Statistical Inference and Multivariate Analysis (MA324)

Lecture Slides
Lecture 10

Point Estimation: Sufficiency, Factorization Theorem

Indian Institute of Technology Guwahati

Jan-May 2023

Information and Sufficiency

- Aim is to estimate unknown parameter $\boldsymbol{\theta}$ based on a realization of a RS using a suitable statistic or estimator.
- The RS $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ **has all the "information"** regarding unknown parameter $\boldsymbol{\theta}$. One should use a **statistic that has same amount** of "information" that the data have regarding $\boldsymbol{\theta}$.
- We can take $\boldsymbol{T}(\boldsymbol{X}) = \boldsymbol{X}$. However, it is **not interesting** in most of the situations as one should take a summary of the data that capture all the "information".
- Therefore, in most of the cases, we will consider a function $\boldsymbol{T} : \chi^n \to \mathbb{R}^m$, **where** $m < n$. In most of the times, the value of m is much smaller than that of n. Such summary or statistic is **as good as the whole data** and is called **sufficient** for $\boldsymbol{\theta}$.

Sufficient Statistic

- If a quantity **vary with the change in another** quantity, then there is **some information** in the first quantity regarding the second. On the other hand, if the first quantity **do not change** with the second quantity, then the first **does not have any information** regarding the second.
- Similarly, if the **distribution of a statistic does not involve the unknown parameter** θ, then the statistic does **not have any information** regarding θ. Motivated by this understanding, a sufficient statistic for θ can be defined as follows.

Def: [**Sufficient Statistic**] A statistic $T = T(X)$ is called a **sufficient statistic for unknown parameter** θ if the **conditional distribution** of X given $T = t$ **does not include** θ for all t in the support of T.

- Thus, given the value t of a sufficient statistic T, **conditionally there is no information left** in X regarding θ.

Sufficient Statistic

- In other words, $\boldsymbol{X}$ is trying to tell us a story regarding $\boldsymbol{\theta}$ and **any statistic is a gist** of the story. If we have the **gist** $\boldsymbol{T}$, a sufficient statistic, the **original story is redundant** as the **gist has all the information** that the original story has regarding $\boldsymbol{\theta}$.
- Note that $\boldsymbol{X}$ is a sufficient statistic. However, we are interested in a **summary statistic** in most of the situations.

Example 1: Let $X_1, X_2, \ldots, X_n \overset{i.i.d.}{\sim} Bernoulli(p)$, $p \in (0, 1)$. Take $T = \sum_{i=1}^{n} X_i$. Show that T is a sufficient statistic for p.

- We can verify if a statistic is sufficient or not using the definition of sufficient statistic. That means that we **first need to guess a correct statistic** and then we can use the definition to **show that it is actually a sufficient** statistic for the unknown parameters.

- However, the next theorem gives **necessary and sufficient conditions**, which can be used to find a sufficient statistic. Therefore, the next theorem is very useful.

Theorem (Neyman-Fisher Factorization Theorem)

Let $X_1, \ldots, X_n$ be RS with JPMF or JPDF $f_{\boldsymbol{X}}(\boldsymbol{x}, \boldsymbol{\theta})$, $\boldsymbol{\theta} \in \Theta$. Then $\boldsymbol{T} = \boldsymbol{T}(X_1, \ldots, X_n)$ is sufficient for $\boldsymbol{\theta}$ if and only if

$$f_{\boldsymbol{X}}(\boldsymbol{x}, \boldsymbol{\theta}) = h(\boldsymbol{x}) g_{\boldsymbol{\theta}}\left(\boldsymbol{T}(\boldsymbol{x})\right),$$

where $h(\boldsymbol{x})$ does not involve $\boldsymbol{\theta}$, $g_{\boldsymbol{\theta}}(\cdot)$ depends on $\boldsymbol{x}$ only through $\boldsymbol{T}(\boldsymbol{x})$.

Example 2: Let $X_1, X_2, \ldots, X_n \overset{i.i.d.}{\sim} Poi(\lambda)$, $\lambda > 0$. Find a sufficient statistic for λ.

Example 3: Let $X_1, X_2, \ldots, X_n \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$, $\mu \in \mathbb{R}$ and $\sigma > 0$. Find a sufficient statistic for (μ, σ^2)