

Definition 1 (Statistical Experiment). A statistical experiment is a triple $(\mathcal{X}, \mathcal{F}, \{\mathbf{P}_\theta\}_{\theta \in \Theta})$ where \mathcal{X} is the sample space, \mathcal{F} is a σ -algebra on \mathcal{X} , and $\{\mathbf{P}_\theta\}$ is a collection of probability measures on \mathcal{X} parametrized by θ in the parameter space Θ .

Definition 2 (Sufficient Statistic). Let $(\mathcal{X}, \mathcal{F}, \{\mathbf{P}_\theta\}_{\theta \in \Theta})$ be an statistical experiment. Let a measurable function $T : (\mathcal{X}, \mathcal{F}) \rightarrow (\mathcal{T}, \mathcal{G})$ be a statistic. The statistic T is said to be sufficient if $\mathbf{P}_\theta(\cdot | T)$ does not depend on θ .

Example 1. Consider an experiment $(\mathcal{X}, \mathcal{F}, \{\mathbf{P}_\theta\}_{\theta \in [0,1]})$ where $\mathcal{X} = \{0, 1\}^n$, \mathcal{F} is the power set of \mathcal{X} , and \mathbf{P}_θ is the joint distribution of n iid Bernoulli(θ) distributions. Define statistic $T(x) := \sum_{i=1}^n x_i$ where x_i denotes the i -th component of x . We check that T is a sufficient statistic: Given any $x \in \mathcal{X}$, let $t := T(x)$ and we have

$$\mathbf{P}_\theta\{X = x | T = t\} = \frac{\mathbf{P}_\theta\{X = x\}}{\mathbf{P}_\theta\{T = t\}} = \frac{\prod_{i=1}^n \theta^{x_i} (1 - \theta)^{1-x_i}}{\binom{n}{t} \theta^t (1 - \theta)^{n-t}} = \binom{n}{t}^{-1}.$$

Since the conditional distribution does not depend on θ , T is a sufficient statistic.

Theorem 1 (Fisher-Neyman Factorization). Consider a statistical experiment $(\mathcal{X}, \mathcal{F}, \{\mathbf{P}_\theta\}_{\theta \in \Theta})$. A statistic $T : (\mathcal{X}, \mathcal{F}) \rightarrow (\mathcal{T}, \mathcal{G})$ is a sufficient statistic iff there exists measurable functions $\{g_\theta\}_{\theta \in \Theta}$ defined on $(\mathcal{T}, \mathcal{G})$ and h defined on $(\mathcal{X}, \mathcal{F})$ such that the probability density function can be decomposed as

$$f(x | \theta) = g_\theta(T(x))h(x). \quad (1)$$

Proof. We only show the proof for discrete random variables. ¹ In this case, we have the probability density function $f(x | \theta) = \mathbf{P}_\theta\{X = x\}$.

(\Rightarrow) Suppose that T is sufficient. Given any realization $x \in \mathcal{X}$, let $t := T(x)$ and we have the following

$$\begin{aligned} \mathbf{P}_\theta\{X = x\} &= \mathbf{P}_\theta\{X = x, T(X) = t\} \\ &= \mathbf{P}_\theta\{X = x | T(X) = t\} \mathbf{P}_\theta\{T(X) = t\} \\ &= \underbrace{\mathbf{P}\{X = x | T(X) = t\}}_{h(x)} \underbrace{\mathbf{P}_\theta\{T(X) = t\}}_{g_\theta(t)} \end{aligned}$$

where the last equality is obtained by the definition of sufficiency.

(\Leftarrow) Suppose we have the factorization (1). Given any realization $x \in \mathcal{X}$, let $t := T(x)$ and we have

$$\begin{aligned} \mathbf{P}_\theta\{X = x | T(X) = t\} &= \frac{\mathbf{P}_\theta\{X = x, T(X) = t\}}{\mathbf{P}_\theta\{T(X) = t\}} \\ &= \frac{g_\theta(t)h(x)}{\mathbf{P}_\theta\{T(X) = t\}} \end{aligned}$$

¹For continuous random variables, one must deal with measure-theoretic technicalities in the proof. One can find a proof for general random variables in *Chapter 2.2 Statistics and Sufficiency* from [1].

where the second equality is obtained by using the factorization on the numerator. Now consider the denominator, we have

$$\mathbf{P}_\theta\{T(X) = t\} = \sum_{x': T(x')=t} \mathbf{P}_\theta\{X = x'\} = \sum_{x': T(x')=t} g_\theta(t)h(x')$$

where the second equality is obtained by factorization. Substitute the result back and we have

$$\begin{aligned} \mathbf{P}_\theta\{X = x \mid T(X) = t\} &= \frac{g_\theta(t)h(x)}{\sum_{x': T(x')=t} g_\theta(t)h(x')} \\ &= \frac{h(x)}{\sum_{x': T(x')=t} h(x')} \end{aligned}$$

where the final expression does not depend on θ . Hence T is a sufficient statistic for θ . #

Example 2 (**Example 1** Continued). Now we check that $T(x) = \sum_{i=1}^n x_i$ is a sufficient statistic using **Theorem 1**:

$$\mathbf{P}_\theta\{X = x\} = \prod_{i=1}^n \theta^{x_i} (1 - \theta)^{1-x_i} = \underbrace{\theta^{T(x)} (1 - \theta)^{n-T(x)}}_{g_\theta(T(x))}.$$

In this case, $h(x) = 1$ is a constant function. Hence, T is a sufficient statistic.

Definition 3 (Sufficiency Principle). Let $(\mathcal{X}, \mathcal{F}, \{\mathbf{P}_\theta\}_{\theta \in \Theta})$ be a statistical experiment and let T be a sufficient statistic. The sufficiency principle states that inference about θ should depend only on T . That is, if two samples x and x' satisfies $T(x) = T(x')$, then they should lead to the same inference about θ .

Remark 1. The concept of sufficient statistic and the sufficiency principle are both due to Sir Ronald Fisher in the early 20th century.

References

- [1] Jun Shao. *Mathematical Statistics*. New York: Springer, 1999. ISBN: 0-387-98674-X.