Theory of Statistical Inference - Lecture III.2
STA422 and STA2162

Michael Evans
University of Toronto
https://utstat.utoronto.ca/mikevans/sta422/sta4222026.html

2026

III.2 Pure Likelihood Inference

- Edwards (1992) Likelihood

- see Richard Royall (1997) Statistical Evidence: A likelihood paradigm

- a good discussion of the basic problems and what is wrong with many approaches to solve these

- Royall emphasizes that characterizing statistical evidence is the goal and makes a proposal

- inference base $(\{f_\theta : \theta \in \Theta\}, x)$ gives rise to likelihood function $L(\cdot \,|\, x) : \Theta \to [0, \infty)$ given by

$$L(\theta \,|\, x) = c f_\theta(x)$$

for any constant $c > 0$

- actually uses the relative likelihood

$$L^{rel}(\theta \,|\, x) = \frac{f_\theta(x)}{\sup\{f_\theta(x) : \theta \in \Theta\}}$$

- typically $L^{rel}(\theta \,|\, x) = f_\theta(x)/f_{\theta(x)}(x)$ where $\theta(x)$ is the MLE

- evidential interpretation whenever $\theta_1 \succcurlyeq \theta_2$, then this indicates that the likelihood contains at least as much evidence that θ_1 is the true value as it does for θ_2

E - estimate the true value by $\theta(x)$ and, for some $\gamma \in [0, 1]$, quote a likelihood region $C_\gamma(x) = \{\theta : L^{rel}(\theta \,|\, x) \geq \gamma\}$ and the "size" of $C_\gamma(x)$ provides an assessment of the accuracy of $\theta(x)$

- how to choose γ?

> **Law of the Likelihood** *(Hacking(1965)) If the ratio* $L(\theta_1 \,|\, x)/L(\theta_2 \,|\, x) > 1$, *then the data* x *is evidence supporting* θ_1 *over* θ_2 *and the ratio measures the strength of that evidence.*

- Royall argues for $\gamma = 1/8$ as representing *fairly strong evidence* (I presume in favor of any value $\theta \in C_{1/8}(x)$)

- why?

- suppose $\theta \in C_{1/8}(x)$ and $\theta' \in \Theta$ then $L^{rel}(\theta \,|\, x) \geq 1/8$ and

$$\frac{L(\theta' \,|\, x)}{L(\theta \,|\, x)} \leq \frac{L(\theta(x) \,|\, x)}{L(\theta \,|\, x)} = \frac{1}{L^{rel}(\theta \,|\, x)} \leq 8$$

and so for a value $\theta \in C_{1/8}(x)$ no other value $\theta' \in \Theta$ will have evidential support for it over θ by a factor greater than 8

- suppose $\theta \notin C_{1/8}(x)$ then $L^{rel}(\theta \,|\, x) < 1/8$ which implies there is at least one value in $C_{1/8}(x)$ having at least 8 times the evidential support, namely, $\theta(x)$

- why 8?

- consider two urns: urn I contains all white balls and urn II contains $1/2$ white and $1/2$ black balls

- an urn is selected (we don't know which) and balls are drawn with replacement

- the first ball is W so $L(I \,|\, x)/L(II \,|\, x) = 2$, when the second ball drawn is W then $L(I \,|\, x)/L(II \,|\, x) = 4$ and when the third ball drawn is W, then $L(I \,|\, x)/L(II \,|\, x) = 8$ and Royall claims most would conclude that this is fairly strong evidence in favor of urn I over urn II

- Royall also argues that $\gamma = 1/32$ represents *quite strong evidence*

- **H** - assess whether there is evidence in favor of or against $H_0 : \theta = \theta_0$ by the value $L^{rel}(\theta_0 \,|\, x)$ and compare this to $1/8$ $(1/32)$

- criticized because the values 8 or 32 seem arbitrary and there does not seem to be any reason to regard the urn model experiment as paradigmatic for the characterization of evidence

- there is another significant problem, namely, what about inference for $\psi = \Psi(\theta)$ when there are nuisance parameters?

- a general solution to this must produce a likelihood ordering for ψ, i.e., there is some function of the data, say $T(x)$, that induces model $\{f_{T,\psi} : \psi \in \Psi(\Theta)\}$ so that we have the likelihood

$$L(\psi \,|\, T(x)) = cf_{T,\psi}(T(x))$$

- the general approach here is to use the *profile likelihood*

$$L^{\Psi}(\psi \,|\, x) = \sup_{\theta \in \Psi^{-1}\{\psi\}} L(\theta \,|\, x)$$

- but in general the profile likelihood is not a likelihood (based on a model) and it produces odd answers at times

Example III.2.1 *A profile likelihood that is not a likelihood*

- model is given by

θ	$f_\theta(1)$	$f_\theta(2)$
0	1/2	1/2
1	1/3	2/3
2	1/5	4/5

- suppose interest is in $\Psi = I_{H_0}$ where $H_0 = \{0, 1\}$ so profile likelihood equals

$$L^{\Psi}(0 \,|\, 1) = 1/5, \quad L^{\Psi}(1 \,|\, 1) = 1/2, \quad L^{\Psi}(0 \,|\, 2) = 4/5, \quad L^{\Psi}(1 \,|\, 2) = 2/3$$

- to be a likelihood we need a function $T : \{1, 2\} \to \{1, 2\}$ such that the likelihoods based on the model induced by T are proportional to $L^{\Psi}(\cdot \,|\, 1)$ and $L^{\Psi}(\cdot \,|\, 2)$

- $L^{\Psi}(\cdot \,|\, 1)$ and $L^{\Psi}(\cdot \,|\, 2)$ are not proportional, so T must be 1-1 to produce two distinct likelihood functions

- there is no such T because the fact that it is 1-1 implies that its model is effectively given by the table ■

Exercise III.2.1 Suppose we are given the inference base inference base $(\{f_\theta : \theta \in \Theta\}, x)$ and we want to predict an independent future value $y \sim g_\theta$ with the same true value of θ as for the f_θ that produced x. A natural *predictive likelihood* for y is obtained from the joint density $f_\theta(x)g_\theta(y)$ via treating θ as a nuisance parameter to obtain the profile likelihood

$$L^Y(y \,|\, x) = \sup_{\theta \in \Theta} f_\theta(x)g_\theta(y).$$

from the joint likelihood for (θ, y). We can also form the profile likelihood for θ from this, namely,

$$L^\Theta(\theta \,|\, x) = \sup_{\theta \in \Theta} f_\theta(x)g_\theta(y)$$

and in general $L^\Theta(\theta \,|\, x) \neq cL(\theta \,|\, x) = cf_\theta(x)$ for any constant $c > 0$. As a specific example, suppose x is a sample of n from the model $\{N(0, \sigma^2) : \sigma^2 > 0\}$ and $y \sim N(0, \sigma^2)$ independent of x and unobserved. Determine the predictive likelihood for y, the profile likelihood for σ^2 and the standard likelihood for σ^2. Determine the MLE of σ^2 and the profile MLE of σ^2.

Personal Opinion

I believe that the likelihood ordering of the θ values must be incorporated as part of any theory of inference about θ in the sense that any valid inference must reflect this. In particular, any proper characterization of statistical evidence has to reflect this. While it is clear that there must be evidence in favor of the MLE $\theta(x)$, the pure likelihood approach does not clearly specify a cut-off γ such that θ values in the likelihood region $C_{\gamma}(x)$ have evidence in their favor and those outside don't. The problem with nuisance parameters further convinces me that the likelihood alone does not characterize statistical evidence properly.