Article

Approximated Information Analysis in Bayesian Inference

Jung In Seo [1] and Yongku Kim [2],*

[1] Department of Statistics, Yeungnam University, Gyeongsan 712-749, Korea;
  E-Mail: leehoo1928@ynu.ac.kr

[2] Department of Statistics, Kyungpook National University, Daegu 702-701, Korea

* Author to whom correspondence should be addressed; E-Mail: kim.1252@knu.ac.kr;
  Tel.: +82-53-950-5365.

Academic Editor: Kevin H. Knuth

Abstract: In models with nuisance parameters, Bayesian procedures based on Markov Chain Monte Carlo (MCMC) methods have been developed to approximate the posterior distribution of the parameter of interest. Because these procedures require burdensome computations related to the use of MCMC, approximation and convergence in these procedures are important issues. In this paper, we explore Gibbs sensitivity by using an alternative to the full conditional distribution of the nuisance parameter. The approximate sensitivity of the posterior distribution of interest is studied in terms of an information measure, including Kullback–Leibler divergence. As an illustration, we then apply these results to simple spatial model settings.

Keywords: Bayesian sensitivity; Gibbs sampler; Kullback–Leibler divergence; Laplace approximation

1. Introduction

Let d denote the data, which can be scalar- or vector-valued, and suppose that $d \sim p(d|\theta, \beta)$, where $\theta \in \Theta$ is the parameter of interest and $\beta \in \boldsymbol{B}$ is a nuisance parameter. Realizations from the joint posterior distribution $\pi(\theta, \beta|d)$ can be produced by independent sampling based on $\pi(\theta, \beta|d) = \pi(\theta|\beta, d) \times \pi(\beta|d)$ or, if $\pi(\beta|d)$ is intractable, by Gibbs sampling based on the full conditional distributions $\pi(\theta|\beta, d)$ and $\pi(\beta|\theta, d)$. Since β is a nuisance parameter, our primary interest

is in the marginal posterior distribution $\pi(\theta|d) = \int \pi(\theta, \beta|d)d\beta$. In general, it may often be feasible to integrate out some nuisance parameters either analytically or numerically. Missing data problems are brought into this framework by augmenting the observed data d with latent data β.

The latent variable/nuisance parameter scenario is a commonly studied one in the literature. One issue in nuisance parameter problems is the relationship between $\pi(\beta|d)$ and $\pi(\theta|d)$. Our main interest lies in the impacts on the sensitivity of inferences based on the target marginal posterior distribution $\pi(\theta|d)$ compared with an approximation based on choosing $\pi^*(\beta|d)$, an alternative to the posterior distribution of the nuisance parameter, instead of $\pi(\beta|d)$. The point is that $\pi^*(\beta|d)$ is a flexible and manageable approximation to an unmanageable $\pi(\beta|d)$. For application to simple spatial model settings, we can consider the Gaussian approximation or the Laplace approximation to $\pi(\beta|d)$. For example, we can use

$$\pi^*(\beta|d) \propto \left.\frac{\pi(\theta, \beta|d)}{\tilde{\pi}_G(\theta|\beta, d)}\right|_{\theta=\hat{\theta}(\beta)},$$

where $\tilde{\pi}_G(\theta|\beta, d)$ is the Gaussian approximation to $\pi(\theta|\beta, d)$. Under "standard conditions," the Laplace approximation of a marginal posterior density has the error rate $O(n^{-1})$ [1].

State-of-the-art Markov Chain Monte Carlo (MCMC) approaches to posterior inference typically revolve around reparameterizations. Yu and Meng introduce an alternative strategy for boosting MCMC efficiency by simply interweaving—but not alternating—two parameterizations, namely the centered parameterization and the non-centered parameterization, to ensure effective MCMC implementation [2]. Filippone and Girolamipresent a pseudo-marginal MCMC approach to account for uncertainty in the model parameters when making model-based predictions on out-of-sample data [3]. Attias presents the Variational Bayes framework, which provides a solution for the structure of models with latent variables [4]. Here, Kullback–Leibler divergence is minimized between the posterior and a typically exponential family approximation [4]. Expectation propagation is similar in nature [5]. The integrated nested Laplace approximation is considered for the approximate Bayesian inference in latent Gaussian models [6].

In this paper, we explore Gibbs sensitivity by using an alternative to the full conditional distribution of the nuisance parameter. The approximate sensitivity of the posterior distribution of interest is studied in terms of an information measure including Kullback–Leibler divergence. As an illustration, we apply the proposed approach to PRUDENCE (Prediction of Regional scenarios and Uncertainties for Defining EuropeaN Climate change risks and Effects; http://prudence.dmi.dk/) ensemble of regional climate models over Central Europe (about 8000 grid points), which involves the analysis of large quantities of data. Furrer and Sain combine two techniques, namely tapering and backfitting, to model and analyze these spatial datasets [7]. Kim and Kimpropose an approximate likelihood function of the spatial correlation parameter based on PRUDENCE data [8]. This paper thus provides some information analysis for their approaches.

The rest of the paper is organized as follows. In Section 2, we describe the approximation setting in the Bayesian computation and discuss some of its sensitivity issues. Various theoretical results on the information analysis are provided in the subsequent section. As an illustration, the results from Section 3 are applied to simple spatial model settings in Section 4, and Section 5 concludes the study.

2. Sensitivity Issues

Let $\pi(\theta|\beta, d)$ and $\pi(\beta|\theta, d)$ be the full conditional distributions for the joint posterior distribution $\pi(\theta, \beta|d)$, which can be expressed in terms of these full conditional distributions. The consistency conditions on the full conditional distributions are required to reconstruct the joint posterior distribution (e.g., see the Hammersley–Clifford Theorem in [9]). Let $\pi^*(\beta|\theta, d)$ be an approximated full conditional distribution of $\pi(\beta|\theta, d)$. Under the regulation condition with reference points β_0 and θ_0, the joint posterior distributions can be written in the form

$$\pi(\theta, \beta|d) \propto \frac{\pi(\beta|\theta = \theta_0, d)}{\pi(\beta = \beta_0|\theta = \theta_0, d)} \frac{\pi(\theta|\beta, d)}{\pi(\theta = \theta_0|\beta, d)} \propto g(\beta)\pi(\theta|\beta, d)$$

and

$$\pi^*(\theta, \beta|d) \propto \frac{\pi^*(\beta|\theta = \theta_0, d)}{\pi^*(\beta = \beta_0|\theta = \theta_0, d)} \frac{\pi(\theta|\beta, d)}{\pi(\theta = \theta_0|\beta, d)} \propto g^*(\beta)\pi(\theta|\beta, d),$$

where

$$g(\beta) = \frac{\pi(\beta|\theta = \theta_0, d)}{\pi(\theta = \theta_0|\beta, d)} \text{ and } g^*(\beta) = \frac{\pi^*(\beta|\theta = \theta_0, d)}{\pi(\theta = \theta_0|\beta, d)}.$$

Therefore,

$$\pi(\beta|d) = Cg(\beta) \text{ and } \pi^*(\beta|d) = C^*g^*(\beta),$$

where C and C^* are normalizing constants, that is,

$$C = \left[\int g(\beta)d\beta\right]^{-1} \text{ and } C^* = \left[\int g^*(\beta)d\beta\right]^{-1}.$$

Note that C and C^* can be explicated by

$$C = \frac{\pi(\beta = \beta_0, \theta = \theta_0|d)}{\pi(\beta = \beta_0|\theta = \theta_0, d)} \text{ and } C^* = \frac{\pi^*(\beta = \beta_0, \theta = \theta_0|d)}{\pi^*(\beta = \beta_0|\theta = \theta_0, d)},$$

where $\pi(\beta = \beta_0, \theta = \theta_0|d)$ and $\pi^*(\beta = \beta_0, \theta = \theta_0|d)$ can be obtained from a Monte Carlo computation.

Using an approximating density $\pi^*(\beta|d)$ instead of the true density $\pi(\beta|d)$ leads to an approximate joint posterior distribution, $\pi^*(\theta, \beta|d) = \pi(\theta|\beta, d)\pi^*(\beta|d)$. Define a simple pointwise difference measure between the corresponding marginal posterior densities $\pi(\theta|d)$ and $\pi^*(\theta|d)$ as

$$d_\theta = \big|\pi(\theta|d) - \pi^*(\theta|d)\big|,$$

where $\pi(\theta|d) = \int \pi(\theta, \beta|d)d\beta$ and $\pi^*(\theta|d) = \int \pi^*(\theta, \beta|d)d\beta$.

Alternatively, the bounds for the (log) ratio of $\pi(\theta|d)$ and $\pi^*(\theta|d)$ can be obtained in terms of the full conditional distribution of the nuisance parameter $\pi(\beta|\theta, d)$ (see [10]).

Theorem 1. *Suppose that for some set $B_0 \subseteq \boldsymbol{B}$ and $\alpha \in \Re$, $1 \leq \frac{\pi(\beta|d)}{\pi^*(\beta|d)} \leq 1 + \alpha$ for all $\beta \in B_0$. Then, for all $\theta \in \Theta$,*

$$P^*(\beta \in B_0|\theta, d) \leq \frac{\pi(\theta|d)}{\pi^*(\theta|d)} \leq \frac{1 + \alpha}{P(\beta \in B_0|\theta, d)},$$

where P and P^* are the probability measures for the conditional distributions of β given θ and d, based on $\pi(\beta|d)$ and $\pi^*(\beta|d)$, respectively.

Proof. Since

$$
\begin{aligned}
\pi(\theta|d) &\geq \int_{B_0} \pi(\theta|\beta,d)\pi(\beta|d)d\beta \\
&\geq \int_{B_0} \pi(\theta|\beta,d)\pi^*(\beta|d)d\beta \\
&= \pi^*(\theta|d)P^*(\beta \in B_0|\theta,d),
\end{aligned}
$$

and

$$
\begin{aligned}
\pi^*(\theta|d) &\geq \int_{B_0} \pi(\theta|\beta,d)\pi^*(\beta|d)d\beta \\
&\geq \frac{1}{1+\alpha}\int_{B_0} \pi(\theta|\beta,d)\pi(\beta|d)d\beta \\
&= \frac{P(\beta \in B_0|\theta,d)}{1+\alpha}\pi(\theta|d),
\end{aligned}
$$

thus,

$$
P^*(\beta \in B_0|\theta,d) \leq \frac{\pi(\theta|d)}{\pi^*(\theta|d)} \quad \text{and} \quad \frac{\pi(\theta|d)}{\pi^*(\theta|d)} \leq \frac{1+\alpha}{P(\beta \in B_0|\theta,d)}.
$$

Suppose that there exists $\gamma \in \Re$ such that $\frac{\pi(\beta|d)}{\pi^*(\beta|d)} \leq \gamma$ for all $\beta \in \boldsymbol{B}$. Then, for all $\beta \in \boldsymbol{B}$,

$$
\frac{P(\beta \in \boldsymbol{B} \setminus B_0|\theta,d)}{P(\beta \in B_0|\theta,d)} \leq \gamma\frac{P^*(\beta \in \boldsymbol{B} \setminus B_0|\theta,d)}{P^*(\beta \in B_0|\theta,d)}
$$

and

$$
\frac{\pi(\beta|\theta,d)}{\pi^*(\beta|\theta,d)} \leq \frac{\pi(\beta|\theta,d)}{\pi^*(\beta|d)P^*(\beta \in B_0|\theta,d)}.
$$

$\square$

This result can be extended to the Kullback–Leibler divergence of $\pi^*(\theta|d)$ from $\pi(\theta|d)$.

In practice, the marginal posterior distribution of the parameter of interest is hard to calculate analytically. Suppose $\pi(\theta|\beta,d)$ is a smooth (positive) function of β. By using Laplace's method, $\pi(\theta|d)$ is approximated as

$$
\pi(\theta|d) = \int \pi(\theta|\beta,d)\exp[-nh(\beta)]d\beta, \text{ where } h(\beta) = -\frac{1}{n}\log \pi(\beta|d).
$$

Assume $\hat{\beta}$ maximizes $\log\pi(\beta|d)$, and define $\Sigma = \left(-\frac{\partial^2}{\partial\beta^2}h(\beta)\right)^{-1}_{\beta=\hat{\beta}}$. Then, $\pi(\theta|d)$ can be well approximated by Laplace's method. That is,

$$
\pi(\theta|d) \approx \pi(\theta|\hat{\beta},d)\pi(\hat{\beta}|d)\left(\frac{2\pi}{n}\right)^{m/2}|\Sigma|^{1/2}\left(1 + \frac{\frac{\partial^2}{\partial\beta^2}\pi(\theta|\beta,d)}{2\pi(\theta|\beta,d)}\frac{\Sigma}{n}\right)_{\beta=\hat{\beta}},
$$

where m is the dimension of β. More generally, we have

$$
\pi(\theta|d) = \hat{\pi}(\theta|d)\left[1 + O_\theta\left(\frac{1}{n}\right)\right], \text{ where } \hat{\pi}(\theta|d) = \pi(\theta|\hat{\beta},d)\pi(\hat{\beta}|d)\left(\frac{2\pi}{n}\right)^{m/2}|\Sigma|^{1/2}.
$$

Laplace's method requires three conditions, referred to as Laplace regularity: (1) the integrals in the equation must exist and be finite; (2) the determinant of the Hessians must be bounded away from

zero at the optimizers; and (3) the log-likelihood must be differentiable on the parameters and all the partial derivatives be bounded in the neighborhood of the optimizers. These conditions imply, under mild assumptions, the asymptotic normality of the posterior.

Based on the above results, the bounds for the differences between the marginal posterior distributions can be approximated. Suppose that $|\pi(\beta|d) - \pi^*(\beta|d)|$ has a unique maximum at $\hat{\beta}$. Then,

$$\int |\pi(\theta|d) - \pi^*(\theta|d)|d\theta \leq \big(\frac{2\pi}{n}\big)^{m/2}|\tilde{\Sigma}|^{1/2}|\pi(\hat{\beta}|d) - \pi^*(\hat{\beta}|d)|\Big[1 + O\big(\frac{1}{n}\big)\Big],$$

where $O(n^{-1})$ does not depend on θ.

3. Information Analysis

3.1. Approximation to Kullback–Leibler Divergence

Kullback–Leibler divergence, or relative entropy, is a quantity that measures the difference between two probability distributions [11]. The Kullback–Leibler divergence between $\pi(\theta, \beta|d)$ and $\pi^*(\theta, \beta|d)$ is the same as that between $\pi(\beta|d)$ and $\pi^*(\beta|d)$. Theorem 1 provides the upper bound for the Kullback–Leibler divergence between $\pi(\theta|d)$ and $\pi^*(\theta|d)$, denoted by $I(\pi(\theta|d), \pi^*(\theta|d))$:

$$I(\pi(\theta|d), \pi^*(\theta|d)) \leq \log(1 + \alpha) - E_{\theta|d}\log P(\beta \in B_0|\theta, d).$$

For sufficiently large $P(\beta \in B_0|\theta, d)$ and small α,

$$\log(1 + \alpha) - E_{\theta|d}\log P(\beta \in B_0|\theta, d) \approx \int_{\beta \notin B_0} \pi(\beta|\theta, d)\pi(\theta|d)d\theta.$$

Thus, the upper bound can be approximated by $1 - P(\beta \in B_0|d) = P(\beta \notin B_0|d)$. However, the analytic calculation of the Kullback–Leibler divergence between $\pi(\theta|d)$ and $\pi^*(\theta|d)$ is usually difficult. Here, we introduce an approximation to Kullback–Leibler divergence based on Laplace's method along with its convergence properties.

Theorem 2. *Suppose that $\hat{\beta}$ and $\hat{\beta}^*$ maximize $\pi(\beta|d)$ and $\pi^*(\beta|d)$, respectively. Then, we have*

$$I(\pi, \pi^*) \equiv I(\pi(\theta|d), \pi^*(\theta|d)) = \hat{I}_1(\pi, \pi^*)\Big[1 + O\big(\frac{1}{n}\big)\Big],$$

where

$$\hat{I}_1(\pi, \pi^*) = \int \big(\frac{2\pi}{n}\big)^{m/2}|\Sigma|^{1/2}\pi(\theta|\hat{\beta}, d)\pi(\hat{\beta}|d)log\frac{|\Sigma|^{1/2}\pi(\theta|\hat{\beta}, d)\pi(\hat{\beta}|d)}{|\Sigma^*|^{1/2}\pi(\theta|\hat{\beta}^*, d)\pi^*(\hat{\beta}^*|d)}d\theta$$

and

$$\Sigma = \frac{\partial^2}{\partial\beta^2}\frac{1}{n}\log\pi(\beta|d)\Big|_{\beta=\hat{\beta}} \quad and \quad \Sigma^* = \frac{\partial^2}{\partial\beta^2}\frac{1}{n}\log\pi^*(\beta|d)\Big|_{\beta=\hat{\beta}^*}.$$

It is noted that $\hat{I}_1(\pi, \pi^*) \geq 0$ when $|\Sigma| \geq |\Sigma^*|$. Furthermore, if $\hat{\beta} = \hat{\beta}^*$ and $|\Sigma|^{1/2} = |\Sigma^*|^{1/2}$, then

$$\hat{I}_1(\pi, \pi^*) = \big(\frac{2\pi}{n}\big)^{m/2}|\Sigma|^{1/2}\pi(\hat{\beta}|d)\log\frac{\pi(\hat{\beta}|d)}{\pi^*(\hat{\beta}|d)} \geq 0.$$

Proof. Since

$$I(\pi, \pi^*) \equiv \int \pi(\theta|d) \log \frac{\pi(\theta|d)}{\pi^*(\theta|d)} d\theta$$

$$= \int \hat{\pi}(\theta|d) \Big[1 + O_\theta\big(\frac{1}{n}\big)\Big] \log \frac{\hat{\pi}(\theta|d)\Big[1 + O_\theta\big(\frac{1}{n}\big)\Big]}{\hat{\pi}^*(\theta|d)\Big[1 + O_\theta\big(\frac{1}{n}\big)\Big]} d\theta$$

$$= \int \hat{\pi}(\theta|d) \log \frac{\hat{\pi}(\theta|d)}{\hat{\pi}^*(\theta|d)} \Big[1 + O_\theta\big(\frac{1}{n}\big)\Big] d\theta$$

$$= \int \hat{\pi}(\theta|d) \log \frac{\hat{\pi}(\theta|d)}{\hat{\pi}^*(\theta|d)} d\theta \Big[1 + O\big(\frac{1}{n}\big)\Big]$$

and

$$\int \hat{\pi}(\theta|d) \log \frac{\hat{\pi}(\theta|d)}{\hat{\pi}^*(\theta|d)} d\theta = \int \pi(\theta|\hat{\beta}, d) \pi(\hat{\beta}|d) \big(\frac{2\pi}{n}\big)^{m/2} |\Sigma|^{1/2} \times \log \frac{\pi(\theta|\hat{\beta}, d) \pi(\hat{\beta}|d) \big(\frac{2\pi}{n}\big)^{m/2} |\Sigma|^{1/2}}{\pi(\theta|\hat{\beta}^*, d) \pi^*(\hat{\beta}^*|d) \big(\frac{2\pi}{n}\big)^{m/2} |\Sigma^*|^{1/2}} d\theta,$$

where

$$\Sigma = \Big(-\frac{\partial^2}{\partial\beta^2} \frac{1}{n} \log \pi(\beta|d)\Big)^{-1}_{\beta=\hat{\beta}} \quad \text{and} \quad \Sigma^* = \Big(-\frac{\partial^2}{\partial\beta^2} \frac{1}{n} \log \pi^*(\beta|d)\Big)^{-1}_{\beta=\hat{\beta}^*},$$

we have

$$I(\pi, \pi^*) = \hat{I}_1(\pi, \pi^*) \Big[1 + O\big(\frac{1}{n}\big)\Big],$$

where

$$\hat{I}_1(\pi, \pi^*) = \int \big(\frac{2\pi}{n}\big)^{m/2} |\Sigma|^{1/2} \pi(\theta|\hat{\beta}, d) \pi(\hat{\beta}|d) \log \frac{|\Sigma|^{1/2} \pi(\hat{\beta}|d)}{|\Sigma^*|^{1/2} \pi^*(\hat{\beta}|d)} d\theta$$

$$= \big(\frac{2\pi}{n}\big)^{m/2} |\Sigma|^{1/2} \pi(\hat{\beta}|d) \log \frac{|\Sigma|^{1/2} \pi(\hat{\beta}|d)}{|\Sigma^*|^{1/2} \pi^*(\hat{\beta}|d)}.$$

□

Under the weak conditions allowing the exchange of integral and limit, it can be shown that $\hat{I}_1(\pi, \pi^*)$ converges to $I(\pi, \pi^*)$. Suppose that both $\pi(\beta|d)$ and $\pi^*(\beta|d)$ are maximized at $\hat{\beta}$, a posterior mode. Then,

$$|\hat{I}_1(\pi, \pi^*) - I(\pi, \pi^*)| \to 0 \text{ as } n \to \infty.$$

Since $\sup_\theta |\hat{\pi}(\theta|d) - \pi(\theta|d)| \to 0$ and $\sup_\theta |\hat{\pi}^*(\theta|d) - \pi^*(\theta|d)| \to 0$ as $n \to \infty$, continuity implies that

$$\Big| \int \hat{\pi}(\theta|d) \log \frac{\hat{\pi}(\theta|d)}{\hat{\pi}^*(\theta|d)} d\theta - \int \pi(\theta|d) \log \frac{\pi(\theta|d)}{\pi^*(\theta|d)} d\theta \Big|$$

$$\leq \int \Big| \hat{\pi}(\theta|d) \log \frac{\hat{\pi}(\theta|d)}{\hat{\pi}^*(\theta|d)} - \pi(\theta|d) \log \frac{\pi(\theta|d)}{\pi^*(\theta|d)} \Big| d\theta \to 0 \quad \text{as} \quad n \to \infty.$$

Therefore,

$$\Big| \big(\frac{2\pi}{n}\big)^{m/2} |\Sigma|^{1/2} \pi(\hat{\beta}|d) \log \frac{|\Sigma|^{1/2} \pi(\hat{\beta}|d)}{|\Sigma^*|^{1/2} \pi^*(\hat{\beta}|d)} - I(\pi, \pi^*) \Big| \to 0 \quad \text{as} \quad n \to \infty.$$

Note that the first-order approximation $\hat{I}_1(\pi, \pi^*)$ depends only on the marginal posterior distribution of the nuisance parameters $\pi(\beta|d)$ and $\pi^*(\beta|d)$ but not the full conditional distribution of the parameter of interest $\pi(\theta|\beta, d)$. Based on the asymptotic properties of the posterior distributions (which can be achieved easily under fairly general conditions when the true value of the parameter is in the support of the prior), a Gaussian distribution with mean $\hat{\beta}$, the generalized MLE of β, and variance $\left(- \frac{\partial^2}{\partial \beta^2} \log \pi(\beta|d) \right)^{-1}_{\beta = \hat{\beta}}$ can be considered as $\pi^*(\beta|d)$. That is,

$$\pi^*(\beta|d) \propto \exp\left(- \frac{1}{2}(\beta - \hat{\beta})' H_\beta (\beta - \hat{\beta}) \right), \text{ where } H_\beta \equiv \frac{\partial^2}{\partial \beta^2} \log \pi(\beta|d).$$

For a sufficiently large n, $\pi^*(\beta|d)$ is maximized at $\hat{\beta}$, the posterior mode of $\pi(\beta|d)$, and $\frac{\partial^2}{\partial \beta^2} \log \pi(\beta|d)\big|_{\beta=\hat{\beta}} \approx \frac{\partial^2}{\partial \beta^2} \log \pi^*(\beta|d)\big|_{\beta=\hat{\beta}}$.

3.2. Higher-Order Approximations to Kullback–Leibler Divergence

To improve accuracy, second-order approximations can be considered in Laplace's method [1]. Let $\hat{\beta}_1$, $\hat{\beta}_1^*$, Σ_1, and Σ_1^* be defined as above in the case of the first-order approximations. Let $\hat{\beta}_2$ and $\hat{\beta}_2^*$ maximize $\log[\pi(\theta|\beta, d)\pi(\beta|d)]$ and $\log[\pi(\theta|\beta, d)\pi^*(\beta|d)]$, respectively. While the second-order approximations to $\pi(\theta|d)$ and $\pi(\theta|d)^*$ are more accurate, the second-order approximation to $I(\pi, \pi^*)$ is quite difficult to calculate since both $\hat{\beta}_2$ and $\hat{\beta}_2^*$ are functions of θ, that is, $\hat{\beta}_2 = \hat{\beta}_2(\theta)$ and $\hat{\beta}_2^* = \hat{\beta}_2^*(\theta)$.

Assume that both $\hat{\beta}$ and $\hat{\beta}^*$ maximize $\pi(\theta|\beta, d)\pi(\beta|d)$ and $\pi(\theta|\beta, d)\pi^*(\beta|d)$, respectively, regardless of θ. Then, it can be shown that

$$I(\pi, \pi^*) = \hat{I}_2(\pi, \pi^*)\left[1 + O\left(\frac{1}{n^2}\right) \right]$$

or

$$|\hat{I}_2(\pi, \pi^*) - I(\pi, \pi^*)| \to 0 \text{ as } n \to \infty,$$

where

$$\hat{I}_2(\pi, \pi^*) \equiv \int \frac{|\Sigma_2|^{1/2}}{|\Sigma_1|^{1/2}} \frac{\pi(\theta|\hat{\beta}_2, d)\pi(\hat{\beta}_2|d)}{\pi(\hat{\beta}_1|d)} \log \frac{\frac{|\Sigma_2|^{1/2}}{|\Sigma_1|^{1/2}} \frac{\pi(\theta|\hat{\beta}_2, d)\pi(\hat{\beta}_2|d)}{\pi(\hat{\beta}_1|d)}}{\frac{|\Sigma_2^*|^{1/2}}{|\Sigma_1^*|^{1/2}} \frac{\pi(\theta|\hat{\beta}_2^*, d)\pi^*(\hat{\beta}_2^*|d)}{\pi^*(\hat{\beta}_1^*|d)}} d\theta,$$

$$\Sigma_2 = \left(- \frac{\partial^2}{\partial \beta^2} \frac{1}{n} \log[\pi(\theta|\beta, d)\pi(\beta|d)] \right)^{-1}_{\beta = \hat{\beta}_2},$$

and

$$\Sigma_2^* = \left(- \frac{\partial^2}{\partial \beta^2} \frac{1}{n} \log[\pi(\theta|\beta, d)\pi^*(\beta|d)] \right)^{-1}_{\beta = \hat{\beta}_2^*}.$$

If both $\log[\pi(\theta|\beta, d)\pi(\beta|d)]$ and $\log[\pi(\theta|\beta, d)\pi^*(\beta|d)]$ are maximized at $\hat{\beta}_2$, then the second-order approximation to $I(\pi, \pi^*)$ can be simplified to

$$\hat{I}_2(\pi, \pi^*) = \frac{|\Sigma_2|^{1/2}}{|\Sigma_1|^{1/2}} \frac{\pi(\hat{\beta}_2|d)}{\pi(\hat{\beta}_1|d)} \log \frac{\frac{|\Sigma_2|^{1/2}}{|\Sigma_1|^{1/2}} \frac{\pi(\hat{\beta}_2|d)}{\pi(\hat{\beta}_1|d)}}{\frac{|\Sigma_2^*|^{1/2}}{|\Sigma_1^*|^{1/2}} \frac{\pi^*(\hat{\beta}_2^*|d)}{\pi^*(\hat{\beta}_1^*|d)}}.$$

$\hat{I}_2(\pi, \pi^*) \geq 0$ when $\frac{|\Sigma_2|}{|\Sigma_1|} \geq \frac{|\Sigma_2^*|}{|\Sigma_1^*|}$ and $\pi(\hat{\beta}_1|d) \geq \pi^*(\hat{\beta}_1^*|d)$.

3.3. Approximation to Other Information Measures

Instead of Kullback–Leibler divergence, other useful information measures based on uncertainty functions or the entropy function can be used. One information measure based on uncertainty functions [12] is

$$I = EVar(\beta|d) - EVar^*(\beta|d) \text{ or } I = E \log \frac{Var(\beta|d)}{Var^*(\beta|d)}.$$

Another information measure is based on Renyi's entropy function [13] :

$$I_\alpha(\pi(\theta|d), \pi^*(\theta|d)) = \frac{1}{1-\alpha} \log \int [\pi(\theta|d)]^\alpha [\pi^*(\theta|d)]^{1-\alpha} d\theta.$$

Theorem 3. *Suppose that $\hat{\beta}$ maximizes both $\pi(\beta|d)$ and $\pi^*(\beta|d)$. Then,*

$$I_\alpha(\pi(\theta|d), \pi^*(\theta|d)) = \frac{1}{1-\alpha} \log J\left(1 + O\left(\frac{1}{n}\right)\right),$$

where

$$J \equiv \left(\frac{2\pi}{n}\right)^{m/2} |\Sigma|^{1/2} [\pi(\hat{\beta}|d)]^\alpha [\pi^*(\hat{\beta}|d)]^{1-\alpha}.$$

Proof. Let

$$\delta_\alpha(\pi(\theta|d), \pi^*(\theta|d)) = \exp\left[(1-\alpha) I_\alpha((\pi(\theta|d), \pi^*(\theta|d)))\right].$$

Then, it suffices to show that

$$
\begin{aligned}
\delta_\alpha(\pi(\theta|d), \pi^*(\theta|d)) &= \int [\pi(\theta|d)]^\alpha [\pi^*(\theta|d)]^{1-\alpha} d\theta \\
&= \int \left[\int \pi(\theta|\beta, d)\pi(\beta|d)d\beta\right]^\alpha \left[\int \pi(\theta|\beta, d)\pi^*(\beta|d)d\beta\right]^{1-\alpha} d\theta \\
&= \int \pi(\theta|\hat{\beta}, d)\left(\frac{2\pi}{n}\right)^{m/2} |\Sigma|^{1/2} \left(1 + O_\theta\left(\frac{1}{n}\right)\right) [\pi(\hat{\beta}|d)]^\alpha [\pi^*(\hat{\beta}|d)]^{1-\alpha} d\theta, \\
&= \left(\frac{2\pi}{n}\right)^{m/2} |\Sigma|^{1/2} [\pi(\hat{\beta}|d)]^\alpha [\pi^*(\hat{\beta}|d)]^{1-\alpha} \left(1 + O\left(\frac{1}{n}\right)\right).
\end{aligned}
$$

□

Under similar conditions to those in Theorem 2, it can be shown that

$$\left|\frac{1}{1-\alpha} \log J - I_\alpha(\pi(\theta|d), \pi^*(\theta|d))\right| \to 0 \text{ as } n \to \infty.$$

4. Illustrative Example

We consider an approach to approximating the likelihood function of the spatial correlation parameters in the Gaussian random field. Consider the simple Gaussian random field $\mathbf{Z} \sim MVN(\mathbf{0}, \Sigma(\theta))$, where $\Sigma(\theta)$ is parameterized by a variance term and a correlation function. That is, $\Sigma(\theta) = \sigma^2 R(\theta)$. Then, the likelihood function of (σ^2, θ) is

$$L(\sigma^2, \theta) \propto |\sigma^2 R(\theta)|^{-1/2} \exp\left(-\frac{1}{2\sigma^2}\mathbf{Z}'R^{-1}(\theta)\mathbf{Z}\right). \tag{1}$$

Note that in the problem with a large spatial domain, it is not computationally feasible to compute the likelihood function of the spatial correlation parameters because of R^{-1}.

The proposed approaches are also illustrated in regional climate models, which are used to model the evolution of a climate system over a limited area. These models address smaller spatial regions than global climate models. However, the higher resolution of regional climate models better captures the impact of local features such as lakes and mountains as well as the subgrid-scale atmospheric process. The PRUDENCE project involves regional models over Europe from various climate research centers (http://prudence.dmi.dk/) and employs a major archive of data of a 25-km resolution covering the 1951–2100 transient periods. In the analysis, spatial parameters are estimated based on the approximated likelihood approach using PRUDENCE data (about 8000 grid points). Here, the mean trend in the surface temperature change is modeled as follows:

$$T(s) = \beta_0 + \beta_1 I_{land/sea}(s) + \beta_2 P(s) + \beta_3 lon(s) + \beta_4 lat(s) + \beta_5 elev(s), \tag{2}$$

where $I_{land/sea}(s)$ is an indicator function for the sea and land, $P(s)$ is the amount of seasonal precipitation, $lon(s)$ is the longitude, $lat(s)$ is the latitude, and $elev(s)$ is the elevation in location s. For the detrended surface temperature field, we consider a stationary Gaussian spatial process with an exponential covariance function, $\sigma^2 \exp\left(-\frac{d^2}{2\xi^2}\right)$. For simplicity, we also assume that σ^2 is known and $\theta = 2\xi^2$.

Now we consider an approximated likelihood function for θ. Considering a log transformation for the correlation function ρ leads to

$$\hat{\theta} \sim_{approx} N\left(\theta, \theta^{-4}\left(\sum_{d \in \mathcal{D}} \frac{d^2}{\sigma_d^2}\right)^{-1}\right), \tag{3}$$

where $\hat{\theta}$ is the MLE of the spatial parameter θ,

$$\hat{\theta} = -\frac{\sum_{d \in \mathcal{D}} \frac{d^2}{\sigma_d^2}}{\sum_{d \in \mathcal{D}} \frac{d}{\sigma_d} \log \rho(d)}. \tag{4}$$

Note that the model coefficients $(\beta_0, \beta_1, \beta_2, \beta_3, \beta_4)$ are the parameters of interest and that the correlation parameter θ is the nuisance parameter in this example. Here, the full conditional distribution of the parameter of interest, $\pi(\beta_0, \beta_1, \beta_2, \beta_3, \beta_4 | \theta, d)$, can be obtained based on the model of the mean trend in the surface temperature change, while the true marginal density of the nuisance parameter, $\pi(\beta|d)$, and approximate marginal density of the nuisance parameter, $\pi^*(\beta|d)$, can be computed by using likelihood functions (1) and (3), respectively. Therefore, the full conditional distribution of the parameter of interest, $(\beta_0, \beta_1, \beta_2, \beta_3, \beta_4)$, can be expressed as

$$[\beta_0, \beta_1, \beta_2, \beta_3, \beta_4 | \mathbf{D}, \theta] \propto |\sigma^2 R(\theta)|^{-1/2} \exp\left(-\frac{1}{2\sigma^2}(\mathbf{D} - \mathbf{T})' R^{-1}(\theta)(\mathbf{D} - \mathbf{T})\right), \tag{5}$$

where $\mathbf{D} = \{d(1), d(2), \ldots, d(n)\}$ is the observation vector and $\mathbf{T} = \{T(1), T(2), \ldots, T(n)\}$ is the mean trend vector in (2). The approximated full conditional distribution of the correlation parameter θ is of the form

$$[\theta | \mathbf{D}, \beta_0, \beta_1, \beta_2, \beta_3, \beta_4] \propto \exp\left(-\frac{1}{2}\frac{(\hat{\theta} - \theta)^2}{\theta^4 \sum_{d \in \mathcal{D}} \frac{d^2}{\sigma_d^2}}\right), \tag{6}$$

whereas the exact full conditional distribution is quite similar to (5). Further, $\hat{\theta}$ is the MLE of the spatial parameter θ in (4). The reference points for $(\beta_0, \beta_1, \beta_2, \beta_3, \beta_4)$ and θ are randomly chosen in the neighborhood of the MLEs of the parameters. For more details, see [8].

In the next step, various grid points ($n = 400, 900, 1600$) are randomly chosen and then eliminated as in the simulation study. Table 1 provides the first-order-approximated Kullback–Leibler divergence along with the exact Kullback–Leibler divergence under various settings. The estimated information measures are quite efficient and competitive because the seasonal mean surface temperature fields from the global climate model are already smoothed. Further, the approximated likelihood function of the correlation parameter is not well estimated, particularly when the number of observations is less than 100.

Sampled Grid Size	First-Order-Approximated Kullback–Leibler Distance	Exact Kullback–Leibler Distance
400	0.482	0.425
900	0.279	0.251
1600	0.121	0.108

Table 1. First-order-approximated Kullback–Leibler divergence and the exact Kullback–Leibler divergence under the various settings of the grid points.

5. Summary

We introduced various ways of checking the sensitive effects on the target posterior distribution of the parameter of interest by using an alternative to the full conditional distribution of the nuisance parameter. By using Laplace's method, approximated Kullback–Leibler divergence between $\pi(\theta|d)$ and $\pi^*(\theta|d)$ was also calculated in terms of the entropy of $\pi(\beta|d)$ and $\pi^*(\beta|d)$ at the generalized MLE $\hat{\beta}$. Other information measures provided similar results. However, it is still difficult to check analytically the robustness of the marginal posterior distribution of interest, $\pi(\theta|d)$, according to the choice of the full conditional distribution of the nuisance parameter, $\pi(\beta|\theta, d)$. Nonetheless, for the general class of available marginal posterior distributions of the nuisance parameters, the sensitivity and robustness to the marginal posterior distribution of the parameters of interest can be checked approximately. In addition, we can find a reasonable and flexible substitute for the complicated full conditional distribution of the nuisance parameter under the sensitivity and robustness criteria to the target distribution and then perform inference based on this substitute. Our approach can be applied to future sensitivity analysis on the posterior predictive distribution by using an approximated posterior distribution, Bayes factor or marginal density by using the choice of prior distribution, and expected loss or utility by using an approximated posterior distribution.

Author Contributions

Jung In Seo and Yongku Kim conceived the idea and developed the mothod presented in this paper. Both authors performed data analysis and wrote the paper. Both authors have read and approved the final manuscript.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Tierney, L.; Kadane, J. B. Accurate approximations for posterior moments and marginal densities. *J. Am. Stat. Assoc.* **1986**, *81*, 82–86.
2. Yu, Y.; Meng, X.L. To center or not to center: That is not the question—An Ancillarity-Sufficiency Interweaving Strategy (ASIS) for boosting MCMC efficiency. *J. Comput. Graph. Stat.* **2011**, *20*, 531–570.
3. Filippone, M.; Girolami, M. Pseudo-marginal Bayesian inference for Gaussian Processes. *IEEE Trans. Pattern. Anal. Mach. Intell.* **2014**, doi:TPAMI.2014.2316530.
4. Attias, H. Inferring parameters and structure of latent variable models by variational Bayes. In Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence (UAI'99), Stockholm, Sweden, 30 July–1 August 1999; Morgan Kaufmann Publishers Inc.: San Francisco, CA, USA, 1999; pp. 21–30.
5. Minka, T. P. Expectation propagation for approximate Bayesian inference. In Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence, Seattle, WA, USA, 2–5 August 2001; Morgan Kaufmann Publishers Inc.: San Francisco, CA, USA, 2001; pp. 362–369.
6. Rue, H.; Martino, S.; Chopin, N. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *J. R. Stat. Soc. Ser. B* **2009**, *71*, 319–392.
7. Furrer, R.; Sain, S.R. Spatial model fitting for large datasets with applications to climate and microarray problems. *Stat. Comput.* **2008**, *19*, 113–128.
8. Kim, Y.; Kim, D.H. An approximate likelihood function of spatial correlation parameters. **2015**, *submitted*.
9. Cressie, N. *Statistics for Spatial Data*, Revised ed.; Wiley: New York, NY, USA, 1993.
10. Dickey, J.M. Approximate posterior distributions. *J. Am. Stat. Assoc.* **1976**, *71*, 680–689.
11. Kullback, S.; Leibler, R.A. On information and sufficiency. *Ann. Math. Stat.* **1951**, *22*, 79–86.
12. Shannon, C.E. A Mathematical Theory of Communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423, 623–656,
13. Schneier, B *Applied Cryptography*, 2nd ed.; John Wiley and Sons: New York, NY, USA, 1996.