Math 362: Mathematical Statistics II

Le Chen
le.chen@emory.edu
chenle02@gmail.com

Emory University
Atlanta, GA

Last updated on Spring 2021
Last compiled on January 15, 2023

2021 Spring

Chapter 5. Estimation

Chapter 5. Estimation

§ 5.6 Sufficient Estimators

Rationale: Let $\widehat{\theta}$ be an estimator to the unknown parameter θ. Whether does $\widehat{\theta}$ contain all information about θ?

Equivalently, how can one reduce the random sample of size n, denoted by $(X_1, \cdots, X_n)$, to a function without losing any information about θ?

E.g., let's choose the function $h(X_1, \cdots, X_n) := \frac{1}{n}\sum_{i=1}^{n} X_i$, the sample mean. In many cases, $h(X_1, \cdots, X_n)$ contains all relevant information about the true mean $\mathbb{E}(X)$. In that case, $h(X_1, \cdots, X_n)$, as an estimator, is sufficient.

Definition. Let $(X_1, \cdots, X_n)$ be a random sample of size n from a discrete population with a unknown parameter θ, of which $\widehat{\theta}$ (resp. θ_e) be an estimator (resp. estimate). We call $\widehat{\theta}$ and θ_e **sufficient** if

$$\mathbb{P}\left(X_1 = k_1, \cdots, X_n = k_n \,\middle|\, \widehat{\theta} = \theta_e\right) = b(k_1, \cdots, k_n) \qquad \text{(Sufficency-1)}$$

is a function that does not depend on θ.

In case for random sample $(Y_1, \cdots, Y_n)$ from the continuous population, (Sufficency-1) should be

$$f_{Y_1, \cdots, Y_n | \widehat{\theta} = \theta_e}\left(y_1, \cdots, y_n \,\middle|\, \widehat{\theta} = \theta_e\right) = b(y_1, \cdots, y_n)$$

Note: $\widehat{\theta} = h(X_1, \cdots, X_n)$ and $\theta_e = h(k_1, \cdots, k_n)$.
or $\widehat{\theta} = h(Y_1, \cdots, Y_n)$ and $\theta_e = h(y_1, \cdots, y_n)$.

Equivalently,

Definition. ... $\widehat{\theta}$ (or θ_e) is **sufficient** if the likelihood function can be factorized as:

$$L(\theta) = \begin{cases} \prod_{i=1}^{n} p_X(k_i;\theta) = g(\theta_e,\theta)\; b(k_1,\cdots,k_n) & \text{Discrete} \\ \prod_{i=1}^{n} f_Y(y_i;\theta) = g(\theta_e,\theta)\; b(y_1,\cdots,y_n) & \text{Continous} \end{cases} \qquad \text{(Sufficency-2)}$$

where g is a function of two arguments only and b is a function that does not depend on θ.

E.g. 1. A random sample of size n from Bernoulli(P). $\widehat{p} = \sum_{i=1}^{n} X_i$. Check sufficiency of $\widehat{p}$ for p by (Sufficency-1):

Case I: If $k_1, \cdots, k_n \in \{0,1\}$ such that $\sum_{i=1}^{n} k_i \neq c$, then

$$\mathbb{P}\left(X_1 = k_1, \cdots, X_n = k_n \,\middle|\, \widehat{p} = c\right) = 0.$$

Case II: If $k_1, \cdots, k_n \in \{0,1\}$ such that $\sum_{i=1}^{n} k_i = c$, then

$$
\begin{aligned}
&\mathbb{P}\left(X_1 = k_1, \cdots, X_n = k_n \,\middle|\, \widehat{p} = c\right) \\
&= \frac{\mathbb{P}\left(X_1 = k_1, \cdots, X_n = k_n, \widehat{p} = c\right)}{\mathbb{P}(\widehat{p} = c)} \\
&= \frac{\mathbb{P}\left(X_1 = k_1, \cdots, X_n = k_n, X_n + \sum_{i=1}^{n-1} X_i = c\right)}{\mathbb{P}\left(\sum_{i=1}^{n} X_i = c\right)} \\
&= \frac{\mathbb{P}\left(X_1 = k_1, \cdots, X_{n-1} = k_{n-1}, X_n = c - \sum_{i=1}^{n-1} k_i\right)}{\mathbb{P}\left(\sum_{i=1}^{n} X_i = c\right)} \\
&= \frac{\left(\prod_{i=1}^{n-1} p^{k_i}(1-p)^{1-k_i}\right) \times p^{c-\sum_{i=1}^{n-1} k_i}(1-p)^{1-c+\sum_{i=1}^{n-1} k_i}}{\binom{n}{c} p^c (1-p)^{n-c}} \\
&= \frac{1}{\binom{n}{c}}.
\end{aligned}
$$

In summary,

$$\mathbb{P}\left(X_1 = k_1, \cdots, X_n = k_n \,\middle|\, \widehat{p} = c\right) = \begin{cases} \frac{1}{\binom{n}{c}} & \text{if } k_i \in \{0,1\} \text{ s.t. } \sum_{i=1}^n k_i = c, \\ 0 & \textit{otherwise}. \end{cases}$$

Hence, by (Sufficency-1), $\widehat{p} = \sum_{i=1}^n X_i$ is a sufficient estimator for p.

E.g. 1'. As in E.g. 1, check sufficiency of $\widehat{p}$ for p by (Sufficency-2):

Notice that $p_e = \sum_{i=1}^{n} k_i$. Then

$$
\begin{aligned}
L(p) = \prod_{i=1}^{n} p_X(k_i; p) &= \prod_{i=1}^{n} p^{k_i}(1-p)^{1-k_i} \\
&= p^{\sum_{i=1}^{n} k_i}(1-p)^{n-\sum_{i=1}^{n} k_i} \\
&= p^{p_e}(1-p)^{n-p_e}
\end{aligned}
$$

Therefore, p_e (or $\widehat{p}$) is sufficient since (Sufficency-2) is satisfied with

$$
g(p_e, p) = p^{p_e}(1-p)^{n-p_e} \quad \text{and} \quad b(k_1, \cdots, k_n) = 1.
$$

Comment 1. The estimator $\widehat{p}$ is sufficient but not unbiased since $\mathbb{E}(\widehat{p}) = np \neq p$.

2. Any one-to-one function of a sufficient estimator is again a sufficient estimator. E.g., $\widehat{p}_2 := \frac{1}{n}\widehat{p}$, which is a unbiased, sufficient, and MVE.

3. $\widehat{p}_3 := X_1$ is not sufficient!

E.g. 2. Poisson(λ), $p_X(k;\lambda) = e^{-\lambda}\lambda^k/k!$, $k = 0, 1, \cdots$. Show that $\widehat{\lambda} = (\sum_{i=1}^n X_i)^2$ is sufficient for λ for a sample of size n.

Sol: The Corresponding estimate is $\lambda_e = (\sum_{i=1}^n k_i)^2$.

$$
\begin{aligned}
L(\lambda) &= \prod_{i=1}^n \frac{e^{-\lambda}\lambda^{k_i}}{k_i!} \\
&= e^{-n\lambda}\lambda^{\sum_{i=1}^n k_i}\left(\prod_{i=1}^n k_i!\right)^{-1} \\
&= \underbrace{e^{-n\lambda}\lambda^{\sqrt{\lambda_e}}}_{g(\lambda_e,\lambda)} \times \underbrace{\left(\prod_{i=1}^n k_i!\right)^{-1}}_{b(k_1,\cdots,k_n)}.
\end{aligned}
$$

Hence, $\widehat{\lambda}$ is sufficient estimator for λ. $\square$

E.g. 3. Let $Y_1, \cdots, Y_n$ be a random sample from $f_Y(y;\theta) = \frac{2y}{\theta^2}$ for $y \in [0,\theta]$. Whether is the MLE $\widehat{\theta} = Y_{max}$ sufficient for θ?

Sol: The corresponding estimate is $\theta_e = y_{max}$.

$$
\begin{aligned}
L(\theta) = \prod_{i=1}^{n} \frac{2y}{\theta^2} I_{[0,\theta]}(y_i) &= 2^n \theta^{-2n} \left(\prod_{i=1}^{n} y_i \right) \times \prod_{i=1}^{n} I_{[0,\theta]}(y_i) \\
&= 2^n \theta^{-2n} \left(\prod_{i=1}^{n} y_i \right) \times I_{[0,\theta]}(y_{max}) \\
&= \underbrace{2^n \theta^{-2n} I_{[0,\theta]}(\theta_e)}_{=g(\theta_e,\theta)} \times \underbrace{\prod_{i=1}^{n} y_i}_{=b(y_1,\cdots,y_k)} .
\end{aligned}
$$

Hence, $\widehat{\theta}$ is a sufficient estimator for θ. $\square$

Note: MME $\widehat{\theta} = \frac{3}{2}\overline{Y}$ is NOT sufficient for θ!