Revised Lindley-Jeffreys Paradox: Section 4.1.3

Ronald Christensen

Department of Mathematics and Statistics

University of New Mexico

July 16, 2010

Abstract

KEY WORDS:

4.1.3 Lindley-Jeffreys Paradox

The moral of the Lindley-Jeffreys paradox is that if you pick a stupid prior, you can get a stupid posterior. Our discussion is not in the form of the traditional Lindley-Jeffreys paradox but it makes the same point.

Suppose $y|\theta \sim N(\theta, 1)$ and we want to test the hypothesis $H_0 : \theta = 0$ versus the alternative $H_1 : \theta \neq 0$. We need to specify a prior probability distribution on θ values. As previously discussed, if we use a continuous distribution, the prior probability of $\theta = 0$ is 0, so there will be no chance of ever accepting the null hypothesis. We thus specify $q_0 = \Pr(\theta = 0) = 0.5$, and distribute the rest of the probability on $\theta \neq 0$. To distribute the rest of the probability we specify that $1 - q_0 = \Pr(\theta \neq 0) = 0.5$ and further specify a continuous distribution for θ given $\theta \neq 0$.

To make our point, we need a distribution that is highly dispersed. Such distributions are often thought to convey ignorance about the particular value of $\theta \neq 0$, but we have reservations about using such models to convey ignorance. For no other reason than to simplify computations, we use a normal conditional distribution, in particular, $\theta|\,\theta \neq 0$ is $N(0, \sigma_0^2)$. To have high dispersion, we need σ_0^2 large.

Example 2.3.3 gave a Bayesian analysis for normal data with known variance and a normal prior. Computing the marginal distribution of y in that case gives

$$y|\,\theta \neq 0 \sim N(0, 1 + \sigma_0^2).$$

Given data y, to decide between the hypotheses we calculate the posterior probability of H_0:

$$\Pr(\theta = 0|y) = \frac{q_0 f(y|\theta = 0)}{q_0 f(y|\theta = 0) + (1 - q_0) f(y|\theta \neq 0)}.$$

With $q_0 = 0.5$, it is not difficult to see that $\Pr(\theta = 0|y) > 0.5$ if and only if $f(y|\theta = 0) > f(y|\theta \neq 0)$. These are both normal densities:

$$f(y|\theta = 0) = \frac{1}{\sqrt{2\pi}} \exp\left[-y^2/2\right]$$

and

$$f(y|\theta \neq 0) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{1+\sigma_0^2}} \exp\left[-y^2/2(\sigma_0^2+1)\right].$$

Note that $f(y|\theta = 0) > f(y|\theta \neq 0)$ if and only if

$$y^2 < \log(1+\sigma_0^2)(\sigma_0^2+1)/\sigma_0^2.$$

Clearly, as σ_0^2 increases, arbitrarily large values of y cause us to accept H_0. In particular, with σ_0 as small as 100, an observed value of $y = 3$ causes us to choose $\theta = 0$ rather than $\theta \neq 0$.

The paradoxical thing about this is that seeing $y = 3$ from a $N(0, 1)$ distribution is an exceptionally unusual event. In significance testing, one would easily reject the model with $\theta = 0$ when seeing $y = 3$. The problem is that the prior distribution is putting so much probability on very large values of θ that, even though it is very unlikely that one would see $y = 3$ from a $N(0, 1)$, it is more likely to come from the $N(0, 1)$ distribution than it is from the distribution of y given the alternative, namely, the $N(0, 10001)$ distribution.

We believe that there is no such thing as a prior distribution that embodies ignorance; however, there are some distributions that are more easily overwhelmed by the data than others, and these are often prior distributions with large variability. In any case, the moral of the Lindley-Jeffreys Paradox is that you need to actually think about an appropriate distribution for θ given the alternative. For this problem, trying to choose a convenient distribution containing little information wreaks havoc.