Journal of the Korean Statistical Society 44 (2015) 261–270

Contents lists available at ScienceDirect

Journal of the Korean Statistical Society

journal homepage: www.elsevier.com/locate/jkss

Approximated sensitivity analysis in posterior predictive distribution

Yongku Kim [a], L. Mark Berliner [b], Dal Ho Kim [a,*]

[a] *Department of Statistics, Kyungpook National University, South Korea*

[b] *Department of Statistics, The Ohio State University, United States*

ARTICLE INFO

Article history:
Received 1 April 2014
Accepted 22 September 2014
Available online 22 October 2014

AMS 2000 subject classifications:
62F15
62B30

Keywords:
Bayesian sensitivity
Kullback–Leibler divergence
Laplace approximation
Posterior predictive distribution

ABSTRACT

In Bayesian statistics, a model can be assessed by checking that the model fits the data, which is addressed by using the posterior predictive distribution for a discrepancy, an extension of classical test statistics to allow dependence on unknown (nuisance) parameters. Posterior predictive assessment of model fitness allows more direct assessment of the discrepancy between data and the posited model. The sensitivity analysis revealed that the effect of priors on parameter inferences is different from their effect on marginal density and predictive posterior distribution. In this paper, we explore the effect of the prior (or posterior) distribution on the corresponding posterior predictive distribution. The approximate sensitivity of the posterior predictive distribution is studied in terms of information measure including the Kullback–Leibler divergence. As an illustration, we applied these results to the simple spatial model settings.

1. Introduction

Assessing the plausibility of assumptions is always fundamental, especially in Bayesian data analyses. A Bayesian analysis can be very misleading when the model is far from plausible. Therefore any meaningful Bayesian analysis should include a check to see if the posited model should be excluded because it fails to provide a reasonable summary of the data at hand. In Bayesian statistics, a model can be checked in at least three ways: (1) examining sensitivity of inferences to reasonable changes in the prior distribution and the likelihood; (2) checking that the posterior inferences are reasonable, given the substantive context of the model; and (3) checking that the model fits the data (for a general review, Berger, 1985; Bernardo & Smith, 2000). Especially, the third of these concerns is addressed by using the posterior predictive distribution for a discrepancy, an extension of classical test statistics to allow dependence on unknown (nuisance) parameters. Posterior predictive assessment was introduced by Guttman (1967), applied by Rubin (1981), and given a formal Bayesian definition by Rubin (1984). Posterior predictive assessment of model fitness allows more direct assessment of the discrepancy between data and the posited model. The posterior predictive distribution is the distribution of unobserved observations (prediction) conditional on the observed data. That is, the posterior predictive distribution is an integral of the likelihood function with respect to the posterior distribution. A sample from a posterior predictive distribution can be generated based on draws from the posterior distribution of parameter of interest. The posterior predictive distribution can be used to check whether the model is consistent with data. For more information about using predictive distribution as a model checking tool see Gelman, Carlin, Stern, and Rubin (2004).

* Corresponding author.
E-mail address: dalkim@knu.ac.kr (D.H. Kim).

http://dx.doi.org/10.1016/j.jkss.2014.09.002
1226-3192/

The sensitivity analysis revealed that the effect of priors on parameter inferences is different from their effect on Bayes factors. In obtaining posteriors, priors can usually be conceived as representing some minimal amount of information, whose influence will be outweighed as more data are observed. When the data are highly informative or precise, the posterior density converges to the likelihood, and the effect of the prior may be negligible. In contrast, the marginal likelihood is highly sensitive to prior assumptions (Aitkin, 2001; Sinharay & Stern, 2002). The pervasive nature of this divergence means that one is often forced to choose between inferences by posterior distributions, or inferences by Bayes factors (Kass & Raftery, 1995).

Let $\mathbf{y}$ denote the observed data, which can be scalar or vector valued and suppose that $\mathbf{y} \sim p(\mathbf{y}|\theta)$, where $\theta \in \Theta$ is m-dimensional unknown model parameter. To avoid confusion with the observed data, $\mathbf{y}$, define $\mathbf{z}$ as the replicated data that could have been observed, or, to think predictively, as the data that would appear if the experiment that produced $\mathbf{y}$ today were replicated tomorrow with the same model, the same (unknown) value of θ that produced $\mathbf{y}$. In the Bayesian framework, the inference for θ is provided by its posterior distribution, $\pi(\theta|\mathbf{y}) \propto p(\mathbf{y}|\theta)\pi(\theta)$, where $\pi(\theta)$ is the prior distribution. Correspondingly, the reference distribution of the future observation $\mathbf{z}$, given $p(\mathbf{z}|\theta)$, is its posterior predictive distribution,

$$p(\mathbf{z}|\mathbf{y}) = \int p(\mathbf{z}|\theta)\pi(\theta|\mathbf{y})d\theta. \tag{1}$$

Note that $p(\mathbf{z}|\theta, \mathbf{y})$ can be used instead of the reference distribution $p(\mathbf{z}|\theta)$ in the spatial prediction. In general, it may be often feasible to integrate out the parameter either analytically or numerically. Main interest lies on the impacts on sensitivity of Bayesian inferences based on two posterior candidates $\pi(\theta|\mathbf{y})$ and $\pi^*(\theta|\mathbf{y})$. Using alternate model probability distribution $p^*(\mathbf{y}|\theta)$ or prior distribution function $\pi^*(\theta)$ instead of the current distribution function $p(\mathbf{y}|\theta)$ or $\pi(\theta)$ leads to another posterior predictive distribution $p^*(\mathbf{z}|\mathbf{y})$. The point is that the posterior distribution $\pi(\theta|\mathbf{y})$ is a flexible and manageable approximation by considering the Gaussian approximation or the Laplace approximation to $\pi(\theta|\mathbf{y})$. Under standard conditions, the Laplace approximation of a posterior density has error rare $O(n^{-1})$ (Tierney & Kadane, 1987).

In this paper, we explore the sensitivity of predictive posterior distribution to model probability or prior distribution. The approximate sensitivity of the posterior distribution of interest is studied in terms of information measure including the Kullback–Leibler divergence.

2. Sensitivity analysis

Using an approximating density $\pi^*(\theta|\mathbf{y})$ instead of the true density $\pi(\theta|\mathbf{y})$ leads to an approximate joint distribution, $p^*(\mathbf{z}, \theta|\mathbf{y}) = p(\mathbf{z}|\theta)\pi^*(\theta|\mathbf{y})$. Define a simple pointwise difference measure between the corresponding posterior predictive densities $p(\mathbf{z}|\mathbf{y})$ and $p^*(\mathbf{z}|\mathbf{y})$ as

$$d_{\mathbf{z}} = \left|p(\mathbf{z}|\mathbf{y}) - p^*(\mathbf{z}|\mathbf{y})\right|, \tag{2}$$

where $p(\mathbf{z}|\mathbf{y}) = \int p(\mathbf{z}, \theta|\mathbf{y})d\theta$ and $p^*(\mathbf{z}|\mathbf{y}) = \int p^*(\mathbf{z}, \theta|\mathbf{y})d\theta$.

As a special case, the posterior distribution $\pi_n(\mathbf{z}|\mathbf{y})$ can be chosen directly by approximation, such as the second order approximation or the Edgeworth expansion (Wallace, 1958). That is, $\pi(\mathbf{z}|\mathbf{y}) = \pi_n(\mathbf{z}|\mathbf{y}) + \epsilon_n(\theta)$, where the $\epsilon_n(\theta)$ term is of $o(n^{-1/2})$, $o(n^{-1})$, $o(n^{-2})$, $O(n^{-1/2})$, $O(n^{-1})$ or $O(n^{-2})$ according to approximation method used.

Since for each $\mathbf{z}$,

$$p(\mathbf{z}|\mathbf{y}) - p_n(\mathbf{z}|\mathbf{y}) = \int p(\mathbf{z}|\theta)\pi(\theta|\mathbf{y})d\theta - \int p(\mathbf{z}|\theta)\pi_n(\theta|\mathbf{y})d\theta \tag{3}$$

$$= \int p(\mathbf{z}|\theta)\pi(\theta|\mathbf{y})d\theta - \int p(\mathbf{z}|\theta)\left[\pi(\theta|\mathbf{y}) - \epsilon_n(\theta)\right]d\theta \tag{4}$$

$$= \int p(\mathbf{z}|\theta) \cdot \epsilon_n(\theta)d\theta, \tag{5}$$

$d_{\mathbf{z}} = |p(\mathbf{z}|\mathbf{y}) - p_n(\mathbf{z}|\mathbf{y})| = \left|\int p(\mathbf{z}|\theta) \cdot \epsilon_n(\theta)d\theta\right| \le \int p(\mathbf{z}|\theta)|\epsilon_n(\theta)|d\theta$. For some $e > 0$, if $\int p(\mathbf{z}|\theta)|\epsilon_n(\theta)|d\theta \le e$ for all $\mathbf{z}$, then $d_{\mathbf{z}} < e$ for all $\mathbf{z}$. Let $\hat{d} = \int d_{\mathbf{z}}d\mathbf{z}$ and $\tilde{d} = \sup_{\mathbf{z}}\{d_{\mathbf{z}}\}$, then

$$\hat{d} \le \int |\epsilon_n(\theta)|d\theta \quad \text{and} \quad \tilde{d} \le \sup_{\mathbf{z}} \int p(\mathbf{z}|\theta)|\epsilon_n(\theta)|d\theta. \tag{6}$$

Suppose that $\hat{\theta}$ is the generalized MLE of θ. Then $\pi(\mathbf{z}|\mathbf{y})$ can be approximated by an approximating Gaussian distribution $\pi_n(\theta|\hat{\theta})$. Under the regularity conditions for asymptotic normality of the posterior distribution,

$$d_{\mathbf{z}} = |p(\mathbf{z}|\mathbf{y}) - p_n(\mathbf{z}|\mathbf{y})| \to 0 \quad \text{as } n \to \infty \text{ provided } \int p(\mathbf{z}|\theta)d\theta < \infty, \tag{7}$$

since for each $\theta \in \Theta$,

$$\epsilon_n(\theta) = \pi_n(\theta|\hat{\theta}) - \pi(\theta|\mathbf{y}) \to 0 \quad \text{as } n \to \infty. \tag{8}$$