Lecture 19: Sufficiency

Mathematical Statistics I, MATH 60061/70061

Tuesday November 30, 2021

Reference: Casella & Berger, 6.1-6.2.1

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 $T(X)$ is sufficient for θ if the conditional distribution of 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.

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.

Poisson sufficient statistic

Suppose $X_1, \ldots, X_n$ are iid $\mathrm{Pois}(\theta)$, where $\theta > 0$. Then $T(\boldsymbol{X}) = X_1 + \cdots + X_n$ is a sufficient statistic for θ.

Poisson sufficient statistic

Suppose $X_1, \ldots, X_n$ are iid $\mathrm{Pois}(\theta)$, where $\theta > 0$. Then $T(\boldsymbol{X}) = X_1 + \cdots + X_n$ is a sufficient statistic for θ.

The PMF of $\boldsymbol{X}$, for $x_i = 0, 1, 2, \ldots$, is given by

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

Recall that $T(\boldsymbol{X}) \sim \mathrm{Pois}(n\theta)$ (Lecture 5). Thus, the PMF of T, for $t = 0, 1, 2, \ldots$, is

$$f_T(t \mid \theta) = \frac{(n\theta)^t e^{-n\theta}}{t!}.$$

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

$$f_{\boldsymbol{X}|T}(\boldsymbol{x} \mid t) = \frac{f_{\boldsymbol{X}}(\boldsymbol{x} \mid \theta)}{f_T(t \mid \theta)} = \frac{\frac{\theta^{\sum_{i=1}^n x_i} e^{-n\theta}}{\prod_{i=1}^n x_i!}}{\frac{(n\theta)^t e^{-n\theta}}{t!}} = \frac{t!}{n^t \prod_{i=1}^n x_i!},$$

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 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 $\theta \in \Theta$.

We will prove the result for the discrete case only.

Necessity of the Factorization Theorem

Suppose T is sufficient. By the definition of sufficient statistic,

$$f_{\boldsymbol{X}|T}(\boldsymbol{x} \mid t) = P(\boldsymbol{X} = \boldsymbol{x} \mid T(\boldsymbol{X}) = t)$$

is free of θ. Let

$$\begin{aligned} g(t \mid \theta) &= P_\theta(T(\boldsymbol{X}) = t) \\ h(\boldsymbol{x}) &= P(\boldsymbol{X} = \boldsymbol{x} \mid T(\boldsymbol{X}) = t). \end{aligned}$$

Then, because $\{\boldsymbol{X} = \boldsymbol{x}\} \subset \{T(\boldsymbol{X}) = t\}$,

$$\begin{aligned} f_{\boldsymbol{X}}(\boldsymbol{x} \mid \theta) &= P_\theta(\boldsymbol{X} = \boldsymbol{x}) \\ &= P_\theta(\boldsymbol{X} = \boldsymbol{x}, T(\boldsymbol{X}) = t) \\ &= P_\theta(T(\boldsymbol{X}) = t) P(\boldsymbol{X} = \boldsymbol{x} \mid T(\boldsymbol{X}) = t) \\ &= g(t \mid \theta) h(\boldsymbol{x}). \end{aligned}$$

Sufficiency of the Factorization Theorem

Suppose the factorization holds. To establish that $T = T(\boldsymbol{X})$ is sufficient, it suffices to show that

$$f_{\boldsymbol{X}|T}(\boldsymbol{x} \mid t) = P(\boldsymbol{X} = \boldsymbol{x} \mid T(\boldsymbol{X}) = t)$$

is free of θ. Denoting $T(\boldsymbol{x}) = t$, we have

$$\begin{aligned} f_{\boldsymbol{X}|T}(\boldsymbol{x} \mid t) = P(\boldsymbol{X} = \boldsymbol{x} \mid T(\boldsymbol{X}) = t) &= \frac{P_\theta(\boldsymbol{X} = \boldsymbol{x}, T(\boldsymbol{X}) = t)}{P_\theta(T(\boldsymbol{X}) = t)} \\ &= \frac{P_\theta(\boldsymbol{X} = \boldsymbol{x}) I(T(\boldsymbol{X}) = t)}{P_\theta(T(\boldsymbol{X}) = t)} \\ &= \frac{g(t \mid \theta) h(\boldsymbol{x}) I(T(\boldsymbol{X}) = t)}{P_\theta(T(\boldsymbol{X}) = t)}, \end{aligned}$$

because the factorization holds by assumption.

Sufficiency of the Factorization Theorem (cont.)

Now, write

$$P_\theta(T(\boldsymbol{X}) = t) = P_\theta(\boldsymbol{X} \in A_t),$$

where $A_t = \{\boldsymbol{x} \in \mathcal{X} : T(\boldsymbol{x}) = t\}$. Note that

$$\begin{aligned} P_\theta(\boldsymbol{X} \in A_t) &= \sum_{\boldsymbol{x} \in \mathcal{X} : T(\boldsymbol{x}) = t} P_\theta(\boldsymbol{X} = \boldsymbol{x}) \\ &= \sum_{\boldsymbol{x} \in \mathcal{X} : T(\boldsymbol{x}) = t} g(t \mid \theta) h(\boldsymbol{x}) \\ &= g(t \mid \theta) \sum_{\boldsymbol{x} \in \mathcal{X} : T(\boldsymbol{x}) = t} h(\boldsymbol{x}). \end{aligned}$$

Therefore,

$$f_{\boldsymbol{X}|T}(\boldsymbol{x} \mid t) = \frac{g(t \mid \theta) h(\boldsymbol{x}) I(T(\boldsymbol{X}) = t)}{g(t \mid \theta) \sum_{\boldsymbol{x} \in \mathcal{X} : T(\boldsymbol{x}) = t} h(\boldsymbol{x})} = \frac{h(\boldsymbol{x}) I(T(\boldsymbol{X}) = t)}{\sum_{\boldsymbol{x} \in \mathcal{X} : T(\boldsymbol{x}) = t} h(\boldsymbol{x})},$$

which is free of θ.

Poisson sufficient statistic (cont.)

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

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

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

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

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

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

Sufficient statistics in the Exponential family

Suppose $X_1, \ldots, X_n$ 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 $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_d)$, $d \leq k$. Then

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

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 $j = 1, \ldots, k$.