planetmath.org

Math for the people, by the people.

sufficient statistic

Canonical name	SufficientStatistic
Date of creation	2013-03-22 15:02:42
Last modified on	2013-03-22 15:02:42
Owner	CWoo (3771)
Last modified by	CWoo (3771)
Numerical id	11
Author	CWoo (3771)
Entry type	Definition
Classification	msc 62B05
Synonym	sufficient estimator
Synonym	minimally sufficient statistic
Synonym	minimal sufficient
Synonym	minimally sufficient
Defines	minimal sufficient statistic
Defines	equivalent statistic

Let $\{f_\theta\}$ be a statistical model with parameter θ. Let $\boldsymbol{X} = (X_1, \dots, X_n)$ be a random vector of random variables representing n observations. A statistic $T = T(\boldsymbol{X})$ of $\boldsymbol{X}$ for the parameter θ is called a *sufficient statistic*, or a *sufficient estimator*, if the conditional probability distribution of $\boldsymbol{X}$ given $T(\boldsymbol{X}) = t$ is not a function of θ (equivalently, does not depend on θ).

In other words, all the information about the unknown parameter θ is captured in the sufficient statistic T. If, say, we are interested in finding out the percentage of defective light bulbs in a shipment of new ones, it is enough, or *sufficient*, to count the number of defective ones (sum of the X_i's), rather than worrying about which individual light bulbs are the defective ones (the vector $(X_1, \dots, X_n)$). By taking the sum, a certain "reduction" of data has been achieved.

Examples

1. Let $X_1, \dots, X_n$ be n independent observations from a uniform distribution on integers $1, \dots, \theta$. Let $T = \max\{X_1, \dots, X_n\}$ be a statistic for θ. Then the conditional probability distribution of $\boldsymbol{X} = (X_1, \dots, X_n)$ given $T = t$ is

$$P(\boldsymbol{X} \mid t) = \frac{P(X_1 = x_1, \dots, X_n = x_n, \max\{X_n\} = t)}{P(\max\{X_n\} = t)}.$$

The numerator is 0 if $\max\{x_n\} \neq t$. So in this case, $P(\boldsymbol{X} \mid t) = 0$ and is not a function of θ. Otherwise, the numerator is θ^{-n} and $P(\boldsymbol{X} \mid t)$ becomes

$$\frac{\theta^{-n}}{P(\max\{X_n\} = t)} = (\theta^n P(X_{(1)} \leq \dots \leq X_{(n)} = t))^{-1},$$

where $X_{(i)}$'s are the rearrangements of the X_i's in a non-decreasing order from $i = 1$ to n. For the denominator, we first note that

$$\begin{aligned} P(X_{(1)} \leq \dots \leq X_{(n)} = t) &= P(X_{(1)} \leq \dots \leq X_{(n)} \leq t) - P(X_{(1)} \leq \dots \leq X_{(n)} < t) \\ &= P(X_{(1)} \leq \dots \leq X_{(n)} \leq t) - P(X_{(1)} \leq \dots \leq X_{(n)} \leq t-1). \end{aligned}$$

From the above equation, we find that there are $t^n - (t-1)^n$ ways to form non-decreasing finite sequences of n positive integers such that the maximum of the sequence is t. So

$$(\theta^n P(X_{(1)} \leq \dots \leq X_{(n)} = t))^{-1} = (\theta^n (t^n - (t-1)^n) \theta^{-n})^{-1} = (t^n - (t-1)^n)^{-1}$$

again is not a function of θ. Therefore, $T = \max\{X_i\}$ is a sufficient statistic for θ. Here, we see that a reduction of data has been achieved by taking only the largest member of set of observations, not the entire set.

2. If we set $T(X_1, \dots, X_n) = (X_1, \dots, X_n)$, then we see that T is trivially a sufficient statistic for *any* parameter θ. The conditional probability distribution of $(X_1, \dots, X_n)$ given T is 1. Even though this is a sufficient statistic by definition (of course, the individual observations provide as much information there is to know about θ as possible), and there is no loss of data in T (which is simply a list of all observations), there is really no reduction of data to speak of here.

3. The sample mean
$$\overline{X} = \frac{X_1 + \cdots + X_n}{n}$$
of n independent observations from a normal distribution $N(\mu, \sigma^2)$ (both μ and σ^2 unknown) is a sufficient statistic for μ. This is the result of the factorization criterion. Similarly, one sees that any partition of the sum of n observations X_i into m subtotals is a sufficient statistic for μ. For instance,
$$T(X_1, \dots, X_n) = (\sum_{i=1}^{j} X_i, \sum_{i=j+1}^{k} X_i, \sum_{i=k+1}^{n} X_i)$$
is a sufficient statistic for μ.

4. Again, assume there are n independent observations X_i from a normal distribution $N(\mu, \sigma^2)$ with unknown mean and variance. The sample variance
$$\frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2$$
is *not* a sufficient statistic for σ^2. However, if μ is a known constant, then
$$\frac{1}{n-1} \sum_{i=1}^{n} (X_i - \mu)^2$$
is a sufficient statistic for σ^2.

A sufficient statistic for a parameter θ is called a *minimal sufficient statistic* if it can be expressed as a function of any sufficient statistic for θ.

Example. In example 3 above, both the sample mean $\overline{X}$ and the finite sum $S = X_1 + \cdots + X_n$ are minimal sufficient statistics for the mean μ. Since, by the factorization criterion, any sufficient statistic T for μ is a vector whose coordinates form a partition of the finite sum, taking the sum of these coordinates is just the finite sum S. So, we have just expressed S as a function of T. Therefore, S is minimal. Similarly, $\overline{X}$ is minimal.

Two sufficient statistics T_1, T_2 for a parameter θ are said to be equivalent provided that there is a bijection g such that $g \circ T_1 = T_2$. $\overline{X}$ and S from the above example are two equivalent sufficient statistics. Two minimal sufficient statistics for the same parameter are equivalent.