Likelihood Principle

Definition 6.5

Let $\boldsymbol{X}$ *be a random sample with joint pdf (or pmf)* $f(\boldsymbol{x} \mid \theta)$. *If* $\boldsymbol{x}$ *is the observed value of* $\boldsymbol{X}$ *then the likelihood function is the function of* θ *given by*

$$L(\theta \mid \boldsymbol{x}) = f(\boldsymbol{x} \mid \theta)$$

- We can use the likelihood to compare the probability of observing $\boldsymbol{X} = \boldsymbol{x}$ under different parameter values.

- If $L(\theta_1 \mid \boldsymbol{x}) > L(\theta_2 \mid \boldsymbol{x})$ then we may think of θ_1 as being more plausible than θ_2 given that we did actually observe $\boldsymbol{X} = \boldsymbol{x}$.

The Likelihood Principle

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

$$L(\theta \mid \boldsymbol{x}) = C(\boldsymbol{x}, \boldsymbol{y}) L(\theta \mid \boldsymbol{y})$$

then the inference drawn from $\boldsymbol{x}$ should be identical to those drawn from $\boldsymbol{y}$.

- Note that if $T(\boldsymbol{X})$ is a sufficient statistic and $\boldsymbol{x}$ and $\boldsymbol{y}$ are such that $T(\boldsymbol{x}) = T(\boldsymbol{y})$ then the sufficiency principle says that inference should be the same and so does the likelihood principle.

- One type of inference which satisfies the likelihood principle is fiducial inference.

- In such inference, $L(\theta \mid \boldsymbol{x})$ is considered to be proportional to a pdf (or pmf) for θ.

- There is no guarantee that $L(\theta \mid \boldsymbol{x})$ can be normalized.

- This type of inference has a long history but is very controversial.

- We shall see later that fiducial inference is equivalent to Bayesian inference with a constant prior on θ.

- A more formal likelihood principle can be found by considering a random experiment E to be made up of the triple $\{\boldsymbol{X}, \theta, f(\boldsymbol{x} \mid \theta)\}$.

- The inference or conclusions drawn from such an experiment when we observe $\boldsymbol{X} = \boldsymbol{x}$ is denoted as $\mathsf{Ev}(\mathsf{E}, \boldsymbol{x})$.

Formal Sufficiency Principle

Consider an experiment $E = \{\boldsymbol{X}, \theta, f(\boldsymbol{x} \mid \theta)\}$ *and suppose that, in this experiment,* $T(\boldsymbol{X})$ *is a sufficient statistic for* θ. *If* $\boldsymbol{x}$ *and* $\boldsymbol{y}$ *are sample points such that* $T(\boldsymbol{x}) = T(\boldsymbol{y})$ *then* $\mathsf{Ev}(\mathsf{E}, \boldsymbol{x}) = \mathsf{Ev}(\mathsf{E}, \boldsymbol{y})$.

Conditionality Principle

Suppose that $E_1 = \{\boldsymbol{X}_1, \theta, f_1(\boldsymbol{x}_1 \mid \theta)\}$ *and* $E_2 = \{\boldsymbol{X}_2, \theta, f_2(\boldsymbol{x}_2 \mid \theta)\}$ *are two experiments with common parameter* θ*. Consider another experiment in which we first randomly select one of* E_1 *or* E_2 *and then perform that experiment. Call this experiment* $E^* = \{\boldsymbol{X}^*, \theta, f^*(\boldsymbol{x}^* \mid \theta)\}$ *where* $\boldsymbol{X}^* = (j, \boldsymbol{X}_j)$*. Then*

$$\mathrm{Ev}(\mathrm{E}^*, (\mathrm{j}, \boldsymbol{x}_\mathrm{j})) = \mathrm{Ev}(\mathrm{E}_\mathrm{j}, \boldsymbol{x}_\mathrm{j})$$

- Inference should depend only on the actual experiment performed and not on any other possible experiments.

Formal Likelihood Principle

Suppose that $E_1 = \{\boldsymbol{X}_1, \theta, f_1(\boldsymbol{x}_1 \mid \theta)\}$ *and* $E_2 = \{\boldsymbol{X}_2, \theta, f_2(\boldsymbol{x}_2 \mid \theta)\}$ *are two experiments with common parameter* θ *and that* $\boldsymbol{x}_1$ *and* $\boldsymbol{x}_2$ *are sample points from* E_1 *and* E_2 *respectively. Let* L_1 *and* L_2 *be the likelihood functions corresponding to* f_1 *and* f_2 *respectively. If*

$$L_1(\theta \mid \boldsymbol{x}_1) = C(\boldsymbol{x}_1, \boldsymbol{x}_2) L_2(\theta \mid \boldsymbol{x}_2)$$

then $\mathrm{Ev}(\mathrm{E}_1, \boldsymbol{x}_1) = \mathrm{Ev}(\mathrm{E}_2, \boldsymbol{x}_2)$*.*

Likelihood Principle Corollary

If $E = \{\boldsymbol{X}, \theta, f(\boldsymbol{x} \mid \theta)\}$ *is an experiment and* $\boldsymbol{x}$ *is the observed sample then* $\mathrm{Ev}(\mathrm{E}, \boldsymbol{x})$ *should depend on* E *and* $\boldsymbol{x}$ *only through the likelihood function* $L(\theta \mid \boldsymbol{x})$*.*

The Equivariance Principle

Definition 6.6

A procedure is measurement equivariant if a change in the measurement scale of the observations causes the same change in inference.

Definition 6.7

If two inference problems have the same mathematical structure in terms of the parameter space, set of pdfs (or pmfs) and set of allowable inferences and consequences of incorrect inference then the formal invariance says that the same inference procedure should be used in the two problems.

Equivalence Principle

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

- This means that when we change measurement scale, our inference procedure should make the same change in our estimate of θ.
- Furthermore, the inference should only depend on the mathematical structure and so inference based on a point $\boldsymbol{x}$ should be identical irrespective of the scale on which it was measured.

- Application of the equivariance principle relies on a set of transformations of the sample space being a group of transformations on the sample space $\mathcal{X}$.

Definition 6.8

A set of functions $\{g(\boldsymbol{x}) : g \in \mathcal{G}\}$ mapping $\mathcal{X}$ to itself is called a group of transformations of $\mathcal{X}$ if

1. (inverse) For every $g \in \mathcal{G}$ there exists a $g_1 \in \mathcal{G}$ such that

$$g_1(g(\boldsymbol{x})) = \boldsymbol{x} \quad \textit{for every } \boldsymbol{x} \in \mathcal{X}.$$

2. (composition) For every $g_1 \in \mathcal{G}$ and $g_2 \in \mathcal{G}$ there exists a function $g_2 \in \mathcal{G}$ such that

$$g_1(g(\boldsymbol{x})) = g_2(\boldsymbol{x}) \quad \textit{for every } \boldsymbol{x} \in \mathcal{X}.$$

- Note that this definition implies that the identity function $e(\boldsymbol{x}) = \boldsymbol{x}$ is also an element of $\mathcal{G}$.

Definition 6.9

Let $\mathcal{F} = \{f(\boldsymbol{x} \mid \theta) : \theta \in \boldsymbol{\Theta}\}$ *be a set of pdfs (or pmfs) for* $\boldsymbol{X}$ *and let* $\mathcal{G}$ *be a group of transformations of the sample space* $\mathcal{X}$*. Then* $\mathcal{F}$ *is* *invariant under the group* $\mathcal{G}$ *if for every* $\theta \in \boldsymbol{\Theta}$ *and* $g \in \mathcal{G}$ *there exists a unique* $\theta_1 \in \boldsymbol{\Theta}$

$$X \sim f(\boldsymbol{x} \mid \theta) \;\Rightarrow\; Y \sim f(\boldsymbol{y} \mid \theta_1).$$