Lecture 02: Sufficient and Minimal Sufficient Statistics

Mathematical Statistics II, MATH 60062/70062

Tuesday January 25, 2022

Reference: Casella & Berger, 6.2.1-6.2.2

Sample and statistic

Suppose that $X_1, \ldots, X_n$ is an iid sample. A **statistic**,

$$T = T(\boldsymbol{X}) = T(X_1, \ldots, X_n),$$

is a function of the sample $\boldsymbol{X} = (X_1, \ldots, X_n)$. The only restriction is that T cannot depend on unknown parameters.

The statistic T forms a **partition** of $\mathcal{X}$, the support of $\boldsymbol{X}$. Specifically, T partitions $\mathcal{X} \subseteq \mathbb{R}^n$ into sets

$$A_t = \{\boldsymbol{x} \in \mathcal{X} : T(\boldsymbol{x}) = t\},$$

for $t \in \mathcal{T}$. All points in A_t are treated the same if we are interested in T only.

Data reduction

The statistic T summarizes the data $\boldsymbol{X}$ in that one can report

$$T(\boldsymbol{x}) = t \iff \boldsymbol{x} \in A_t$$

instead of reporting $\boldsymbol{x}$ itself. Thus, T provides a **data reduction**. The data $\boldsymbol{x}$ are reduced in a way to be more easily understood without losing the *meaning* associated with the set of observations.

In **statistical inference**, suppose $X_1, \ldots, X_n$ is an iid sample from $f_X(x \mid \theta)$, where $\theta \in \Theta$. We would like to use the sample $\boldsymbol{X}$ to learn about which member (or members) of this family might be reasonable. We are interested in statistics T that reduce the data $\boldsymbol{X}$ while capturing all the information about θ contained in the sample.

Sufficient statistic

A statistic $T = T(\boldsymbol{X})$ is a **sufficient statistic** for a parameter θ if it captures "all of the information" about θ contained in the sample. In other words, we do not lose any information about θ by reducing the sample $\boldsymbol{X}$ to the statistic T.

Formally, a statistic $T(\boldsymbol{X})$ is sufficient for θ if the conditional distribution of $\boldsymbol{X}$ given T does not depend on θ; i.e., the ratio

$$f_{\boldsymbol{X}|T}(\boldsymbol{x} \mid t) = \frac{f_{\boldsymbol{X}}(\boldsymbol{x} \mid \theta)}{f_T(t \mid \theta)}$$

is free of θ, for all $\boldsymbol{x} \in \mathcal{X}$. This means, *after conditioning on T, we have removed all information about θ from the sample $\boldsymbol{X}$.*

Binomial sufficient statistic

Suppose $X_1, \ldots, X_n$ are iid $\mathrm{Bern}(\theta)$ with parameter $0 < \theta < 1$. Then $T(\boldsymbol{X}) = X_1 + \cdots + X_n$ is a sufficient statistic for θ.

The PMF of $\boldsymbol{X}$ is given by

$$f_{\boldsymbol{X}}(\boldsymbol{x} \mid \theta) = \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i}.$$

Note that $T(\boldsymbol{X})$ counts the number of X_i's that equal 1, so $T(\boldsymbol{X})$ has a $\mathrm{Bin}(n, \theta)$ distribution,

$$f_T(t \mid \theta) = \binom{n}{t}\theta^t(1-\theta)^{n-t}.$$

With $t = \sum_{i=1}^n x_i$, the conditional distribution

$$f_{\boldsymbol{X}|T}(\boldsymbol{x} \mid t) = \frac{f_{\boldsymbol{X}}(\boldsymbol{x} \mid \theta)}{f_T(t \mid \theta)} = \frac{\theta^t(1-\theta)^{n-t}}{\binom{n}{t}\theta^t(1-\theta)^{n-t}} = \frac{1}{\binom{n}{\sum x_i}},$$

which is free of θ. Therefore, $T(\boldsymbol{X}) = \sum_{i=1}^n X_i$ is a sufficient statistic.

Sufficient order statistics

Suppose $X_1, \ldots, X_n$ are iid from a continuous distribution with PDF $f_X(x \mid \theta)$, where $\theta \in \Theta$. The vector of order statistics, $\boldsymbol{T} = \boldsymbol{T}(\boldsymbol{X}) = (X_{(1)}, \ldots, X_{(n)})$, is always sufficient.

The joint distribution of the n order statistics is

$$\begin{aligned} f_{X_{(1)},\ldots,X_{(n)}}(x_1, \ldots, x_n \mid \theta) &= n! f_X(x_1 \mid \theta) \ldots f_X(x_n \mid \theta) \\ &= n! f_{\boldsymbol{X}}(\boldsymbol{x} \mid \theta), \end{aligned}$$

for $-\infty < x_1 < \cdots < x_n < \infty$. Therefore, the ratio

$$\frac{f_{\boldsymbol{X}}(\boldsymbol{x} \mid \theta)}{f_{\boldsymbol{T}}(\boldsymbol{t} \mid \theta)} = \frac{f_{\boldsymbol{X}}(\boldsymbol{x} \mid \theta)}{n! f_{\boldsymbol{X}}(\boldsymbol{x} \mid \theta)} = \frac{1}{n!},$$

which is free of θ. So $\boldsymbol{T} = \boldsymbol{T}(\boldsymbol{X}) = (X_{(1)}, \ldots, X_{(n)})$ is a sufficient statistic.

Sufficient order statistics

- Reducing the sample $\boldsymbol{X} = (X_1, \ldots, X_n)$ to $\boldsymbol{T}(\boldsymbol{X}) = (X_{(1)}, \ldots, X_{(n)})$ is not much of a reduction.
- However, in some parametric families (e.g., Cauchy, Logistic, etc.), it is not possible to reduce $\boldsymbol{X}$ any further without losing information about θ.
- In some situations, it may be that the parametric form of $f_X(x \mid \theta)$ is not specified. We should not expect more with so little information provided about the population.

Factorization Theorem

So far, we've used the definition of sufficiency directly by showing that the conditional distribution of $\boldsymbol{X}$ given T is free of θ. What if we need to find a sufficient statistic?

Factorization Theorem: A statistic $T = T(\boldsymbol{X})$ is **sufficient** for θ if and only if there exists functions $g(t \mid \theta)$ and $h(\boldsymbol{x})$ such that

$$f_{\boldsymbol{X}}(\boldsymbol{x} \mid \theta) = g(t \mid \theta)h(\boldsymbol{x}),$$

for all sample points $\boldsymbol{x} \in \mathcal{X}$ and all $\theta \in \Theta$.

Poisson sufficient statistic

Suppose $X_1, \ldots, X_n$ are iid $\mathrm{Pois}(\theta)$, where $\theta > 0$,

$$f_X(x \mid \theta) = \frac{\theta^x e^{-\theta}}{x!}.$$

The PMF of $\boldsymbol{X}$ is

Poisson sufficient statistic

Suppose $X_1, \ldots, X_n$ are iid $\mathrm{Pois}(\theta)$, where $\theta > 0$,

$$f_X(x \mid \theta) = \frac{\theta^x e^{-\theta}}{x!}.$$

The PMF of $\boldsymbol{X}$ is

$$\begin{aligned} f_{\boldsymbol{X}}(\boldsymbol{x} \mid \theta) &= \prod_{i=1}^{n} \frac{\theta^{x_i} e^{-\theta}}{x_i!} \\ &= \frac{\theta^{\sum_{i=1}^n x_i} e^{-n\theta}}{\prod_{i=1}^n x_i!} \\ &= \underbrace{\theta^{\sum_{i=1}^n x_i} e^{-n\theta}}_{g(t|\theta)} \underbrace{\frac{1}{\prod_{i=1}^n x_i!}}_{h(\boldsymbol{x})}, \end{aligned}$$

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

By the Factorization Theorem, $T = T(\boldsymbol{X}) = \sum_{i=1}^n X_i$ is sufficient.

Uniform sufficient statistic

Suppose $X_1, \ldots, X_n$ are iid $\mathrm{Unif}(0, \theta)$, where $\theta > 0$. The PMF of $\boldsymbol{X}$ is

$$\begin{aligned} f_{\boldsymbol{X}}(\boldsymbol{x} \mid \theta) &= \prod_{i=1}^{n} \frac{1}{\theta} I(0 < x_i < \theta) \\ &= \frac{1}{\theta^n} \prod_{i=1}^{n} I(0 < x_i < \theta) \\ &= \underbrace{\frac{1}{\theta^n} I(x_{(n)} < \theta)}_{g(t|\theta)} \underbrace{\prod_{i=1}^{n} I(x_i > 0)}_{h(\boldsymbol{x})}, \end{aligned}$$

where $t = x_{(n)}$.

By the Factorization Theorem, $T = T(\boldsymbol{X}) = X_{(n)}$ is sufficient.

Sufficient statistics in the Exponential family

Suppose $X_1, \dots, X_n$ are iid from the **Exponential family**

$$f_X(x \mid \boldsymbol{\theta}) = h(x)c(\boldsymbol{\theta}) \exp\left(\sum_{j=1}^{k} w_j(\boldsymbol{\theta}) t_j(x)\right),$$

where $\boldsymbol{\theta} = (\theta_1, \dots, \theta_d)$, $d \leq k$. Then

$$\boldsymbol{T} = \boldsymbol{T}(\boldsymbol{X}) = \left(\sum_{i=1}^{n} t_1(X_i), \sum_{i=1}^{n} t_2(X_i), \dots, \sum_{i=1}^{n} t_k(X_i)\right)$$

is sufficient for $\boldsymbol{\theta}$.

Sufficient statistics in the Exponential family

Use the Factorization Theorem. The PDF of $\boldsymbol{X}$ is

$$\begin{aligned} f_{\boldsymbol{X}}(\boldsymbol{x} \mid \boldsymbol{\theta}) &= \prod_{i=1}^{n} h(x_i) c(\boldsymbol{\theta}) \exp\left(\sum_{j=1}^{k} w_j(\boldsymbol{\theta}) t_j(x_i)\right) \\ &= \left(\prod_{i=1}^{n} h(x_i)\right) [c(\boldsymbol{\theta})]^n \exp\left(\sum_{j=1}^{k} w_j(\boldsymbol{\theta}) \sum_{i=1}^{n} t_j(x_i)\right) \\ &= h^*(\boldsymbol{x}) g(t_1^*, t_2^*, \ldots, t_k^* \mid \boldsymbol{\theta}), \end{aligned}$$

where $t_j^* = \sum_{i=1}^{n} t_j(x_i)$ for $j = 1, \ldots, k$.

Binomial sufficient statistic

Suppose $X_1, \ldots, X_n$ are iid $\mathrm{Bern}(\theta)$ with parameter $0 < \theta < 1$.
For $x = 0, 1$, the PMF of X is

Binomial sufficient statistic

Suppose $X_1, \ldots, X_n$ are iid $\mathrm{Bern}(\theta)$ with parameter $0 < \theta < 1$.
For $x = 0, 1$, the PMF of X is

$$
\begin{aligned}
f_X(x \mid \theta) &= \theta^x (1-\theta)^{1-x} \\
&= (1-\theta)\left(\frac{\theta}{1-\theta}\right)^x \\
&= (1-\theta)\exp\left(\log\left(\frac{\theta}{1-\theta}\right) x\right) \\
&= \underbrace{h(x)}_{1}\,\underbrace{c(\theta)}_{1-\theta}\exp[\,\underbrace{w_1(\theta)}_{\log\left(\frac{\theta}{1-\theta}\right)}\;\underbrace{t_1(x)}_{x}].
\end{aligned}
$$

Binomial sufficient statistic

Suppose $X_1, \ldots, X_n$ are iid $\mathrm{Bern}(\theta)$ with parameter $0 < \theta < 1$. For $x = 0, 1$, the PMF of X is

$$
\begin{aligned}
f_X(x \mid \theta) &= \theta^x (1-\theta)^{1-x} \\
&= (1-\theta)\left(\frac{\theta}{1-\theta}\right)^x \\
&= (1-\theta)\exp\left(\log\left(\frac{\theta}{1-\theta}\right)x\right) \\
&= \underbrace{h(x)}_{1}\underbrace{c(\theta)}_{1-\theta}\exp[\ \underbrace{w_1(\theta)}_{\log\left(\frac{\theta}{1-\theta}\right)}\ \underbrace{t_1(x)}_{x}].
\end{aligned}
$$

Therefore,

$$
T = T(\boldsymbol{X}) = \sum_{i=1}^n t_1(X_i) = \sum_{i=1}^n X_i
$$

is sufficient.

There are many sufficient statistics in any problem

The complete sample, $\boldsymbol{X}$, is a sufficient statistic, since

$$f_{\boldsymbol{X}}(\boldsymbol{x} \mid \theta) = g(T(\boldsymbol{x}) \mid \theta)h(\boldsymbol{x}),$$

where $T(\boldsymbol{x}) = \boldsymbol{x}$, $g(\boldsymbol{x} \mid \theta) = f_{\boldsymbol{X}}(\boldsymbol{x} \mid \theta)$, and $h(\boldsymbol{x}) = 1$ for all $\boldsymbol{x}$.

Any one-to-one function of a sufficient statistic is a sufficient statistic. Suppose $T = T(\boldsymbol{X})$ is sufficient, and define $T^*(\boldsymbol{X}) = r(T(\boldsymbol{X}))$, where r is a one-to-one function with inverse r^{-1}. Then

$$\begin{aligned} f_{\boldsymbol{X}}(\boldsymbol{x} \mid \theta) &= g(T(\boldsymbol{x}) \mid \theta)h(\boldsymbol{x}) \\ &= g(r^{-1}(T^*(\boldsymbol{x})) \mid \theta)h(\boldsymbol{x}) \\ &= g^{-1}(T^*(\boldsymbol{x}) \mid \theta)h(\boldsymbol{x}), \end{aligned}$$

where g^{-1} is the composition of g and r^{-1}.

Normal sufficient statistics

Suppose $X_1, \ldots, X_n$ are iid $\mathcal{N}(\mu, \sigma_0^2)$, where $-\infty < \mu < \infty$ and σ_0^2 is known. Each of the following statistics is sufficient:

- $T_1(\boldsymbol{X}) = \bar{X}$
- $\boldsymbol{T}_2(\boldsymbol{X}) = (X_1, \sum_{i=2}^n X_i)$
- $\boldsymbol{T}_3(\boldsymbol{X}) = (X_{(1)}, \ldots, X_{(n)})$
- $\boldsymbol{T}_4(\boldsymbol{X}) = \boldsymbol{X}$

How much data reduction is possible?

Minimal sufficient statistics

A statistic $T = T(\boldsymbol{X})$ is a **minimal sufficient statistic** for a parameter θ if, for any other sufficient statistic $T^*(\boldsymbol{X})$, $T(\boldsymbol{x})$ is a function of $T^*(\boldsymbol{x})$.

This means that if you know $T^*(\boldsymbol{x})$, you can calculate $T(\boldsymbol{x})$, and

$$T^*(\boldsymbol{x}) = T^*(\boldsymbol{y}) \implies T(\boldsymbol{x}) = T(\boldsymbol{y}).$$

A minimal sufficient statistic achieves the *greatest possible data reduction*. In terms of partition sets formed by statistics, a minimal sufficient statistic admits the coarsest possible partition.

Minimal sufficient statistic

Using the definition to find a minimal sufficient statistic is impractical. The following result by Lehmann and Scheffé gives an easier way to find a minimal sufficient statistic.

Suppose $\boldsymbol{X} \sim f_{\boldsymbol{X}}(\boldsymbol{x} \mid \theta)$, where $\theta \in \Theta$. Suppose there exists a function $T(\boldsymbol{x})$ such that, for all $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{X}$,

$$\frac{f_{\boldsymbol{X}}(\boldsymbol{x} \mid \theta)}{f_{\boldsymbol{X}}(\boldsymbol{y} \mid \theta)} \text{ is free of } \theta \iff T(\boldsymbol{x}) = T(\boldsymbol{y}).$$

Then $T(\boldsymbol{X})$ is a minimal sufficient statistic.

Normal minimal sufficient statistic

Suppose $X_1, \dots, X_n$ are iid $\mathcal{N}(\mu, \sigma_0^2)$, where $-\infty < \mu < \infty$ and σ_0^2 is known. The PDF of $\boldsymbol{X}$ is

$$
\begin{aligned}
f_{\boldsymbol{X}}(\boldsymbol{x} \mid \theta) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-(x_i-\mu)^2/2\sigma_0^2} \\
&= \left(\frac{1}{\sqrt{2\pi\sigma_0^2}} \right)^n e^{-\sum_{i=1}^{n}(x_i-\mu)^2/2\sigma_0^2},
\end{aligned}
$$

where

$$
\sum_{i=1}^{n}(x_i - \mu)^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2 + n(\bar{x} - \mu)^2.
$$

The ratio

$$
\frac{f_{\boldsymbol{X}}(\boldsymbol{x} \mid \theta)}{f_{\boldsymbol{X}}(\boldsymbol{y} \mid \theta)} = \frac{\left(\frac{1}{\sqrt{2\pi\sigma_0^2}} \right)^n \exp\left[-\left(\sum_{i=1}^{n}(x_i - \bar{x})^2 + n(\bar{x} - \mu)^2 \right) / 2\sigma_0^2 \right]}{\left(\frac{1}{\sqrt{2\pi\sigma_0^2}} \right)^n \exp\left[-\left(\sum_{i=1}^{n}(y_i - \bar{y})^2 + n(\bar{y} - \mu)^2 \right) / 2\sigma_0^2 \right]},
$$

is free of μ if and only if $\bar{x} = \bar{y}$. Therefore, $T(\boldsymbol{X}) = \bar{X}$ is a **minimal sufficient statistic**.

Uniform minimal sufficient statistic

Suppose $X_1, \ldots, X_n$ are iid $\mathrm{Unif}(\theta, \theta+1)$, where $-\infty < \theta < \infty$. The PDF of $\boldsymbol{X}$ is

$$f_{\boldsymbol{X}}(\boldsymbol{x} \mid \theta) = I(x_{(1)} > \theta) I(x_{(n)} < \theta + 1) \prod_{i=1}^{n} I(x_i \in \mathbb{R}).$$

The ratio

$$\frac{f_{\boldsymbol{X}}(\boldsymbol{x} \mid \theta)}{f_{\boldsymbol{X}}(\boldsymbol{y} \mid \theta)} = \frac{I(x_{(1)} > \theta) I(x_{(n)} < \theta + 1) \prod_{i=1}^{n} I(x_i \in \mathbb{R})}{I(y_{(1)} > \theta) I(y_{(n)} < \theta + 1) \prod_{i=1}^{n} I(y_i \in \mathbb{R})},$$

is free of θ if and only if $(x_{(1)}, x_{(n)}) = (y_{(1)}, y_{(n)})$. Therefore, $\boldsymbol{T}(\boldsymbol{X}) = (X_{(1)}, X_{(n)})$ is a minimal sufficient statistic.

Uniform minimal sufficient statistic

Suppose $X_1, \ldots, X_n$ are iid $\mathrm{Unif}(\theta, \theta+1)$, where $-\infty < \theta < \infty$. The PDF of $\boldsymbol{X}$ is

$$f_{\boldsymbol{X}}(\boldsymbol{x} \mid \theta) = I(x_{(1)} > \theta) I(x_{(n)} < \theta + 1) \prod_{i=1}^{n} I(x_i \in \mathbb{R}).$$

The ratio

$$\frac{f_{\boldsymbol{X}}(\boldsymbol{x} \mid \theta)}{f_{\boldsymbol{X}}(\boldsymbol{y} \mid \theta)} = \frac{I(x_{(1)} > \theta) I(x_{(n)} < \theta + 1) \prod_{i=1}^{n} I(x_i \in \mathbb{R})}{I(y_{(1)} > \theta) I(y_{(n)} < \theta + 1) \prod_{i=1}^{n} I(y_i \in \mathbb{R})},$$

is free of θ if and only if $(x_{(1)}, x_{(n)}) = (y_{(1)}, y_{(n)})$. Therefore, $\boldsymbol{T}(\boldsymbol{X}) = (X_{(1)}, X_{(n)})$ is a minimal sufficient statistic.

- Any one-to-one function of a minimal sufficient statistic is also a minimal sufficient statistic.
- So, $\boldsymbol{T}^*(\boldsymbol{X}) = (X_{(n)} - X_{(1)}, (X_{(1)} + X_{(n)})/2)$ is also a minimal sufficient statistic.