Shaky foundations

What they don't tell you as an under-grad

Russell Millar
University of Auckland

27 Aug 2014

The likelihood principle

The likelihood principle (LP):
If $\boldsymbol{y}_1$ observed from experiment E_1, and $\boldsymbol{y}_2$ observed from experiment E_2, have the same likelihood functions (to within a constant), i.e., $f_1(\boldsymbol{y}_1;\boldsymbol{\theta}) = cf_2(\boldsymbol{y}_2;\boldsymbol{\theta})$, for all $\boldsymbol{\theta} \in \Theta$, then the information content about $\boldsymbol{\theta}$ is the same from both $(E_1, \boldsymbol{y}_1)$ and $(E_2, \boldsymbol{y}_2)$. $\square$

The LP is not liked by frequentists. For example, frequentist inference for negative binomial and binomial experiments is different, although they have the same form of likelihood.

E.g., If $y_1 = 4$ is observed from a Bin$(12, p)$ experiment, and $y_2 = 12$ is observed from a NegBin$(4, p)$ experiment (where Y_2 is the number of trial required to get 4 successes), these two experiments have the same likelihood function. However, they lead to different frequentist inferences.

The frequentist dilemma

Frequentists do “like” the conditionality principle. In fact, the notion of repeating experiments is reliant upon conditioning on ancillary variables.

The conditionality principle (CP):
Let the mixture experiment, E, consist of $(E_1, \boldsymbol{y}_1)$ (i.e., performing experiment E_1, from which $\boldsymbol{y}_1$ is observed) with probability p, or $(E_2, \boldsymbol{y}_2)$ with probability $1 - p$. If $(E_i, \boldsymbol{y}_i), i = 1, 2$, is the experiment actually performed, then the information content about $\boldsymbol{\theta}$ from the mixture experiment $(E, (E_i, \boldsymbol{y}_i))$ is equal to that from $(E_i, \boldsymbol{y}_i)$. $\square$

Evans et al (1986) argued that CP $\rightarrow$ LP!

They write “... the proof [of CP $\rightarrow$ LP] proceeds precisely because of a well-known problem with CP, the lack of a unique maximal ancillary.”

The Rainbow Stone (*Lichihodae paradoxi*)

Adapted from the drunken sailor example from Stone (1976), which had origins as an attempt to refute the LP.

Rainbow stones have a hard multilayered covering, each layer being of a colour remarkably similar to one of the seven colours of the rainbow. Annual calcific growth rings are clearly evident in the cross-section of a Rainbow stone, enabling clear resolution of adjacent layers of identical colour. The entire sequence of colours, beginning from the centre, is often referred to as the "state" of a stone.

At annual intervals Rainbow stone attempt to grow another layer. The new layer will be any of the seven colours with equal probability $\frac{1}{8}$. With probability $\frac{1}{8}$ the stone will be unable to produce a new layer and remains unchanged. These probabilities are believed to be independent of the stone's previous state.

A Cacti stone, closely related to the Rainbow stone, but having only four possible colours for each layer.

The Rainbow Stone (*Lichihodae paradoxi*)

PROBLEM: To make inference about the previous state of a randomly chosen Rainbow stone, given observation of its current state.

Denoting the observed current state $C_1 C_2 ... C_n,\ n \geq 2$, where each $C_i \in \{Red, Orange, Yellow, Green, Blue, Indigo, Violet\}$, there are just two possible previous states, the observed state minus the top layer,

$$C_1 C_2 ... C_{n-1}$$

and the observed state itself,

$$C_1 C_2 ... C_n.$$

These two previous states each have likelihood $\frac{1}{8}$.

The Rainbow Stone (*Lichihodae paradoxi*)

From a frequentist perspective, Rainbow stone add a layer with probability $\frac{7}{8}$ and the observed state minus the top layer will therefore give the true previous state 7 times out of 8. However, this preference is not borne out by the equal likelihoods given to the two possible states.

Therein lies a potential "counter example" to the likelihood principle (???). The likelihood principle (LP) gives equal support to the two previous states $C_1 C_2 ... C_{n-1}$ and $C_1 C_2 ... C_n$.

The Rainbow Stone (*Lichihodae paradoxi*)

Is this a counter example to LP???

NO, at least not for a Bayesian. In fact this drives home the message to use the LP within the Bayesian paradigm wherein prior information can be formally utilized (Berger and Wolpert 1988).

Standard arguments based on population dynamics (e.g., using a Markov transition matrix) lead to a prior which places 7 times as much prior mass on the previous state $C_1 C_2 \ldots C_{n-1}$ than $C_1 C_2 \ldots C_n$.

Frequentists don't appear to have a coherent answer to the LP.

My personal journey

- Having encountered the LP and CP in the mid 90's, I was receptive to Bayesian concepts upon arriving at U. Akld
- Robert Gentleman and Geoff Nicholls put on informal workshops on MCMC in the late 90's.
- MCMC was a viable approach to fitting nonlinear state-space models of fish population dynamics - fisheries biologists could run these models in BUGS.
- The Bayesian approach is increasingly popular in population modeling due to its ability to formally include uncertainty in prediction.
- Now having great "fun" with Bayesian sensitivity theory... leading to leave-one-out approximations...leading to information criteria such as WAIC (widely appicable information criterion) and MWAIC (more WAIC).

References

Berger and Wolpert 1988. The likelihood principle. 252p

Stone 1976. Strong inconsistency from uniform priors, JASA.

Evans et al. 1986. On principles and arguments to likelihood, Can. J S