Mathematical Statistics II

Clémonell Bilayi-Biakana
Department of Mathematics and Statistics
University of Ottawa

January 13, 2020.

- Let X be a random variable such that $X \sim f_\theta(x)$, where $\theta \in \Theta$ is an unknown parameter and $f_\theta(x)$ is a known functional form of the pmf/pdf of X.
- Let $\boldsymbol{X} = (X_1, \ldots, X_n)$ be a random sample drawn for satistical purposes. Let $\boldsymbol{x} = (x_1, \ldots, x_n)$ denote its whose observed value. If the sample size n is large then the observed sample $\boldsymbol{x}$ is a long list of numbers that may be hard to interpret.
- To get around this, **data reduction** is needed. In other words, an experimenter might wish to summarize the information in a sample by determining a few key features of the sample values.

- This is usually done by computing **statistics**, that is Borel measurable functions

$$\begin{array}{rccc} T & : & \mathbb{R}^n & \rightarrow & \mathbb{R} \\ & & \boldsymbol{x} & \mapsto & T(\boldsymbol{x}). \end{array}$$

- In what follows, we study principles of data reduction that do not discard important information about the unknown parameter θ and methods that successfully discard information that is irrelevant as far as gaining knowledge about θ is concerned. These include:
  - i) Likelihood Principle
  - ii) Sufficiency Principle
  - iii) Equivariance Principle.

Likelihood Principle

Definition

Let X be a random variable such that $X \sim f(x\|\theta)$, $\theta \in \Theta$. Let $f(\mathbf{x}\|\theta)$ denote the joint pdf or pmf of the sample $\mathbf{X}$. Then given $\mathbf{X} = \mathbf{x}$ is observed,

$$\theta \mapsto L(\theta\|\mathbf{x}) = f(\mathbf{x}\|\theta) = \prod_{i=1}^{n} f(x_i\|\theta)$$

is called the **likelihood function**.

Remark

If $\mathbf{X}$ is discrete, then $L(\theta\|\mathbf{x}) = P_\theta(\mathbf{X} = \mathbf{x})$.

Remarks

- *If $\theta_1, \theta_2 \in \Theta$ such that*

$$P_{\theta_1}(\boldsymbol{X} = \boldsymbol{x}) = L(\theta_1 \| \boldsymbol{x}) > L(\theta_2 \| \boldsymbol{x}) = P_{\theta_2}(\boldsymbol{X} = \boldsymbol{x}),$$

  then the sample we actually observed is more likely to occur at $\theta = \theta_1$ than at $\theta = \theta_2$.
- *Therefore, that comparison of the likelihood function at two different parameters θ_1 and θ_2 gives an approximate comparison of the probability of the observed sample value, $\boldsymbol{x}$.*

Likelihood Principle

Principle

If $\mathbf{x}$ and $\mathbf{y}$ are two sample points such that:

$$L(\theta \| \mathbf{x}) \propto L(\theta \| \mathbf{y})$$

$$\Leftrightarrow \exists\ C(\mathbf{x}, \mathbf{y}) : \forall\ \theta, L(\theta \| \mathbf{x}) = C(\mathbf{x}, \mathbf{y}) L(\theta \| \mathbf{y})$$

then the conclusions drawn from $\mathbf{x}$ and $\mathbf{y}$ should be identical.

Remarks

- *The likelihood principle says that even if two sample points contain have only proportional likelihood, then they contain equivalent information about θ.*
- *The likelihood principle specifies how the likelihood should be used as data reduction device.*

Sufficiency Principle

Definition

A statistic $T : \mathbb{R}^n \to \mathbb{R}$ is sufficient for θ if for all $\boldsymbol{x}, \theta$,

$$P_\theta(\boldsymbol{X} = \boldsymbol{x} \| T(\boldsymbol{X}) = T(\boldsymbol{x})) \text{ is free from } \theta$$

Theorem

If $p(\boldsymbol{x}\|\theta)$ is the joint pdf or pmf of $\boldsymbol{X}$ and $q(t\|\theta)$ is the pdf or pmf of $T(\boldsymbol{X})$. Then $T(\boldsymbol{X})$ is sufficient for θ if

$$\forall \boldsymbol{x}, \quad \frac{p(\boldsymbol{x}\|\theta)}{q(T(\boldsymbol{x}\|\theta)} \text{ is free from } \theta.$$

Sufficiency Principle

Exercise

Let $X_1, \ldots, X_n \overset{i.i.d}{\sim} \mathcal{N}(\mu, \sigma^2)$, with σ^2-known. Show that $\bar{X}$ is a sufficient statistics for μ.

Theorem (Factorization Theorem)

Let $f(\boldsymbol{x} \| \theta)$ be the joint pdf or pmf of $\boldsymbol{X} = (X_1, \ldots, X_n)$. A statistic $T(\boldsymbol{X})$ is sufficient for $\theta \Leftrightarrow \exists\, g(t \| \theta)$ and $h(\boldsymbol{x})$ s.t.

$$f(\boldsymbol{x} \| \theta) = g(T(\boldsymbol{x}) \| \theta) h(\boldsymbol{x}).$$

Exercise

Let $X_1, \ldots, X_n \overset{i.i.d}{\sim} \mathcal{N}(\mu, \sigma^2)$, with σ^2-known. Find a sufficient statistics for μ.

Sufficiency Principle

Exercise

Let $X_1, \dots, X_n \overset{i.i.d}{\sim} \mathcal{N}(\mu, \sigma^2)$*, with* μ, σ^2*-unknown. Find a sufficient statistics for* μ*.*

Theorem

Let $X_1, \dots, X_n \overset{i.i.d}{\sim} f(x \| \boldsymbol{\theta})$*, where* $\boldsymbol{\theta} = (\theta_1, \dots, \theta_d), (d \leqslant k)$*. Assume that* $f(\cdot \| \boldsymbol{\theta})$ *belongs to the the exponential family, i.e*

$$f(x \| \boldsymbol{\theta}) = a(\boldsymbol{\theta}) b(x) e^{\sum_{j=1}^{k} c_j(\boldsymbol{\theta}) T_j(x)},$$

Then $T(\boldsymbol{X}) = \left(\sum_{j=1}^{n} T_1(X_j), \dots, \sum_{j=1}^{n} T_k(X_j) \right)$ *is a sufficient statistic for* $\boldsymbol{\theta}$*.*

Sufficiency Principle

Principle (Sufficiency)

If $T : \mathbb{R}^n \to \mathbb{R}$ is a sufficient for θ, then any inference about θ should depend on the sample $\boldsymbol{X}$ only through $T(\boldsymbol{x})$.

Remarks

- *The rationale behind the sufficiency principle is that if $\boldsymbol{x}$ and $\boldsymbol{y}$ are two sample points such that $T(\boldsymbol{x}) = T(\boldsymbol{y})$; then the inference about θ should be the same whether $\boldsymbol{X} = \boldsymbol{x}$ or $\boldsymbol{X} = \boldsymbol{y}$ is observed.*
- *A* **sufficient statistic** *for a parameter θ is a statistic that in a certain sense,* **captures all the information about** *θ contained in the sample.*

Sufficiency Principle

Definition (Minimal Sufficiency)

A sufficient statistic $T(\boldsymbol{X})$ for θ is called **minimal sufficient** *if for any other sufficient statistic $Q(\boldsymbol{X})$,*

$$T(\boldsymbol{x}) = h[Q(\boldsymbol{x})].$$

Theorem (Lehmann-Scheffe)

Let $f(\boldsymbol{x}\|\theta)$ be the joint pdf or pmf of a sample $\boldsymbol{X}$. Suppose there exists a measurable function $T : \mathbb{R}^n \to \mathbb{R}$ such that

$$\forall\, \boldsymbol{x}, \boldsymbol{y},\ \frac{f(\boldsymbol{x}\|\theta)}{f(\boldsymbol{y}\|\theta)} \text{ is free from } \theta \Leftrightarrow T(\boldsymbol{x}) = T(\boldsymbol{y}).$$

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

Sufficiency Principle

Remarks

- *A minimal sufficient statistic is a statistic that has achieved the maximal amount of data reduction possible while retaining all the information about θ.*
- *Intuitively, a minimal statistic eliminates all the extra neous information in the sample, retaining only that piece with information about θ.*

Example

If $X_1, \ldots, X_n \overset{i.i.d}{\sim} \mathcal{N}(\mu, \sigma^2)$, with μ, σ^2-unknown, then $(\bar{X}, S^2)$ is a minimal sufficient statistic for (μ, σ^2).

Equivariance Principle

Equivariance principle includes two considerations:

i) measurement equivariance

ii) formal invariance.

Measurement equivariance says that the inference made should not depend on the measurement scale that is used.

Formal invariance states that if two inference problems have the same formal structure in terms of the mathematical model used, then the same inference procedure should be used in both problems. The elements of the model that must be the same are:

- Θ, the parameter space,
- $\{f(\boldsymbol{x}\|\theta) : \theta \in \Theta\}$, the set of pdfs or pmfs for the samples.

Example

Two foresters are going to estimate the average diameter of trees in a forest.

- *The first uses data on tree diameters expressed in inches,*
- *and the second uses the same data expressed in meters.*

If both are asked to produce an estimate in inches, what should they do?.

Remark

Measurement equivariance requires that both foresters produce the same estimates.

Example

The sample parameter may be $\Theta = \{\theta : \theta > 0\}$ *in two problems. But in one problem* θ *may be the average price of a dozen eggs in the United States (measured in cents) and in another problem* θ *may refer to the average height of giraffes in Kenya (measured in meters).*

Remarks

- *Formal invariance is concerned only with the mathematical entities involved, not the physical description of the experiment.*
- *Formal invariance equates these two parameter spaces since they both refer to the same set of real numbers.*

Principle (Equivariance)

If $\boldsymbol{Y} = g(\boldsymbol{X})$ *is change of measurement scale such that the model for* $\boldsymbol{Y}$ *has the same formal structure as the model function* $\boldsymbol{X}$*, then an inference procedure should be both* **measurement equivariant** *and* **formally invariant**.

Exercise

Let $X \sim \mathcal{B}(n, p)$*, where* n *is known and* p *unknown. Show that an inference procedure should be both measurement equivariant and formally invariant.*