Bayesian Analysis (2009) **4**, Number 4, pp. 759–762

Inconsistent Bayesian Estimation

Ronald Christensen*

Abstract. A simple example is presented using standard continuous distributions with a real valued parameter in which the posterior mean is inconsistent on a dense subset of the real line.

Keywords: Dirichlet process, Posterior mean

1 Introduction

There has been extensive work on inconsistent Bayesian estimation. Early work was done by Halpern (1974) , Stone (1976) , and Meeden and Ghosh (1981). An important paper was Diaconis and Freedman (1986a), henceforth referred to as DFa, with extensive references and discussion by Barron; Berger; Clayton; Dawid; Doksum and Lo; Doss; Hartigan; Hjort; Krasker and Pratt; LeCam; and Lindley. Follow up work includes Diaconis and Freedman (1986b, 1990, 1993), Datta (1991), Berliner and MacEachern (1993), and Rukhin (1994).

DFa require consistency for every parameter value. They also point out that if their definition of consistency holds, then the posterior mean is consistent ("minor technical details apart"). The purpose of this note is to provide a particularly simple example of an inconsistent Bayes estimate and to draw some conclusions from that example. In particular, the example has a posterior mean that is inconsistent on a dense subset of the real line.

Consider $y_1, \ldots, y_n$ a random sample from a density $f(y|\theta)$. The distribution of $f(y|\theta)$ is Cauchy with median θ when θ is a rational number and Normal with mean θ and variance 1 when θ is irrational. In other words,

$$f(y|\theta) = \begin{cases} \text{Cauchy}(\theta) & \theta \text{ rational} \\ N(\theta, 1) & \theta \text{ irrational.} \end{cases}$$

For the prior density, we take $g(\theta)$ to be absolutely continuous. For the sake of simplicity, take it to be $N(\mu_0, 1)$.

We will show that the posterior distribution of θ given the data is the same as if the entire conditional distribution of y were $N(\theta, 1)$. In other words, the posterior distribution is

$$f(\theta|y_1, \ldots, y_n) \sim N\left(\frac{\mu_0 + n\overline{y}}{n+1}, \frac{1}{n+1}\right).$$

The standard Bayes estimate is the posterior mean, $(\mu_0 + n\overline{y})/(n+1)$, which behaves asymptotically like $\overline{y}$. If the true value of θ is an irrational number, the true sampling

*Department of Mathematics and Statistics, University of New Mexico mailto:fletcher@stat.unm.edu

 DOI:10.1214/09-BA428

distribution is normal and the Bayes estimate is consistent. However, if the true value of θ is a rational number, the true sampling distribution is Cauchy(θ), for which it is well known that $\bar{y}$ is an inconsistent estimate of θ. Thus we have the Bayes estimate inconsistent on a dense set, but a set of prior probability zero.

As the editor has pointed out, except in neighborhoods of $\theta = 0$, the example works just as well with the Cauchy(θ) replaced by a $N(-\theta, 1)$. Then, the posterior mean is consistent but for the wrong value of θ.

Obviously, the key to this example is that, by virtually any concept of proximity for distributions, the conditional distributions $f(y|\theta)$ are discontinuous on a dense set of θs. Not only is the mean function $E(y|\theta)$ discontinuous everywhere in θ but if $F(y|\theta)$ is the cdf of $f(y|\theta)$, measures such as the Kolmogorov-Smirnov distance $\sup_y |F(y|\theta) - F(y|\theta')|$ are never uniformly small in any neighborhood of θs. An interesting aspect of DFa is that, while generally it is possible to get discrete distributions arbitrarily close to continuous ones, DFa illustrate that you cannot always get Ferguson distributions close enough to a continuous target.

It seems quite clear from the calculus behind this example that the proper concern for Bayesians is whether their procedures are consistent with prior probability one. Doob's theorem, see DFa's Corollary A.2, establishes precisely this result. Moreover, there seems to be little remedy for Bayesian inconsistency if one has postulated a prior distribution for which all interesting parameters have collective prior probability zero. We have done that here. Who ever reports numerical values to clients that are not rational numbers? This also seems to be the argument of DFa, that Dirichlet priors put zero prior probability on continuous distributions and therefore the inconsistency of Dirichlet priors with respect to continuous distributions in some applications is a problem. Others might argue that the distribution of any observable phenomenon must be discrete and that continuous models are merely useful approximations, in which case the issue being called in question for Dirichlet processes is the usefulness of continuous approximations.

Nothing in the Bayesian machinery will ensure conditional consistency everywhere. That requires assumptions on the conditional distributions over and above the Bayesian paradigm. However, such assumptions may well be valid considerations when developing models for data.

2 Technical Details

Let $Y = (y_1, \ldots, y_n)'$ and consider the probability $\Pr[\theta \in A \text{ and } Y \in B]$ for arbitrary Borel sets A and B. Let $\mathbf{1}_{[A \times B]}(\theta, Y)$ be the indicator function of the set $A \times B$. The conditional probability $\Pr[\theta \in A | Y = w]$ can be defined as a Y measurable function such that for any set B

$$\int_B \Pr[\theta \in A | Y = w] dP(\theta, Y) = \int \mathbf{1}_{[A \times B]}(\theta, Y) dP(\theta, Y), \tag{1}$$

see Rao (1973, p. 91) or Berry and Christensen (1979).

First of all, the joint distribution of (θ, Y) exists. The joint density (θ, Y) is $h(\theta, Y) \equiv f(Y|\theta)g(\theta)$. This is clearly dominated by taking $g(\theta)$ the same and replacing $f(Y|\theta)$ with a finite multiple of a Cauchy(θ) density. Since the integral exists, we can apply Fubini's theorem.

Let $f^*(y|\theta)$ be the density for a $N(\theta, 1)$ distribution. We show that

$$\Pr[\theta \in A | Y = w] = \int_A f(\theta|Y) d\theta$$

where

$$f(\theta|Y) = \frac{f^*(y|\theta)g(\theta)}{\int f^*(y|\theta)g(\theta)d\theta}.$$

Thus, this version of the posterior probability behaves as if there were no Cauchy components to the sampling distribution at all. The claims of the previous section follow immediately from this result. To see the validity of the result, observe that

$$\int \mathbf{1}_{[A \times B]}(\theta, Y) dP(\theta, Y) = \int_A \int_B f(Y|\theta)g(\theta) dY d\theta.$$

However, $f(Y|\theta)$ and $f^*(Y|\theta)$ are equal almost everywhere, so $\int_B f(Y|\theta)dY = \int_B f^*(Y|\theta)dY$ almost everywhere and

$$\int_A \int_B f(Y|\theta)g(\theta) dY d\theta = \int_A \int_B f^*(Y|\theta)g(\theta) dY d\theta.$$

The distribution associated with $f^*(Y|\theta)$ is perfectly well behaved, so Bayes theorem can be applied to give

$$\int_A \int_B f^*(Y|\theta)g(\theta) dY d\theta = \int_B \int_A f(\theta|Y) f(Y) d\theta dY.$$

It follows that equation (1) holds.

References

Berliner, L. M. and MacEachern, S. N. (1993). "Examples of inconsistent Bayes procedures based on observations on dynamical systems." *Statistics and Probability Letters*, 17: 355–360. 759

Berry, D. A. and Christensen, R. (1979). "Empirical Bayes estimation of a binomial parameter via mixtures of Dirichlet processes." *Annals of Statistics*, 7: 558–568. 760

Datta, S. (1991). "On the consistency of posterior mixtures and its applications." *Annals of Statistics*, 19: 338–353. 759

Diaconis, P. and Freedman, D. (1986a). "On the consistency of Bayes estimates." *Annals of Statistics*, 14: 1–26. 759

— (1986b). "On inconsistent Bayes estimates of location." *Annals of Statistics*, 14: 68–87. 759

— (1990). "On the uniform consistency of Bayes estimates for multinomial probabilities." *Annals of Statistics*, 18: 1317–1327. 759

— (1993). "Nonparametric binary regression: A Bayesian approach." *Annals of Statistics*, 21: 2108–2137. 759

Halpern, E. F. (1974). "Posterior consistency for coefficient estimation and model selection in the general linear hypothesis." *Annals of Statistics*, 2: 703–712. 759

Meeden, G. and Ghosh, M. (1981). "Admissibility in finite problems." *Annals of Statistics*, 9: 846–852. 759

Rao, C. R. (1973). *Linear Statistical Inference and its Applications*. New York: John Wiley and Sons, second edition. 760

Rukhin, A. L. (1994). "Recursive testing of multiple hypotheses: Consistency and efficiency of the Bayes rule." *Annals of Statistics*, 22: 616–633. 759

Stone, M. (1976). "Strong inconsistency from uniform priors." *Journal of the American Statistical Association*, 71: 114–116. 759