Sufficient Statistics

Travis Andersen

February 5, 2021

A sufficient statistic is a statistic that, if known, will result in the same inference on parameters as the whole sample. That is to say that if $\mathbf{x}$ and $\mathbf{y}$ are two possible samples, such that $T(\mathbf{x}) = T(\mathbf{y})$, then inference on θ is the same regardless of whether $\mathbf{x}$ or $\mathbf{y}$ is sampled.

A sufficient statistic can be found by the factorization theorem, which states that $T(\mathbf{x})$ is a sufficient statistic if and only if functions g and h can be found that satisfy the following equation:

$$f(\mathbf{x}|\theta) = g(T(\mathbf{x})|\theta)h(\mathbf{x})$$

Sufficient statistics also have a relationship with exponential families, where for a random sample of size n from an exponential family with pmf/pdf

$$f(\mathbf{x}|\theta) = h(x)c(\theta)\exp(\sum_{i=1}^{k} w_i(\theta)t_i(x)),$$

$T(\mathbf{X})$ defined as follows is a sufficient statistic for θ.

$$T(\mathbf{X}) = (\sum_{j=1}^{n} t_1(X_j), ..., \sum_{j=1}^{n} t_k(X_j))$$

A minimal sufficient statistic is a sufficient statistic that has been reduced as far as possible. A minimal sufficient statistic must be a function of any other sufficient statistic.