SUFFICIENT STATISTICS

Suppose the distribution of a random variable $\underline{X}$ is determined by a parameter $\underline{\theta}$:

$$\underline{X} \sim f_{\underline{\theta}}(\underline{x})$$

The functional form of f is known, but $\underline{\theta}$ is unknown.

[Note: We will use f to denote a pdf, p to denote a pmf, and f when it could be either.]

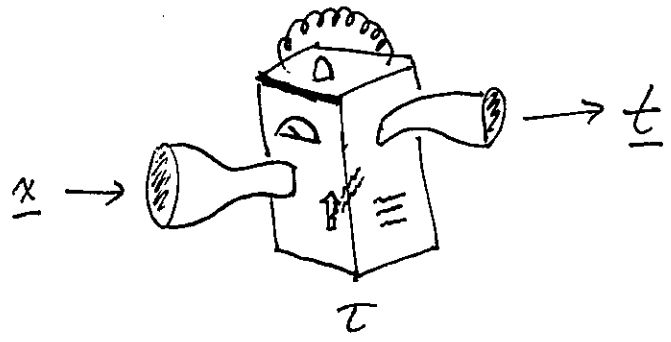
In statistical inference, we observe a realization $\underline{x}$ of $\underline{X}$ and need to answer some question about $\underline{\theta}$, such as

- Is $\underline{\theta} \in \textcircled{H}_1$ or is $\underline{\theta} \in \textcircled{H}_2$? (Detection)
- What is a good guess for $\underline{\theta}$? (Estimation)

If $\underline{x} = [x_1 \cdots x_N]^T$ and $\underline{\theta} = [\theta_1 \cdots \theta_p]^T$ where $p < N$, one might wonder whether it is possible to <u>compress</u> the measurement $\underline{x}$ into a low-dimensional statistic without affecting the quality of the inference about $\underline{\theta}$.

In other words, does there exist $\underline{T} = \tau(\underline{X})$, where the dimension of $\underline{T}$ is $M < N$, such that $\underline{T}$ carries all the useful information about $\underline{\theta}$?

If so, for the purpose of studying $\underline{\theta}$, we could discard the raw measurement $\underline{x}$ and retain only the compressed statistic $\underline{t}$.



Definition Let $\underline{X} \sim f_{\underline{\theta}}(\underline{x})$. The statistic

$\underline{T} = \tau(\underline{X})$ is a <u>sufficient statistic</u> for $\underline{\theta}$ if the conditional distribution of $\underline{X}$ given $\underline{T}$ is independent of $\underline{\theta}$. Equivalently, the functional form of $f(\underline{x}|\underline{t})$ does not involve $\underline{\theta}$.

Interpretations

1. Let $P_{\underline{\theta}}(\underline{x}, \underline{t})$ denote the joint pmf of $(\underline{X}, \underline{T})$. Then

$$P_{\underline{\theta}}(\underline{x}, \underline{t}) = \begin{cases} P_{\underline{\theta}}(\underline{x}) & \text{if } \underline{t} = \tau(\underline{x}) \\ 0 & \text{otherwise} \end{cases}$$

Therefore

$$\begin{aligned} P_{\underline{\theta}}(\underline{x}) &= P_{\underline{\theta}}(\underline{x}, \tau(\underline{x})) \\ &= P_{\underline{\theta}}(\underline{x} \,|\, \tau(\underline{x})) P_{\underline{\theta}}(\tau(\underline{x})) \\ &= p(\underline{x} \,|\, \tau(\underline{x})) P_{\underline{\theta}}(\tau(\underline{x})) \end{aligned}$$

$\Rightarrow$ the dependence of $P_{\underline{\theta}}(\underline{x})$ on θ is manifested entirely in $P_{\underline{\theta}}(\underline{t})$.

[continuous case requires more care, but same conclusion holds – see Scharf]

2. Given $\underline{t} = \tau(\underline{x})$, full knowledge of $\underline{x}$ brings no additional information about $\underline{\theta}$.

3. Any inference strategy based on $f_{\underline{\theta}}(\underline{x})$ may be replaced by a strategy based on $f_{\underline{\theta}}(\underline{t})$.

Example] Bernoulli trials

Suppose we observe $\underline{x} = [x_1 \cdots x_N]^T$ where

$$X_i \overset{iid}{\sim} \text{Bernoulli}(\theta)$$

$\theta \in [0,1]$ is unknown.

Recall

(a) $$p_\theta(x) =$$

Since we can assume $x_i \in \{0,1\}$, we may write

$$p_\theta(x_i) = \theta^{x_i}(1-\theta)^{1-x_i}$$

Therefore

$$\begin{aligned} p_\theta(\underline{x}) &= \prod_{i=1}^{n} p_\theta(x_i) \\ &= \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i} \\ &= \theta^k (1-\theta)^{N-k} \end{aligned}$$

where $k = \sum_{i=1}^{n} x_i$.

Claim: K is a sufficient statistic for θ.

We must show $p_\theta(\underline{x}/k)$ is independent of θ.

From interpretation #1 we know

$$p_\theta(\underline{x}/k) = \frac{P_\theta(\underline{x})}{P_\theta(k)}.$$

Exercise | Complete this argument to establish that K is sufficient for θ.

Solution | K is a Binomial (N, θ) random variable.

Therefore

$$P_\theta(k) = \binom{N}{k} \theta^k (1-\theta)^{N-k}$$

and

$$p(\underline{x} \mid k) = \frac{P_\theta(\underline{x})}{P_\theta(k)}$$

$$= \frac{\theta^k (1-\theta)^{N-k}}{\binom{N}{k} \theta^k (1-\theta)^{N-k}}$$

$$= \frac{1}{\binom{N}{k}}$$

which is independent of θ.

The Fisher-Neyman Factorization Theorem

In the previous example, we had to guess the sufficient statistic and work out the conditional pmf by hand. In general, it is difficult to verify the definition of sufficient statistic directly.

The following theorem allows us to identify and verify sufficient statistics more readily, and can be taken as a working definition of sufficiency.

Theorem Let $f_{\underline{\theta}}(\underline{x})$ be the density or mass function for $\underline{X}$. The statistic $\underline{T} = \tau(\underline{X})$ is sufficient for $\underline{\theta}$ iff there exist functions $g_{\underline{\theta}}(\underline{t})$ and $h(\underline{x})$ such that

$$f_{\underline{\theta}}(\underline{x}) = g_{\underline{\theta}}(\tau(\underline{x})) \cdot h(\underline{x})$$

Note: h is independent of $\underline{\theta}$.

Example Bernoulli trials revisited

$$P_\theta(\underline{x}) = \prod_{i=1}^{N} \theta^{x_i}(1-\theta)^{1-x_i}$$

$$= \theta^{k}(1-\theta)^{N-k}$$

$$= g_\theta(k) \cdot h(\underline{x})$$

where

$$g_\theta(k) = \theta^{k}(1-\theta)^{N-k}$$

$$h(\underline{x}) = 1$$

$\Longrightarrow$ K is sufficient for θ.

Proof of Theorem

We will assume $\underline{X}$ is discrete. The continuous case is slightly more involved - see Scharf or Kay, vol I.

First, assume $\underline{T}$ is sufficient for $\underline{\theta}$. Recall

$$P_{\underline{\theta}}(\underline{x}) = p(\underline{x}/\tau(\underline{x})) \cdot P_{\underline{\theta}}(\tau(\underline{x}))$$

Now take

$$g_{\underline{\theta}}(\underline{t}) = P_{\underline{\theta}}(\underline{t})$$

$$h(\underline{x}) = p(\underline{x}/\tau(\underline{x}))$$

← Independent of $\underline{\theta}$ by sufficiency

For the other direction, assume $p_{\underline{\theta}}(\underline{x})$ may be written

$$p_{\underline{\theta}}(\underline{x}) = g_{\underline{\theta}}(\tau(\underline{x}))h(\underline{x}).$$

We need to show $\underline{T} = \tau(\underline{X})$ is sufficient for $\underline{\theta}$. That is, we need to show $p_{\underline{\theta}}(\underline{x} \mid \underline{t})$ is independent of $\underline{\theta}$.

Again, we will rely on the identity

$$p_{\underline{\theta}}(\underline{x} \mid \underline{t}) = \frac{p_{\underline{\theta}}(\underline{x})}{p_{\underline{\theta}}(\underline{t})}.$$

Since $\underline{X}$ and $\underline{T}$ are discrete, we have

$$p_{\underline{\theta}}(\underline{t}) = \sum_{\underline{x}' : \tau(\underline{x}') = \underline{t}} p_{\underline{\theta}}(\underline{x}').$$

Therefore

$$p_{\underline{\theta}}(\underline{x} \mid \underline{t}) = \frac{g_{\underline{\theta}}(\underline{t}) \cdot h(\underline{x})}{\sum_{\underline{x}' : \tau(\underline{x}') = \underline{t}} g_{\theta}(\underline{t})\, h(\underline{x}')}$$

$$= \frac{h(\underline{x})}{\sum_{\underline{x}' : \tau(\underline{x}') = \underline{t}} h(\underline{x}')} \leftarrow \boxed{\text{Independent of } \underline{\theta}}$$

∎

Note! The FNFT gives us a formula for $P(\underline{x}|\underline{t})$, namely

$$P(\underline{x}|\underline{t}) = \frac{h(\underline{x})}{\sum_{\underline{x}': \tau(\underline{x}') = \underline{t}} h(\underline{x}')}$$

Example! Bernoulli trials, part III

$h(\underline{x}) = 1$, so

$$p(\underline{x}|k) = \frac{1}{\sum_{\underline{x}': \sum_{i=1}^{N} x_i' = k} 1}$$

$$= \frac{1}{\#\{\underline{x}': \sum_{i=1}^{N} x_i' = k\}}$$

$$= \frac{1}{\binom{N}{k}}$$

Let's look at an example of the FNFT for the continuous case:

Example Gaussian with unknown mean

We are given $\underline{x} = [x_1 \cdots x_N]^T$ where

$$x_i \overset{iid}{\sim} N(\theta, \sigma^2)$$

and σ^2 is known.

$$f_\theta(\underline{x}) = \prod_{i=1}^{N} f_\theta(x_i)$$

$$= \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x_i-\theta)^2}{2\sigma^2}\right\}$$

$$= (2\pi\sigma^2)^{-\frac{N}{2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{N}(x_i-\theta)^2\right\}$$

$$= (2\pi\sigma^2)^{-\frac{N}{2}} \exp\left\{\frac{-1}{2\sigma^2}\left[\sum_{i=1}^{N} x_i^2 - 2\theta\sum_{i=1}^{N} x_i + N\theta^2\right]\right\}$$

(b)

where $t =$ $\Rightarrow$ t is sufficient for θ.

Example Gaussian w/ unknown mean and variance.

Now assume

$$x_i \overset{iid}{\sim} N(\theta_1, \theta_2), \quad i = 1, \ldots, N$$

where $\underline{\theta} = [\theta_1 \ \theta_2]^T$ is unknown. Then

$$f_{\underline{\theta}}(\underline{x}) = \underbrace{\left(\frac{1}{2\pi\theta_2}\right)^{\frac{N}{2}} \exp\left\{-\frac{1}{2\theta_2}\left[\sum_{i=1}^{N} x_i^2 - 2\theta_1 \sum_{i=1}^{N} x_i + N\theta_1^2\right]\right\}}_{g_{\underline{\theta}}(\underline{t})} \cdot \underbrace{1}_{h(\underline{x})}$$

where $\underline{t} = \left[\sum_{i=1}^{N} x_i \ \ \sum_{i=1}^{N} x_i^2\right]^T$ is sufficient.

If an invertible function is applied to a sufficient statistic, the result is again a sufficient statistic.

For example, in the Gaussian iid model:

- $\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$ is sufficient for θ_1
- $\left[\bar{x} \quad s^2\right]^T$ is sufficient for $[\theta_1, \theta_2]^T$

  ↳ $= \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2$

Summary

- $\underline{T}$ is sufficient for $\underline{\theta} \iff \underline{T}$ contains all the information about $\underline{\theta}$ present in $\underline{X}$.
- FNFT: $\underline{T}$ is sufficient for $\underline{\theta} \iff$

$$f_{\underline{\theta}}(\underline{x}) = g_{\underline{\theta}}(\tau(\underline{x})) \cdot h(\underline{x})$$

Key

a.
$$\begin{cases} \theta & \text{if } x = 1 \\ 1-\theta & \text{if } x = 0 \\ 0 & \text{else} \end{cases}$$

b.
$$\underbrace{\exp\left\{\frac{-1}{2\sigma^2}\left[-2\theta \sum_{i=1}^{N} x_i + N\theta^2\right]\right\}}_{g_\theta(t)} \underbrace{(2\pi\sigma^2)^{-\frac{N}{2}} \exp\left\{\frac{-\sum x_i^2}{2\sigma^2}\right\}}_{h(\underline{x})}$$

$$t = \sum_{i=1}^{N} x_i$$