - \sum_{i=1}^{n-1} x_i)!} e^{-\lambda}}{\frac{(n\lambda)^t}{t!} e^{-n\lambda}} \\ &= \frac{\frac{1}{x_1!} \times \dots \times \frac{1}{(t - \sum_{i=1}^{n-1} x_i)!}}{\frac{n^t}{t!}}\end{aligned}$$

which is free of λ . □

It can be shown that the sample sum is sufficient in binomial, normal and Poisson case. And the maximal of the data is sufficient for uniform distribution $0, \theta$. The sufficiency describes the lossless property of the data compression. If the statistic loses information of the parameter, then the statistic will not be sufficient.

2 Completeness

We first give the definition.

Definition Given a model parameterized with θ and a sample X , a statistic T is complete if there exists some f such that $\mathbb{E}_\theta f(T) = 0$ for all θ , then $f = 0$ almost surely.

On the other side, a statistic containing ancillary information will not be complete. A complete statistic will only contain the information of the parameter.

Example A binomial sample X from the model with parameter (n, p) is complete.

Proof. If some function f satisfies $\mathbb{E}_p f(X) = 0$ for all p , then

$$\sum_{i=1}^n f(i) \binom{n}{i} p^i (1-p)^{n-i} = 0, \forall p.$$

By compare the coefficient of the polynomials, we have $f(i) = 0$ for all i . □

3 Application

The direct application is the following theorem.

Theorem (Lehmann-Scheffe). Any unbiased estimator based on a sufficient and complete statistic is UMVUE.

and it is direct that we have the following.

Theorem (Rao-Blackwell). Suppose U is an unbiased estimator, and T is sufficient and complete, then $\mathbb{E}(U|T)$ is UMVUE.

These theorems are widely used in theory and applications.