Sufficient Statistics

A. Stepanov
D. Simpson
S. Culpepper

Definition (Sufficient Statistics): Let $X_1, X_2, \dots, X_n$ denote random variables with joint pdf or pmf $f(x_1, x_2, \dots, x_n;\ \theta)$, which depends on the parameter θ. The statistic $Y = u(X_1, X_2, \dots, X_n)$ is said to be **sufficient** for θ if the conditional distribution of $X_1, X_2, \dots, X_n$ given $Y = y$ is independent of θ for all *y*.

Theorem 1 (Factorization Theorem): Let $X_1, X_2, \dots, X_n$ denote random variables with joint pdf or pmf $f(x_1, x_2, \dots, x_n;\ \theta)$ which depends on the parameter θ. The statistic $Y = u(X_1, X_2, \dots, X_n)$ is said to be **sufficient** for θ if and only if

$$f(x_1, x_2, \dots, x_n;\ \theta) = g(u(x_1, x_2, \dots, x_n);\ \theta) \cdot h(x_1, x_2, \dots, x_n)$$

where g depends on $x_1, x_2, \dots, x_n$ only through $u(x_1, x_2, \dots, x_n)$ and $h(x_1, x_2, \dots, x_n)$ does not depend on θ.

Remark (Maximum Likelihood Estimation) The MLE estimates θ by maximizing the joint pdf of $X_1, X_2, \dots, X_n$ evaluated at the observed values and considered as a function of θ. It follows from Theorem 1 that the maximum likelihood estimator $\hat{\theta}$ is always function of the sufficient statistic, because it is the maximizer of

$$g(u(x_1, x_2, \dots, x_n);\ \theta).$$

The factor $h(x_1, x_2, \dots, x_n)$ does not affect the maximization with respect to θ.

Example 1. Let $X_1, X_2, \dots, X_n$ be a random sample from Binomial(1, θ), where $\theta \in (0,1)$ is the unknown parameter of interest (θ is the "success" probability).

The pmf for each independent observation is:

$$f(x;\theta) = \theta^x(1-\theta)^{1-x},\ x = 0,1;\ \text{ zero elsewhere}$$

a) Use the Factorization Theorem to find a sufficient statistic $\boldsymbol{Y} = \boldsymbol{u(X_1, X_2, \dots, X_n)}$ for $\boldsymbol{\theta}$.

$$f(x_1, x_2, \dots, x_n;\ \theta) = \theta^{\sum x_i}(1-\theta)^{n-\sum x_i}$$

$$= \theta^y(1-\theta)^{n-y} = g(y;\theta)$$

where $y = u(x_1, \dots, x_n) = \sum x_i$ and each element of $(x_1, x_2, \dots, x_n)$ is either 0 or 1. By the Factorization Theorem $Y = \sum X_i$ is sufficient for θ.

b) Show $\boldsymbol{P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n |\ Y = y)}$ does not depend on θ.

The model above implies that $Y \sim$Binomial(n, θ) (e.g., derive the m.g.f.).

Therefore the conditional probability is

$$\frac{f(x_1, x_2, \dots, x_n;\ \theta)}{f_Y(y)} = \frac{\theta^y(1-\theta)^{n-y}}{\binom{n}{y}\theta^y(1-\theta)^{n-y}} = \frac{1}{\binom{n}{y}}, \quad \text{if } \sum_{i=1}^{n} x_i = y,$$

and all the $x_i = 0$ or ; and the conditional probability equals 0 otherwise.

We see that the conditional distribution assigns equal probabilities to all the possible ways to arrange *y* 1's and *n-y* 0's among the *n* observations.

Sufficient Statistics

A. Stepanov
D. Simpson
S. Culpepper

Example 2. Let $X_1, X_2, \dots, X_n$ be a random sample from Uniform (0, θ). The pdf for each independent random variable is:

$$f(x;\theta) = \frac{1}{\theta}, \quad 0 < x < \theta, \quad \text{zero elsewhere.}$$

a) Use the Factorization Theorem to find a sufficient statistic $Y = u(X_1, X_2, \dots, X_n)$ for θ.

Define $$1\{A\} = \begin{cases} 1, & \text{if } A \text{ is true} \\ 0, & \text{if } A \text{ is false} \end{cases}$$

Then $f(x;\theta) = \frac{1}{\theta} \cdot 1\{x < \theta\} \cdot 1\{x > 0\}$ and the joint pdf is:

$$f(x_1;\theta) f(x_2;\theta) \cdots f(x_n;\theta) = \frac{1}{\theta^n} \prod_{i=1}^{n} 1\{x_i < \theta\} \prod_{j=1}^{n} 1\{x_j > 0\}$$

$$= \frac{1}{\theta^n} \cdot 1\{\max x_i < \theta\} \cdot 1\{\min x_j > 0\}$$

By the Factorization Theorem, $Y = \max X_i$ is sufficient for θ.

b) Show $f(x_1, x_2, \dots, x_n | y)$ does not depend on θ.

Recall from earlier notes that $Y = \max X_i$ has the pdf

$$f_Y(y) = \frac{n\, y^{n-1}}{\theta^n}, 0 < y < \theta, \text{ zero elsewhere.}$$

The joint conditional pdf is therefore given by

$$\frac{f(x_1, x_2, \dots, x_n;\theta)}{f_Y(y)} = \frac{\frac{1}{\theta^n}}{\frac{n\, y^{n-1}}{\theta^n}} = \frac{1}{n y^{n-1}}, \text{ for } 0 < \text{all } x_i \le y,$$

which does not depend on θ.

Sufficient Statistics

A. Stepanov
D. Simpson
S. Culpepper

Theorem 2 (Exponential Family Models): Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution with a pdf or pmf of the exponential form

$$f(x;\theta) = \exp[p(\theta)k(x) + s(x) + q(\theta)]$$

on a support free of θ. Then the statistic $Y = \sum_{i=1}^{n} k(X_i)$ is sufficient for θ.

Example 3. Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from the distribution with probability density function

$$f(x;\theta) = \begin{cases} \dfrac{1}{\theta}\, x^{\frac{1-\theta}{\theta}}, & 0 \le x \le 1 \\ 0, & \text{otherwise} \end{cases}$$

$0 < \theta < \infty$. Use Theorem 2 to find a sufficient statistic for θ.

Method 1: Rearrange the pdf into exponential family form:

$$f(x;\theta) = \exp\left[\frac{1-\theta}{\theta}\ln(x) - \ln(\theta)\right] \qquad k(x) = \ln(x)$$

$\Rightarrow Y = \sum_{i=1}^{n} \ln(X_i)$ is a sufficient statistic for θ

$\Rightarrow W = e^{Y} = \prod_{i=1}^{n} X_i$ is also a sufficient statistic for θ

Method 2: Use the factorization theorem.

$$\prod_{i=1}^{n} f(x_i;\theta) = \frac{1}{\theta^n}\left(\prod_{i=1}^{n} x_i\right)^{\frac{1-\theta}{\theta}}$$

By Factorization Theorem, $W = \prod_{i=1}^{n} X_i$ is a sufficient statistic for θ. It follows that $Y = \ln(W) = \sum_{i=1}^{n} \ln(X_i)$ is also a sufficient statistic for θ.

Example 4. Let $X_1, X_2, \dots, X_n$ be a random sample of size n from a $N(\mu, \sigma^2)$ distribution. Find joint sufficient statistics for μ and σ.

The pdf is

$$f(x_1, x_2, \dots, x_n; \mu, \sigma) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2\sigma^2}(x_i - \mu)^2\right\}$$

$$= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n} x_i^2 + \frac{\mu}{\sigma^2}\sum_{i=1}^{n} x_i - \frac{n\mu^2}{2\sigma^2}\right\}$$

From the Factorization Theorem or vector version of exponential family representation we see that:

$\left(\sum_{i=1}^{n} X_i,\ \ \sum_{i=1}^{n} X_i^2\right)$ is sufficient for (μ, σ)

$\Rightarrow (\bar{X},\ \sum_{i=1}^{n} X_i^2)$ is also sufficient for (μ, σ)

$\Rightarrow (\bar{X},\ \sum_{i=1}^{n} X_i^2 - n\bar{X}^2)$ is also sufficient for (μ, σ)

$\Rightarrow (\bar{X},\ S^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - n\bar{X}^2\right))$ is also sufficient for (μ, σ)

$\Rightarrow (\bar{X},\ S)$ is also sufficient for (μ, σ)