Published in final edited form as:

J Comput Graph Stat. 2019 ; 28(2): 334–349. doi:10.1080/10618600.2018.1529600.

Partition Weighted Approach for Estimating the Marginal Posterior Density with Applications

Yu-Bo Wang^{*}, Ming-Hui Chen[†], Lynn Kuo[†], and Paul O. Lewis[‡]

^{*}Department of Mathematical Sciences, Clemson University

[†]Department of Statistics, University of Connecticut

[‡]Department of Ecology and Evolutionary Biology, University of Connecticut

Abstract

The computation of marginal posterior density in Bayesian analysis is essential in that it can provide complete information about parameters of interest. Furthermore, the marginal posterior density can be used for computing Bayes factors, posterior model probabilities, and diagnostic measures. The conditional marginal density estimator (CMDE) is theoretically the best for marginal density estimation but requires the closed-form expression of the conditional posterior density, which is often not available in many applications. We develop the partition weighted marginal density estimator (PWMDE) to realize the CMDE. This unbiased estimator requires only a single MCMC output from the joint posterior distribution and the known unnormalized posterior density. The theoretical properties and various applications of the We carry out simulation studies to investigate the empirical performance of the PWMDE and further demonstrate the desirable features of the proposed method with two real data sets from a study of dissociative identity disorder patients and a prostate cancer study, respectively.

Keywords

Bayesian model selection; Conditional marginal density estimator; Ordinal probit regression; Partition weighted kernel estimator; Savage-Dickey density ratio

1 Introduction

Posterior density estimation is one of the most important topics in Bayesian inference because it provides complete information about parameters of interest in the model. Chen (2005a) provides an overview of the usefulness of posterior density estimation in computing Bayes factors, marginal likelihoods, and posterior model probabilities. Posterior density estimation has also been used in various applications, including selection of the best

Supplementary Material: “PWMDESupplementary.pdf” contains the proof of Theorem 3.1 and Table S1: The sensitivity analysis of the conditional working parameter space. (.pdf file)

R codes for Sections 4 and 5: “Section4.zip” and “Section5.zip” contain R codes for the simulation studies and the application to a constrained analysis of variance in dissociative identity disorder patients in the article, respectively. (.r file) The R code used in this article is also available at <https://github.com/ybwstats/Partition-Weighted-Marginal-Density-Estimator>.

DID data set: Data set used in Section 5. “Section5.zip” contains this data set. (.txt file)

predictors for the development of AIDS using historical data (Chen et al., 1999), the development of computational algorithms in molecular population genetics (Stephens and Donnelly, 2000), estimation of the functional Bregman divergence for Bayesian model diagnostics (Goh and Dey, 2014), and the intensity bias correction in endorectal multi-parametric MRI (Lui, 2014; Lui et al., 2015).

When all parameters in the model are of concern, the joint posterior density is investigated using the unnormalized posterior density, which is the product of the likelihood function and a joint prior distribution. The normalizing constant, which is the marginal likelihood when the prior is proper, is often analytically intractable due to complicated model structure. In this case, the computational problem reduces to estimation of the marginal likelihood. For estimating the marginal likelihood, many Monte Carlo methods have been developed, including Meng and Wong (1996), Chib (1995), Chib and Jeliazkov (2001), Friel and Pettitt (2008), Xie et al. (2011), and Fan et al. (2011). More recently Wang et al. (2018) proposed the partition weighted kernel (PWK) estimation, which requires only a single Markov chain Monte Carlo (MCMC) sample from the posterior distribution and the known unnormalized posterior density, both available in the standard output from existing Bayesian software. By assigning a weight to each point of an MCMC sample within a working parameter space where the unnormalized posterior density is bounded away from zero, the PWK estimator is consistent for the reciprocal of the normalizing constant and of finite variance, and its minimum variance estimator can be achieved.

In many practical problems, investigators are only interested in specific rather than all parameters. As a result, the calculation and display of the marginal posterior density of the focal parameters are of most interest. Similar to the joint posterior density, the marginal posterior density is not analytically available for most cases. Hence, several methods have been developed for estimating marginal posterior densities using an MCMC sample from the joint posterior distribution. One common approach is the kernel density estimator (KDE) proposed by Rosenblatt (1956) and Parzen (1962). KDE is easily implemented, but leaves room for more efficient estimators that can make use of more of the available information. Another approach is the conditional marginal density estimator (CMDE) method proposed by Gelfand et al. (1992). This method simply takes the average of the known conditional posterior density of the interested parameters given the other parameters from the MCMC sample. The CMDE is most efficient because it is a Rao-Blackwell estimator. Unfortunately, CMDE requires a closed-form conditional posterior density, which is often known only up to a unknown normalizing constant for many Bayesian problems, especially when the parameter space is constrained. To overcome this difficulty, Chen (1994) proposed the importance weighted marginal density estimation (IWMDE) method, which can be viewed as a generalization of the CMDE and only requires a careful selection of the conditional weight density. Some other approaches were described in Chen (2005a), including two block structures of the IWMDE by Oh (1999) and the Gibbs stopper (GS) estimator by Yu and Tanner (1999); however, when the conditional posterior density is analytically intractable, the realization of the CMDE still remains an open research problem.

Motivated by the PWK estimator by Wang et al. (2018) for estimating marginal posterior density, we propose a new Monte Carlo method, the Partition Weighted marginal density

estimator (PWMDE), which assumes only the availability of the unnormalized posterior density and an MCMC sample from the joint posterior distribution. The PWMDE method is constructed by first partitioning a subset of the support of the conditional posterior distribution, and then assigning a weighted average to ratios of two unnormalized posterior density evaluated at the MCMC sample, except the focused parameters in the numerator are set as fixed. Here, the weights are assigned locally using a representative value of the unnormalized posterior density in each partitioned subset. Both the partition and the weights change at each MCMC sample point. We show that the PWMDE is unbiased, and most of all, its optimal result is realizable and approximates the CMDE, which is known to be the best solution but whose widespread use is limited by unavailability of the conditional posterior density analytically. In addition, we extend the PWMDE method to estimate the conditional posterior density when an MCMC sample from the conditional posterior density is available. We have also carried out simulation studies with multivariate t distributions and show that PWMDE has a better empirical performance than the variation of IWMDE discussed in Chen (1994) in estimating the marginal t distributions. In our first real data example, where closed-form conditional posterior densities are available, we show that our estimator produces results similar to the gold standard (CMDE). In our second real data example, we demonstrate an excellent performance of the PWMDE method in estimating Bayes factors. The proposed PWMDE method has the potential to become a powerful tool for computing posterior densities, Bayes factors, and marginal likelihoods for complex Bayesian models.

Wang et al. (2018) deals with marginal likelihood estimation, whereas this paper deals with posterior density and conditional posterior density estimation. Although both papers assume we have a sample from a single chain MCMC method with associated kernel values, the density estimation is not straightforward. Most of the methods we know require knowing the conditional density form which is usually not available in practice. Our method does not make this assumption. Moreover, we show our locally weighted piecewise constant density estimator is almost as good as the theoretical optimal estimator CMDE proposed by Gelfand et al. (1992) that requires knowing the form of the conditional posterior density. Our PWMDE is also related to the importance weighted marginal density estimator (IWMDE) proposed by Chen (1994) that does not require knowing the conditional posterior density and no partition of the focal parameter space is needed. We show by simulation that our method is more accurate than IWMDE, but at a cost of more computational burden.

The rest of the article is organized as follows: Section 2 is a review of existing methods and a summary of preliminaries needed for our method. In Section 3, we develop the PWMDE method for estimating marginal and conditional posterior densities, and examine its theoretical properties and various applications. Simulation studies are conducted in Section 4 to examine the empirical performance of PWMDE and the other variation of IWMDE. In Section 5, we compare the performances of KDE and PWMDE with the CMDE in the inequality-constrained analysis of variance model. We use the data set from an amnesia study of dissociative identity disorder patients to demonstrate the empirical performance of the PWMDE. In Section 6, we present a novel application of the PWMDE method for computing the marginal posterior densities in Bayesian model selection. A complete Bayesian analysis is carried out under an ordinal probit regression model with latent

variables using the real data from a prostate cancer study. To avoid the phenomenon of the Bartlett's or Lindley's paradox (Jeffreys, 1998; Lindley, 1957), we first apply the PWK estimator to find the best prior distribution in the full model by the empirical Bayes method and we then apply PWMDE to compute the Bayes factors under the best full model using the same MCMC sample used by the PWK estimator for model selection. Finally, we conclude the paper with a brief discussion in Section 7.

2 Preliminaries

We first introduce some notation. Suppose $\boldsymbol{\zeta} = (\boldsymbol{\theta}, \boldsymbol{\xi})$ is a v -dimensional vector of parameters, where $\boldsymbol{\theta}$ is a vector of parameters of interest having length p . Let D denote the data, $L(\boldsymbol{\zeta}|D)$ denote the likelihood function, and $\pi(\boldsymbol{\zeta})$ denote the prior distribution for $\boldsymbol{\zeta}$. Then, the joint posterior density is

$$\pi(\boldsymbol{\zeta}|D) = \frac{L(\boldsymbol{\zeta}|D)\pi(\boldsymbol{\zeta})}{c} = \frac{q(\boldsymbol{\zeta})}{c}, \quad (1)$$

where c is the normalizing constant, and $q(\boldsymbol{\zeta})$ is the unnormalized posterior density. Equation (1) shows that the joint posterior density is known up to a normalizing constant. Hence, when $p = v$, the computation problem is the estimation of c . For this value, the PWK estimator by Wang et al. (2018) can be used. Assume Ω is the support of $\pi(\boldsymbol{\zeta}|D)$ with $\Omega' \subset \Omega$ being the working parameter space, and $\{A_1, A_2, \dots, A_K\}$ forms a partition of Ω' . If we have an MCMC sample $\{\boldsymbol{\zeta}_t = (\boldsymbol{\theta}_t, \boldsymbol{\xi}_t), t = 1, 2, \dots, T\}$ from the joint posterior distribution $\pi(\boldsymbol{\zeta}|D)$, the PWK estimator for $1/c$ is

$$\widehat{c^{-1}} = \frac{\frac{1}{T} \sum_{t=1}^T \sum_{k=1}^K \frac{q(\boldsymbol{\zeta}_k^*)}{q(\boldsymbol{\zeta}_t)} 1\{\boldsymbol{\zeta}_t \in A_k\}}{\sum_{k=1}^K q(\boldsymbol{\zeta}_k^*) V(A_k)}, \quad (2)$$

Where $\boldsymbol{\zeta}_k^*$ is a representative point in region A_k , $1\{\boldsymbol{\zeta}_t \in A_k\}$ is the indicator function, and $V(A_k) = \int_{\Omega'} 1\{\boldsymbol{\zeta} \in A_k\} d\boldsymbol{\zeta}$. As discussed in Wang et al. (2018), $q(\boldsymbol{\zeta}_t)$ in (2) can be made close to $q(\boldsymbol{\zeta}_k^*)$ by increasing the number of partition subsets to improve estimation precision.

When $p < v$, the marginal posterior density of interest is defined by

$$\pi(\boldsymbol{\theta}_0|D) = \int_{\Omega_{\boldsymbol{\theta}_0}} \pi(\boldsymbol{\zeta}|D)|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} d\boldsymbol{\xi} = \int_{\Omega_{\boldsymbol{\theta}_0}} \pi(\boldsymbol{\theta}_0, \boldsymbol{\xi}|D) d\boldsymbol{\xi}, \quad (3)$$

where $\Omega_{\boldsymbol{\theta}_0} = \{\boldsymbol{\xi}: (\boldsymbol{\theta}_0, \boldsymbol{\xi}) \in \Omega\}$. To estimate the marginal posterior density, a nonparametric kernel density estimator (KDE) can be used. It is similar to the KDE of the frequentist literature except the sample is replaced by the MCMC sample of interested parameters. Although it is easily implemented and requires no further assumptions, it may be less

efficient because it does not use the information from the MCMC sample of non-focal parameters and the known structure of the posterior distribution.

Another common approach is the conditional marginal density estimator (CMDE) proposed by Gelfand et al. (1992). Assuming that the analytical form of the conditional posterior density $\pi(\boldsymbol{\theta}|\boldsymbol{\xi}, D)$ is available, (3) can be re-written as

$$\pi(\boldsymbol{\theta}_0|D) = \int_{\Omega} \pi(\boldsymbol{\theta}_0, \boldsymbol{\xi}|D) \pi(\boldsymbol{\xi}|D) d\boldsymbol{\xi} = \int_{\Omega} \pi(\boldsymbol{\theta}_0|\boldsymbol{\xi}, D) \pi(\boldsymbol{\xi}|D) d\boldsymbol{\xi}. \quad (4)$$

Then

$$\hat{\pi}_{\text{CMDE}}(\boldsymbol{\theta}_0|D) = \frac{1}{T} \sum_{t=1}^T \pi(\boldsymbol{\theta}_0|\boldsymbol{\xi}_t, D). \quad (5)$$

It can be shown that under some mild regularity conditions, $\hat{\pi}_{\text{CMDE}}(\boldsymbol{\theta}_0|D)$ is an unbiased and consistent estimator of the marginal posterior density, that is,

$$E(\hat{\pi}_{\text{CMDE}}(\boldsymbol{\theta}_0|D)) = \pi(\boldsymbol{\theta}_0|D),$$

and

$$\lim_{T \rightarrow \infty} \hat{\pi}_{\text{CMDE}}(\boldsymbol{\theta}_0|D) = \pi(\boldsymbol{\theta}_0|D) \text{ a.s.}$$

In addition, the use of the conditional structure of the posterior density makes the CMDE a Rao-Blackwell estimator so that it is optimal for this estimation problem. However, the closed form of $\pi(\boldsymbol{\theta}|\boldsymbol{\xi}, D)$ is often not available in many Bayesian problems. To overcome this difficulty, Chen (1994) proposed the importance weighted marginal density estimation (IWMDE) method, which can be considered as a generalization of CMDE. Consider the following identity:

$$\begin{aligned} \pi(\boldsymbol{\theta}_0|D) &= \int \pi(\boldsymbol{\theta}_0, \boldsymbol{\xi}|D) d\boldsymbol{\xi} = \int \pi(\boldsymbol{\theta}_0, \boldsymbol{\xi}|D) \int w(\boldsymbol{\theta}|\boldsymbol{\xi}) d\boldsymbol{\theta} d\boldsymbol{\xi} \\ &= \iint w(\boldsymbol{\theta}|\boldsymbol{\xi}) \pi(\boldsymbol{\theta}_0, \boldsymbol{\xi}|D) d\boldsymbol{\theta} d\boldsymbol{\xi} = \int \frac{w(\boldsymbol{\theta}_0|\boldsymbol{\xi}) \pi(\boldsymbol{\theta}_0, \boldsymbol{\xi}|D)}{\pi(\boldsymbol{\xi}|D)} \pi(\boldsymbol{\xi}|D) d\boldsymbol{\xi}, \end{aligned} \quad (6)$$

where $w(\boldsymbol{\theta}|\boldsymbol{\xi})$ is a proposed conditional density whose support is contained in the support of $\pi(\boldsymbol{\theta}|\boldsymbol{\xi}, D)$. Using the identity in (6), the IWMDE of $\pi(\boldsymbol{\theta}_0|D)$ is given by

$$\hat{\pi}_{\text{IWMDE}}(\theta_0|D) = \frac{1}{T} \sum_{t=1}^T w(\theta_t|\xi_t) \frac{\pi(\theta_0, \xi_t|D)}{\pi(\theta_t, \xi_t|D)}. \quad (7)$$

The IWMDE method is attractive because it does not require the conditional posterior density to be known, the only requirement is that one needs to choose a good weight function $w(\theta|\xi)$. Under mild regularity conditions, the IWMDE also has the properties of unbiasedness and consistency to the marginal posterior density.

3 The Proposed Method for Estimating the Posterior Densities

3.1 Estimating the Marginal Posterior Density

Chen (1994) showed that the optimal weight function minimizing the variance of (7) is $\pi(\theta|\xi, D)$, and in this case the IWMDE in (7) reduces to CMDE in (5). However, this optimal weight function is unavailable in most cases, and proposing a similar one is nontrivial. To circumvent the difficulties, we exploit the main idea behind the PWK (Wang et al., 2018) estimator to obtain an approximate distribution of the conditional posterior density so that the CMDE can be realized for marginal posterior density estimation.

Let $\Theta_\xi = \{\theta: q(\theta, \xi) > 0\}$ denote the support of the conditional posterior distribution $\pi(\theta|\xi, D)$ and $\tilde{\Theta}_\xi$ be any subset of Θ_ξ such that $\int_{\tilde{\Theta}_\xi} q(\theta, \xi) d\theta > 0$. We call $\tilde{\Theta}_\xi$ the conditional working parameter space. We also let $\{A_k(\xi), k = 1, 2, \dots, K\}$ be the partition of $\tilde{\Theta}_\xi$. We consider a weight function $w(\theta|\xi)$ which satisfies the following two conditions: (i) $w(\theta|\xi) \geq 0$ and (ii) $\int_{\tilde{\Theta}_\xi} w(\theta|\xi) d\theta = 1$. Then, we propose a new estimator of $\pi(\theta_0|D)$ using the idea behind PWK:

$$\hat{\pi}(\theta_0|D) = \frac{1}{T} \sum_{t=1}^T \sum_{k=1}^K w(\theta_t|\xi_t) \frac{q(\theta_0, \xi_t)}{q(\theta_t, \xi_t)} 1\{\theta_t \in A_k(\xi_t)\}. \quad (8)$$

Write

$$\hat{\pi}_t(\theta_0|D) = \sum_{k=1}^K w(\theta_t|\xi_t) \frac{q(\theta_0, \xi_t)}{q(\theta_t, \xi_t)} 1\{\theta_t \in A_k(\xi_t)\}$$

such that $\hat{\pi}(\theta_0|D) = \frac{1}{T} \sum_{t=1}^T \hat{\pi}_t(\theta_0|D)$. Then, we have

$$\begin{aligned}
 E[\hat{\pi}_t(\theta_0|D)] &= \iint \sum_{k=1}^K w(\theta|\xi) \frac{q(\theta_0, \xi)}{q(\theta, \xi)} 1\{\theta \in A_k(\xi)\} \frac{q(\theta, \xi)}{c} d\theta d\xi = \\
 &= \int \frac{q(\theta_0, \xi)}{c} \int w(\theta|\xi) \sum_{k=1}^K 1\{\theta \in A_k(\xi)\} d\theta d\xi = \int \frac{q(\theta_0, \xi)}{c} \int_{\tilde{\Theta}_\xi} w(\theta|\xi) d\theta d\xi = \pi(\theta_0|D),
 \end{aligned} \quad (9)$$

which ensures that $\hat{\pi}(\theta_0|D)$ in (8) is an unbiased estimator of $\pi(\theta_0|D)$. After some algebra, we obtain

$$\text{Var}_w\{\hat{\pi}_t(\theta_0|D)\} = \int \left[q(\theta_0, \xi) \pi(\theta_0, \xi|D) \int_{\tilde{\Theta}_\xi} \frac{w^2(\theta|\xi)}{q(\theta, \xi)} d\theta \right] d\xi - \pi^2(\theta_0|D). \quad (10)$$

Now, we establish the following useful result.

Theorem 3.1 *Let*

$$w_{opt}(\theta|\xi) = \frac{q(\theta, \xi)}{\int_{\tilde{\Theta}_\xi} q(\theta, \xi) d\theta}, \quad (11)$$

which is the conditional posterior density defined on $\tilde{\Theta}_\xi$. Then, we have

$$\text{Var}_{w_{opt}}\{\hat{\pi}_t(\theta_0|D)\} = \int \left[\frac{q(\theta_0, \xi) \pi(\theta_0, \xi|D)}{\int_{\tilde{\Theta}_\xi} q(\theta, \xi) d\theta} \right] d\xi - \pi^2(\theta_0|D) \quad (12)$$

and

$$\text{Var}_{w_{opt}}\{\hat{\pi}_t(\theta_0|D)\} \leq \text{Var}_w\{\hat{\pi}_t(\theta_0|D)\} \quad (13)$$

for any conditional density $w(\cdot)$ defined on $\tilde{\Theta}_\xi$.

The result established in Theorem 3.1 is an extension of Theorem 2.1 in Chen (1994). The proof of this theorem is also similar and provided in the Supplementary Materials.

Remark 3.1: Note that $w_{opt}(\theta|\xi)$ in (11) is not the conditional posterior density $\pi(\theta|\xi, D)$ unless $\tilde{\Theta}_\xi = \Theta_\xi$. Consequently, $\hat{\pi}_{w_{opt}}(\theta_0|D)$, that is equation (8) with plug-in weights of (11), is not the CMDE. From Theorem 3.1, we see that $\text{Var}_{w_{opt}}\{\hat{\pi}_t(\theta_0|D)\}$ decreases when the

conditional working parameter space $\bar{\Theta}_\xi$ gets larger. Once $\bar{\Theta}_\xi = \Theta_\xi$, $\text{Var}\left\{\hat{\pi}_{w_{\text{opt}}}(\theta_0|D)\right\}$ is equal to $[\int \pi(\theta_0|\xi, D)\pi(\theta_0, \xi|D)d\xi - \pi^2(\theta_0|D)]/T$ and exactly the variance of the CMDE. Thus, the CMDE is the best among all of the estimators given in (8). As discussed in Chen (1994), $\pi(\theta|\xi, D)$ is analytically intractable for most applications and thus it is almost always impossible to compute the CMDE in practice. However, the result established in Theorem 3.1 sheds light on how to obtain an estimator of the marginal posterior density that approximates the performance of CMDE.

Although $w_{\text{opt}}(\theta|\xi)$ is analytically intractable, we can borrow the idea behind the PWK estimator to obtain an approximate optimal estimator $\hat{\pi}(\theta_0|D)$ using discretization. First, we take $\bar{\Theta}_\xi$ such that $\int_{\bar{\Theta}_\xi} d\theta < \infty$. Second, we specify $w(\cdot)$ as

$$w_{\text{PWMDE}}(\theta|\xi) = \frac{q(\theta_k^*, \xi)}{\sum_{k=1}^K q(\theta_k^*, \xi)V(A_k(\xi))}, \quad \theta \in A_k(\xi), \quad (14)$$

where θ_k^* is a fixed point in $A_k(\xi)$ and $V(A_k(\xi))$ is the volume of $A_k(\xi)$, that is $\int_{\bar{\Theta}_\xi} 1\{\theta \in A_k(\xi)\}d\theta$. Then, plugging w_{PWMDE} in (8) gives

$$\hat{\pi}_{\text{PWMDE}}(\theta_0|D) = \frac{1}{T} \sum_{t=1}^T \sum_{k=1}^K \frac{q(\theta_k^*, \xi_t)1\{\theta_t \in A_k(\xi_t)\}}{\sum_{k=1}^K q(\theta_k^*, \xi_t)V(A_k(\xi_t))} \frac{q(\theta_0, \xi_t)}{q(\theta_t, \xi_t)}. \quad (15)$$

We see from (15) that the construction of the weight $\frac{q(\theta_k^*, \xi_t)1\{\theta_t \in A_k(\xi_t)\}}{\sum_{k=1}^K q(\theta_k^*, \xi_t)V(A_k(\xi_t))}$ is based on the partition $\{A_k(\xi_t), k = 1, 2, \dots, K\}$. Therefore, we call this the partition weighted marginal density estimator (PWMDE).

Remark 3.2: Under certain ergodic conditions (time-reversible, invariant, and irreducible), we can show that $\hat{\pi}_{\text{PWMDE}}(\theta_0|D)$ is an unbiased and consistent estimator of $\pi(\theta_0|D)$. Notice that $\sum_{k=1}^K q(\theta_k^*, \xi)V(A_k(\xi))$ in (14) is a Riemann approximation of $\int_{\bar{\Theta}_\xi} q(\theta, \xi)d\theta$. Thus, when $q(\theta, \xi)$ is continuous, $\bar{\Theta}_\xi = \Theta_\xi$, $\int_{\bar{\Theta}_\xi} q(\theta, \xi)d\theta < \infty$, $K \rightarrow \infty$ and $A_k(\xi)$ converges to a single point, $\sum_{k=1}^K q(\theta_k^*, \xi)V(A_k(\xi)) \rightarrow \int_{\bar{\Theta}_\xi} q(\theta, \xi)d\theta$, implying that $w_{\text{PWMDE}}(\theta|\xi) \rightarrow w_{\text{opt}}(\theta|\xi)$. Thus, when K is large and $A_k(\xi)$ is small, $\hat{\pi}_{\text{PWMDE}}(\theta_0|D)$ is approximately equal to $\hat{\pi}_{\text{CMDE}}(\theta_0|D)$. Thus, the proposed PWMDE has the potential to be as good as the best solution (CMDE).

Remark 3.3: If Θ_{ξ} is bounded, we suggest $\tilde{\Theta}_{\xi} = \Theta_{\xi}$. If Θ_{ξ} is unbounded, then we apply the elliptical ring approach. Assuming that the dimension of θ is p , the elliptical rings are defined as

$$A_k(\xi) = A_k = \left\{ \theta : r(k-1)/K \leq \left\| \bar{\Sigma}^{-\frac{1}{2}}(\theta - \bar{\theta}) \right\| \leq rk/K \right\}, k = 1, 2, \dots, K,$$

where $\bar{\theta}$ and $\bar{\Sigma}$ are the posterior mean and the posterior covariance matrix of θ , respectively, and r is the radius chosen for the conditional working parameter space. A guide value of r is $\sqrt{\chi_{p,0.95}^2}$ suggested in Wang et al. (2018), which is the square root of the 95th percentile of the chi-square distribution with p degrees of freedom, so that around 95% of an MCMC sample is included in the conditional working parameter space. The volume of $A_k(\xi)$ can be calculated as follows

$$V(A_k) = V_p |\bar{\Sigma}|^{1/2} \left[\left(\frac{rk}{K} \right)^p - \left(\frac{r(k-1)}{K} \right)^p \right], k = 1, 2, \dots, K,$$

where $|\cdot|$ is the determinant and $V_p = \pi^{p/2} \Gamma(p/2+1)$ is the volume of a p -dimensional unit hypersphere. For θ_k^* , we use the midpoint in each A_k along the first eigenvector of $\bar{\Sigma}$.

Remark 3.4: We have discovered the PWMDE continues to work well even with a much smaller working parameter space. In Table S1, we have chosen the 50th percentile, i.e. $\sqrt{\chi_{p,0.50}^2}$ for $\tilde{\Theta}_{\xi}$. The results show a small increases in MCSE and RMSE when compared to Table 1 where 95th percentile was chosen. Hence, it indicates that the performance of the PWMDE is not very sensitive to the choice of $\tilde{\Theta}_{\xi}$.

3.2 Estimating the Conditional Posterior Density

Suppose $\theta = (\theta_1, \theta_2)$. Let $\pi(\theta_2/\theta_1, D)$ denote the conditional posterior density of θ_2 given $\theta_1 = \theta_1$. Using the same notation as in Section 3.1, the PWMDE of $\pi(\theta_2/\theta_1, D)$ is given by

$$\hat{\pi}_{\text{PWMDE}}(\theta_{02}|\theta_{01}, D) = \frac{1}{T} \sum_{t=1}^T \sum_{k=1}^K \frac{q(\theta_{01}, \theta_{2k}^*, \xi_t) 1\{\theta_{2t} \in A_k(\theta_{01}, \xi_t)\}}{\sum_{k=1}^K q(\theta_{01}, \theta_{2k}^*, \xi_t) V(A_k(\theta_{01}, \xi_t))} \frac{q(\theta_{01}, \theta_{02}, \xi_t)}{q(\theta_{01}, \theta_{2t}, \xi_t)}, \quad (16)$$

where $\{(\theta_{2t}, \xi_t), t = 1, 2, \dots, T\}$ is an MCMC sample from $\pi(\theta_2, \xi|\theta_{01}, D)$, θ_{2k}^* is a representative point in $A_k(\theta_{01}, \xi)$, and $V(A_k(\theta_{01}, \xi))$ is the volume of $A_k(\theta_{01}, \xi)$.

Under mild regularity conditions, we can show that $\hat{\pi}_{\text{PWMDE}}(\theta_{02}|\theta_{01}, D)$ is an unbiased estimator of $\pi(\theta_{02}|\theta_{01}, D)$. This result can be established as follows. Write $c(\theta_{01}) = \int q(\theta_{01}, \theta_2, \xi) d\theta_2 d\xi$. Then, we have $\pi(\theta_2, \xi|\theta_{01}, D) = q(\theta_{01}, \theta_2, \xi)/c(\theta_{01})$, $\pi(\theta_2|\theta_{01}, D) = \int \pi(\theta_2, \xi|\theta_{01}, D) d\xi$, and

$$\begin{aligned}
& E[\hat{\pi}_{\text{PWMDE}}(\theta_{02}|\theta_{01}, D)] \\
&= \iint \sum_{k=1}^K \frac{q(\theta_{01}, \theta_{2k}^*, \xi) 1\{\theta_2 \in A_k(\theta_{01}, \xi)\}}{\sum_{k=1}^K q(\theta_{01}, \theta_{2k}^*, \xi) V(A_k(\theta_{01}, \xi))} \frac{q(\theta_{01}, \theta_{02}, \xi)}{q(\theta_{01}, \theta_2, \xi)} \pi(\theta_2, \xi | \theta_{01}, D) d\theta_2 d\xi \\
&= \iint \sum_{k=1}^K \frac{q(\theta_{01}, \theta_{2k}^*, \xi) 1\{\theta_2 \in A_k(\theta_{01}, \xi)\}}{\sum_{k=1}^K q(\theta_{01}, \theta_{2k}^*, \xi) V(A_k(\theta_{01}, \xi))} \frac{q(\theta_{01}, \theta_{02}, \xi)}{c(\theta_{01})} d\theta_2 d\xi \\
&= \int \frac{q(\theta_{01}, \theta_{02}, \xi)}{c(\theta_{01})} d\xi = \pi(\theta_{02} | \theta_{01}, D).
\end{aligned} \tag{17}$$

There are two major applications of (16). First, when the dimension of θ_0 is high, $\hat{\pi}_{\text{PWMDE}}(\theta_0 | D)$ in (15) may not be efficient. The dimension reduction can be achieved via the identity $\pi(\theta_0 | D) = \pi(\theta_{01} | D) \pi(\theta_{02} | \theta_{01}, D)$. Now, instead of estimating $\pi(\theta_0 | D)$ directly, we use (15) and (16) to estimate $\pi(\theta_{01} | D)$ and $\pi(\theta_{02} | \theta_{01}, D)$, respectively. These two estimators require one MCMC sample from the full posterior distribution $\pi(\zeta | D)$ and another MCMC sample from the conditional posterior distribution $\pi(\theta_2, \xi | \theta_{01}, D)$. We note that once the sampling code for the full posterior distribution is readily available, the very same code with minimal changes can be used to generate an MCMC sample from the conditional posterior distribution.

Second, (16) can also be applied to the estimation of the marginal likelihood through Chib's identity (Chib, 1995). Suppose we group ζ into G blocks as $\zeta = (\zeta'_1, \zeta'_2, \dots, \zeta'_G)'$. Following Chib (1995), we have

$$c = \frac{q(\zeta^*)}{\pi(\zeta_1^*, \zeta_2^*, \dots, \zeta_G^* | D)} = \frac{q(\zeta^*)}{\pi(\zeta_1^* | D) \pi(\zeta_2^* | \zeta_1^*, D) \dots \pi(\zeta_G^* | \zeta_1^*, \zeta_2^*, \dots, \zeta_{G-1}^*, D)}, \tag{18}$$

where $\zeta^* = ((\zeta_1^*)', (\zeta_2^*)', \dots, (\zeta_G^*)')'$ is a high posterior density point such as the posterior mean or mode of ζ . In (18), $\pi(\zeta_1^* | D)$ can be estimated by (15) using an MCMC sample from the full posterior density while an estimate of each conditional density $\pi(\zeta_g^* | \zeta_1^*, \zeta_2^*, \dots, \zeta_{g-1}^*, D)$ can be obtained by (16) using an MCMC sample from the conditional posterior distribution $\pi(\zeta_g, \zeta_{g+1}, \dots, \zeta_G | \zeta_1^*, \zeta_2^*, \dots, \zeta_{g-1}^*, D)$ for $g = 2, 3, \dots, G$. This approach does not require knowing the full conditional posterior density analytically (i.e., CMDE) or the closed-form expression of the full conditional density after introducing additional latent variables (namely, the augmented CMDE) as discussed in Chib (1995). Furthermore, compared to the method developed in Chib and Jeliazkov (2001), PWMDE does not require knowing the transition kernel or the proposal density used in the MCMC sampling algorithm.

4 Simulation Studies

We consider an example of a multivariate t distribution and compare the empirical performance of PWMDE with the other variation of IWMDE in estimating the marginal t density. Assume that the joint distribution of ζ is

$$\zeta \sim t_{df=3}(\mu, \Sigma), \quad (19)$$

where $\zeta = (\theta', \xi')'$, $\theta = (\theta_1, \theta_2, \theta_3)'$, $\xi = (\xi_1, \xi_2)'$, df is the degrees of freedom, $\mu = (0, 0, 0, 0, 0)'$ is a location vector, and Σ is a 5×5 positive definite scale matrix. We further assume that the element of the i^{th} row and the j^{th} column in Σ is equal to $3 \times 0.5^{|i-j|}$ for $i, j = 1, 2, \dots, 5$. We wish to estimate the marginal density of θ , which is known as $t_{df=3}((0, 0, 0)', \Sigma_\theta)$, where

$$\Sigma_\theta = \begin{pmatrix} 3 & 1.5 & 0.75 \\ 1.5 & 3 & 1.5 \\ 0.75 & 1.5 & 3 \end{pmatrix}.$$

We note that PWMDE is a special case of IWMDE. In this simulation study, we aim to compare PWMDE to another variation of IWMDE in which the weight function $w(\theta|\xi)$ in (7) is specified as the conditional multivariate normal density, which was one of the choices for $w(\theta|\xi)$ suggested by Chen (1994). Here, the conditional multivariate normal density is constructed based on the MC sample.

We compute the simulation estimate (Mean), the Monte Carlo standard error (MCSE), and the root mean square error (RMSE) of the marginal density estimates in log scale. We generate $M = 1,000$ independent Monte Carlo (MC) samples of size $T = 10,000$ from (19). Let $\pi_B(\theta_0)$ denote the density of $t_{df=3}((0, 0, 0)', \Sigma_\theta)$ evaluated at $\theta = \theta_0$. Let $\hat{d}_m$ be an estimate of $\pi_B(\theta_0)$ based on the m^{th} MC sample for $m = 1, 2, \dots, M$. Then, Mean, MCSE, and RMSE are computed as $\widehat{\log \pi_{t_3}}(\theta_0) = \frac{1}{M} \sum_{m=1}^M \log \hat{d}_m$, $\left\{ \frac{1}{M-1} \sum_{m=1}^M (\log \hat{d}_m - \widehat{\log \pi_{t_3}}(\theta_0))^2 \right\}^{1/2}$, and $\left\{ \frac{1}{M} \sum_{m=1}^M (\log \hat{d}_m - \log \pi_{t_3}(\theta_0))^2 \right\}^{1/2}$, respectively. For PWMDE, we choose $r = \sqrt{\chi_{3,0.95}^2}$ and $K=5, 10, 15$, and 20 , to construct a ring partition. In addition to $\theta_0 = (0, 0, 0)'$, which is the center of the marginal posterior density, we also consider $\theta_0 = (1, 0, 0)'$, $\theta_0 = (0, 1, 0)'$, and $\theta_0 = (0, 0, 1)'$. Table 1 reports these simulation results. We see MCSE and RMSE decrease as K increases. They are smaller than that of the IWMDE even with $K=5$.

To examine how our method can be affected by the conditional working parameter space, we have carried out a similar evaluation except we have chosen $r = \sqrt{\chi_{p,0.50}^2}$, a much smaller working parameter space. We report the results in Table S1 in the Supplementary Materials. They show the MCSE and RMSE only increase slightly when compared to that in Table 1. This study supports the comments made in REMARK 3.4.

To evaluate the performance of PWMDE in high-dimensional problems, we consider two scenarios: (a) $p = 3$ but $v - p$ increases to 10, 20, and 50 (Table 2); (b) under $v - p = 10, 20$, and 50, p increases to 10 (Table 3). Similar to Table 1, we report mean, MCSE, and RMSE, where $\theta_0 = (0, 0, 0)'$ for Table 2, and θ_0 is a 10 dimensional vector of zeros in Table 3. We have also added more choices of K , and computing time in this table. We see PWMDE continues to have smaller MCSE and RMSE than that of IWMDE at a cost of more computing time.

Remark 4.1: For higher dimension, we would recommend using our idea developed in Section 3.2. That is, we evaluate a high-dimensional posterior density using several lower dimension conditional distributions as in the denominator of (18). Although it requires multiple MCMC samples, the computational cost should be manageable for relatively large dimension given that our interest is to evaluate the marginal posterior density only at a few values of θ_0 .

Remark 4.2: Table 3 also illustrates how to select K : investigate the Mean and MCSE for a sequence of increasing K values and stop at the first K when the MCSE becomes stabilized and $\text{MCSE}/|\text{Mean}|$ is less than 5%. So from Table 3, we would choose $K = 25$ for each case, where $v - p = 10, 20$, and 50, respectively.

5 Inequality-Constrained Analysis of Variance

In this section, we apply PWMDE to the inequality-constrained analysis of variance model to find the marginal posterior density. We use the data in Hoijtink et al. (2008) collected to study amnesia in patients with dissociative identity disorder (DID). The experiment was designed to compare memory performance scores among four groups, including DID, mimic normal, symptom simulated, and true amnesic subjects. Chen and Kim (2008) were interested in testing several hypotheses for the mean scores. Based on their best model, we further investigate the marginal posterior density of each mean. In each case, the closed-form of the conditional posterior density is available so that the CMDE can be used for comparison. Furthermore, we obtain two joint marginal posterior densities: the means of the DID and true amnesia groups, and the means of DID and symptom simulated groups. Note that the conditional posterior density of the former is analytically available while the latter requires numerical integration.

The experiment investigated whether a DID patient really suffers from amnesia when switching from one identity to another. To objectively evaluate the subjective experience of amnesia in DID patients and avoid the iatrogenic problem (i.e., patients' behaviors are induced by therapists) or some suggestive influence of media, DID patients ($n_1 = 19$) were implicitly measured for amnesia using the following procedure: after being told a brief story and shown some figures, they were asked to change their identity and then answer recognition questions about the story and figure details. To avoid the issue of symptom simulation in DID patients, three control groups composed of normal healthy people were recruited for comparison. The second (first control) group ($n_2 = 25$) was a normal control group. They were told the story and shown the figures, and were then asked to answer the questions. Without any intervention, their performance would be the performance of normal

people. The same procedure was used for the third (second control) group ($n_3 = 25$), but they were extensively informed about the behaviors of DID in advance and asked to deliberately simulate inter-identity amnesia. They can be viewed as the symptom simulation group. The fourth (third control) group ($n_4 = 25$) were asked to directly answer the questions without experiencing the story and graphs. Due to their complete lack of knowledge, this group represented a true amnesic group.

We model memory performance score y_{ij} for the i^{th} subject in group j as an independent observation from a normal distribution with mean μ_j and variance σ^2 for $i = 1, 2, \dots, n_j$ and $j = 1, \dots, J$. Let Θ denote the corresponding constrained parameter space for θ under the hypothesis, for example $\{\mu_1 < \mu_2 < \dots\}$, and let

$D = \{n, y\} = \left\{ \sum_{j=1}^J n_j, y_{11}, y_{21}, \dots, y_{n_1 1}, \dots, y_{1J}, y_{2J}, \dots, y_{n_J J} \right\}$ denote the observed data. Then,

the likelihood function is given by

$$L(\theta|D) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{j=1}^J \sum_{i=1}^{n_j} (y_{ij} - \mu_j)^2 \right\},$$

where $\theta = (\mu, \sigma^2) = (\mu_1, \dots, \mu_J, \sigma^2)$. For this model, a prior distribution in Chen and Ibrahim (2003) is considered: a power prior based on the prior predictive values of the performance score is assigned for μ given σ^2 , and an inverse gamma prior is assumed for σ^2 . Specifically, the joint prior is

$$\pi(\mu, \sigma^2 | y_0, a_0) \propto \left[(\sigma^2)^{-\frac{n_0}{2}} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{j=1}^J \sum_{i=1}^{n_{0j}} (y_{0ij} - \mu_j)^2 \right\} \right]^{a_0} (\sigma^2)^{-b_{01}-1} \exp \left(-\frac{b_{02}}{\sigma^2} \right) \times 1\{(\mu, \sigma^2) \in \Theta\},$$

where $y_0 = (y_{01}, y_{02}, \dots, y_{0J})$ with $y'_{0j} = (y_{01j}, y_{02j}, \dots, y_{0n_{0j}j})$ denotes the prior predictive values of the response variables, a_0 is a scalar parameter determining the degree of involvement of the prior predictive values in forming the prior distribution, $n_0 = \sum_{j=1}^J n_{0j}$ and $b_{01} > 0$ and $b_{02} > 0$ are prespecified hyperparameters. Under this constrained structure, the marginal posterior density of μ_j is analytically intractable, but can be evaluated by using an MCMC approach.

Assume $\sum_{i=1}^{n_j} y_{0ij} = 0$ for $j = 1, 2, \dots, 4$ and $b_{01} = b_{02} = 0.0001$. For comparing our proposed method with KDE and CMDE, we choose one of the best hypotheses: $\Theta = \{\mu_2 > (\mu_1, \mu_4) > \mu_3, \sigma^2 > 0\}$ under $a_0 = 0.01$ based on Bayesian analyses by Chen and Kim (2008), where Bayesian model selection of different hypotheses on mean scores is carried out with various criteria including L measure, deviance information criterion (DIC), the conditional predictive ordinate (CPO) statistic, and the marginal likelihood or Bayes factor. Under this hypothesis, the CMDE is applicable in estimating the marginal posterior density of each mean because the closed form of the conditional posterior density is available as follows:

$$\pi(\mu_1 | \mu_{(-1)}, \sigma^2, D, y_0, a_0) = \frac{\sqrt{\frac{n_1 + a_0 n_1}{2\pi\sigma^2}} \exp\left(-\frac{(n_1 + a_0 n_1) \left[\mu_1 - \frac{\sum y_{i1}}{n_1 + a_0 n_1}\right]^2}{2\sigma^2}\right)}{\left| \Phi\left(\frac{\mu_2 - \frac{\sum y_{i1}}{n_1 + a_0 n_1}}{\sqrt{\frac{\sigma^2}{n_1 + a_0 n_1}}}\right) - \Phi\left(\frac{\mu_3 - \frac{\sum y_{i1}}{n_1 + a_0 n_1}}{\sqrt{\frac{\sigma^2}{n_1 + a_0 n_1}}}\right) \right|},$$

$$\pi(\mu_2 | \mu_{(-2)}, \sigma^2, D, y_0, a_0) = \frac{\sqrt{\frac{n_2 + a_0 n_2}{2\pi\sigma^2}} \exp\left(-\frac{(n_2 + a_0 n_2) \left[\mu_2 - \frac{\sum y_{i2}}{n_2 + a_0 n_2}\right]^2}{2\sigma^2}\right)}{\left| 1 - \Phi\left(\frac{\max(\mu_1, \mu_4) - \frac{\sum y_{i2}}{n_2 + a_0 n_2}}{\sqrt{\frac{\sigma^2}{n_2 + a_0 n_2}}}\right) \right|},$$

$$\pi(\mu_3 | \mu_{(-3)}, \sigma^2, D, y_0, a_0) = \frac{\sqrt{\frac{n_3 + a_0 n_3}{2\pi\sigma^2}} \exp\left(-\frac{(n_3 + a_0 n_3) \left[\mu_3 - \frac{\sum y_{i3}}{n_3 + a_0 n_3}\right]^2}{2\sigma^2}\right)}{\left| \Phi\left(\frac{\min(\mu_1, \mu_4) - \frac{\sum y_{i3}}{n_3 + a_0 n_3}}{\sqrt{\frac{\sigma^2}{n_3 + a_0 n_3}}}\right) \right|},$$

$$\pi(\mu_4 | \mu_{(-4)}, \sigma^2, D, y_0, a_0) = \frac{\sqrt{\frac{n_4 + a_0 n_4}{2\pi\sigma^2}} \exp\left(-\frac{(n_4 + a_0 n_4) \left[\mu_4 - \frac{\sum y_{i4}}{n_4 + a_0 n_4}\right]^2}{2\sigma^2}\right)}{\left| \Phi\left(\frac{\mu_2 - \frac{\sum y_{i4}}{n_4 + a_0 n_4}}{\sqrt{\frac{\sigma^2}{n_4 + a_0 n_4}}}\right) - \Phi\left(\frac{\mu_3 - \frac{\sum y_{i4}}{n_4 + a_0 n_4}}{\sqrt{\frac{\sigma^2}{n_4 + a_0 n_4}}}\right) \right|},$$

where $\Phi(\cdot)$ denotes the cumulative standard normal distribution function. To implement PWMDE, we first choose the conditional working parameter space $\tilde{\Theta}_{\mu_{(-j)}, \sigma_t^2}$ of each $\pi(\mu_j |$

$\mu_{(-j)}, \sigma^2, D, y_0, a_0)$ to be $\tilde{\Theta}_{\mu_{(-1t)}, \sigma_t^2}^2 = \{\mu_{3t} < \mu_{1t} < \mu_{2t}\},$

$\tilde{\Theta}_{\mu_{(-2t)}, \sigma_t^2}^2 = \left\{ \max(\mu_{1t}, \mu_{4t}) < \mu_{2t} < 15 \right\}, \tilde{\Theta}_{\mu_{(-3t)}, \sigma_t^2}^2 = \{0 < \mu_{3t} < \min(\mu_{1t}, \mu_{4t})\},$ and

$\tilde{\Theta}_{\mu_{(-4t)}, \sigma_t^2}^2 = \{\mu_{3t} < \mu_{4t} < \mu_{2t}\},$ where μ_{jt} is the t^{th} MCMC sample point for the j^{th} parameter.

Notice that the chosen conditional working parameter space for $\tilde{\Theta}_{\mu_{(-jt)}, \sigma_t^2}^2, j = 1, 4$ is exactly

equal to the whole support of the conditional posterior distribution $\pi(\mu_j | \mu_{(-j)}, \sigma^2, D, y_0, a_0), j = 1, 4$ because μ_1 and μ_4 are bounded on both sides. For μ_2 and μ_3 , which are each bounded on one side, we set 15 as the upper bound and 0 as the lower bound for each respective parameter such that the working space can cover most of the MCMC sample. Then, we decide on the number of subsets K in the conditional working parameter space.

Figure 1 includes the four estimated curves of KDE in the first column (using the function, `density(.)`, with the default setting in R software), and PWMDE curves with $K = 20$ and $K = 50$ in the second and third columns, respectively. Figure 2 plots the difference of each estimated curve in Figure 1 from the CMDE estimated curve. It shows that KDE deviates from the CMDE for each parameter in the region around its posterior mode. Whereas, PWMDE can produce results very similar to CMDE even when the number of partition subsets is small. When K is increased to $K = 50$, PWMDE completely overlaps with the CMDE, empirically confirming that PWMDE can be a realization of CMDE as discussed in Remark 3.2.

Next, we examine the performance of PWMDE in the joint marginal posterior density of μ_1 and μ_4 . As in the one-dimensional problem, the CMDE is applicable because the closed form for the conditional posterior density is available:

$$\pi(\mu_1, \mu_4 | \mu_{(-1, -4)}, \sigma^2, D, y_0, a_0) = \prod_{j=1,4} \frac{\sqrt{\frac{n_j + a_0 n_j}{2\pi\sigma^2}} \exp \left[-\frac{\left((n_j + a_0 n_j) \left| \mu_j - \frac{\sum_i y_{ij}}{n_j + a_0 n_j} \right| \right)^2}{2\sigma^2} \right]}{\left[\Phi \left(\frac{\mu_2 - \frac{\sum_i y_{ij}}{n_j + a_0 n_j}}{\sqrt{\frac{\sigma^2}{n_j + a_0 n_j}}} \right) - \Phi \left(\frac{\mu_3 - \frac{\sum_i y_{ij}}{n_j + a_0 n_j}}{\sqrt{\frac{\sigma^2}{n_j + a_0 n_j}}} \right) \right]}.$$

For PWMDE, the conditional working parameter space $\tilde{\Theta}_{\theta_{(-1t, -4t)}}^2$ is set to be $\{\mu_{3t} < (\mu_{1t}, \mu_{4t}) < \mu_{2t}\}$, then the working parameter space in each dimension is equally divided into 20 or 50 pieces so that $K = 400$ or $K = 2, 500$. Figure 3 shows the estimated joint marginal posterior densities by KDE (using the function, `kde2d(.)`, with the default setting in R software), PWMDE ($K=400$), and PWMDE ($K=2,500$) based on an MCMC sample of size $T = 1, 000$. The difference of each from the CMDE estimate is shown below the estimated density surface. Figure 3 shows that KDE is quite different from the CMDE in this two-

dimensional problem, while only minor differences are detectable in the center using PWMDE even with a small number of partition subsets. These slight differences disappear when K is increased.

Finally, we estimate the joint marginal posterior density of μ_1 and μ_3 , where the two parameters of interest are constrained with respect to each other ($\mu_1 > \mu_3$) and the corresponding univariate marginal posterior densities are not far away (Figure 1). Under this special structure, the evaluation of the conditional posterior density, $\pi(\mu_1, \mu_3/\mu_{(-1,-3)}, \sigma^2, D, y_0, a_0)$, requires numerical integration. The conditional parameter space $\Theta_{\mu_{(-1t, -3t)}, \sigma^2}$ for PWMDE is the trapezoid area bordered by lines $\mu_1 = \mu_2$, $\mu_3 = \mu_4$ and $\mu_1 = \mu_3$ and the conditional working parameter space $\tilde{\Theta}_{\mu_{(-1t, -3t)}, \sigma^2}$ is the union of small rectangles inside of $\Theta_{\mu_{(-1t, -3t)}, \sigma^2}$ as shown in Figure 4.

Figure 5 shows the performance of KDE (using the function, `kde2d(.)`, with the default setting in the R software) and PWMDE based on an MCMC sample of size $T = 1,000$. PWMDE performs well even for the smaller number of partition subsets, $K = 590$ (190 for the green area and 400 for the gray area). In contrast, KDE again deviates from the CMDE. Because it does not use the structure of the known unnormalized posterior density, the KDE has positive values in the area beyond the red dashed line, where the joint marginal posterior density values are supposed to be 0.

In order to report the difference between each of the estimates and the CMDE quantitatively in addition to plotting these differences as shown in Figures 2, 3, and 5, we consider an approximately piecewise L_1 -distance. This distance measure between two estimated curves $\hat{\pi}_1(\theta|D)$ and $\hat{\pi}_2(\theta|D)$ is defined by $\Delta(\hat{\pi}_1, \hat{\pi}_2) = \sum_{g=1}^G V(A_g) \max_{\theta \in A_g} |\hat{\pi}_1(\theta|D) - \hat{\pi}_2(\theta|D)|$ where $V(A_g)$ is the volume of A_g and $\{A_g, g = 1, 2, \dots, G\}$ is a partition of the support of the marginal density under consideration. Table 4 shows the results. We see from this table that (i) the distance between the KDE and the CMDE is much larger than the one between the PWMDE and the CMDE for each comparison and (ii) the distance between the PWMDE and the CMDE becomes smaller as K increases for most of the cases except for μ_2 in Figure 2.

6 PWMDE for Bayesian Variable Selection

6.1 The Basic Formulation

Suppose there are p regression parameters, β , in the full model M , and $\beta \in R^p$. In variable selection, we let $\beta^{(m)}$ and $\beta^{(-m)}$ denote the regression parameters included and excluded in the reduced model m , respectively. Then, $\beta^{(M)} = \beta = ((\beta^{(m)})', (\beta^{(-m)})')'$ holds for all m , and $\beta^{(-M)} = \emptyset$. We further let α denote the nuisance parameters in the model. Under this setting, two conditions required for the Savage-Dickey density ratio in Dickey (1971) are given in Chen et al. (1999) as follows:

Condition I. $L(\boldsymbol{\beta}^{(m)}, \boldsymbol{a}/D, m) = L(\boldsymbol{\beta}^{(m)}, \boldsymbol{\beta}^{(-m)} = \mathbf{0}, \boldsymbol{a}/D, M)$, where $L(\boldsymbol{\beta}^{(m)}, \boldsymbol{a}/D, m)$ is the likelihood function under the reduced model m and $L(\boldsymbol{\beta}^{(m)}, \boldsymbol{\beta}^{(-m)} = \mathbf{0}, \boldsymbol{a}/D, M)$ is the likelihood function under the full model evaluated at $\boldsymbol{\beta}^{(m)}, \boldsymbol{\beta}^{(-m)} = \mathbf{0}$ and $\boldsymbol{a}$;

Condition II. $\pi(\boldsymbol{\beta}^{(m)}, \boldsymbol{a}/m) = \pi(\boldsymbol{\beta}^{(m)}, \boldsymbol{a}/\boldsymbol{\beta}^{(-m)} = \mathbf{0}, M)$, where $\pi(\boldsymbol{\beta}^{(m)}, \boldsymbol{a}/m)$ is the prior distribution specified under the reduced model m and $\pi(\boldsymbol{\beta}^{(m)}, \boldsymbol{a}/\boldsymbol{\beta}^{(-m)} = \mathbf{0}, M)$ is the conditional prior distribution of $\boldsymbol{\beta}^{(m)}, \boldsymbol{a}$ given $\boldsymbol{\beta}^{(-m)} = \mathbf{0}$ under the full model.

If both conditions hold, the Bayes factor of the reduced model m over the full model M can be simplified to

$$BF = \frac{c(D|m)}{c(D|M)} = \frac{\pi(\boldsymbol{\beta}^{(-m)} = \mathbf{0}|D, M)}{\pi(\boldsymbol{\beta}^{(-m)} = \mathbf{0}|M)}, \quad (20)$$

where $c(D|m)$ and $c(D|M)$ are the marginal likelihoods under the reduced model and the full model, respectively, and $\pi(\boldsymbol{\beta}^{(-m)} = \mathbf{0}|D, M)$ and $\pi(\boldsymbol{\beta}^{(-m)} = \mathbf{0}|M)$ denote the marginal posterior and prior densities evaluated at $\boldsymbol{\beta}^{(-m)} = \mathbf{0}$ from the full model, respectively. It shows that the Bayes factor can be simplified as the function of the marginal posterior density evaluated at zero values of the parameters excluded from the reduced model m so that the model comparison can be done by using a single MCMC sample from the posterior density of the full model. To estimate the marginal posterior density in (20), we can use the PWMDE in (15) by assigning $\boldsymbol{\theta} = \boldsymbol{\beta}^{(-m)}$, $\boldsymbol{\theta}_0 = \mathbf{0}$, and $\boldsymbol{\xi} = ((\boldsymbol{\beta}^{(m)})', \boldsymbol{a}')'$. Because the parameters in the marginal posterior density are all regression parameters with no constraints, the partition of the conditional working parameter space can be easily constructed via rings.

Suppose $\{(\boldsymbol{\beta}_t^{(m)}, \boldsymbol{\beta}_t^{(-m)}, \boldsymbol{\alpha}_t), t = 1, \dots, T\}$ is an MCMC sample from the posterior distribution under the full model. Let $\boldsymbol{\theta}_t = \boldsymbol{\beta}_t^{(-m)}$ and $\boldsymbol{\xi}_t = ((\boldsymbol{\beta}_t^{(m)})', \boldsymbol{\alpha}_t)'$ for $t = 1, 2, \dots, T$. Then, we have

$$\hat{\pi}_t(\boldsymbol{\theta} = \mathbf{0}|D, M) = \sum_{k=1}^K \frac{q(\boldsymbol{\theta}_k^*, \boldsymbol{\xi}_t) 1\{\boldsymbol{\theta}_t \in A_k\}}{\sum_{k=1}^K q(\boldsymbol{\theta}_k^*, \boldsymbol{\xi}_t) V(A_k)} \frac{q(\boldsymbol{\theta} = \mathbf{0}, \boldsymbol{\xi}_t)}{q(\boldsymbol{\theta}_t, \boldsymbol{\xi}_t)},$$

where $\boldsymbol{\theta}_k^* \in A_k$ is a representative point in the k^{th} ring for $k = 1, 2, \dots, K$. Assume that the marginal prior density $\pi(\boldsymbol{\theta} = \mathbf{0}|M)$ is analytically available. Then, the Bayes factor in log scale is estimated as

$$\log \widehat{BF}_{\text{PWMDE}} = \log \left\{ \frac{1}{T} \sum_{t=1}^T \hat{\pi}_t(\boldsymbol{\theta} = \mathbf{0}|D, M) \right\} - \log \pi(\boldsymbol{\theta} = \mathbf{0}|M). \quad (21)$$

We use the Overlapping Batch Statistics (OBS) of Schmeiser et al. (1990) to estimate the MC standard error of $\log \widehat{BF}_{\text{PWMDE}}$. Write $\hat{\eta} = \log \widehat{BF}_{\text{PWMDE}}$. Let $\hat{\eta}_b$ denote an estimate of the

Bayes factor in log scale via (21) using the b^{th} batch, $\{(\beta_t^{(m)}, \beta_t^{(-m)}, \alpha_t), t = b, \dots, b+B-1\}$ of the MCMC sample for $b = 1, \dots, T-B+1$, where $B < T$ is the batch size. Then, the OBS estimated MC standard error of $\hat{\eta}$ is given by

$$\sqrt{\widehat{Var}(\hat{\eta})} = \left\{ \left[\frac{B}{T-B} \right] \frac{\sum_{b=1}^{T-B+1} (\hat{\eta}_b - \bar{\eta}_{OBS})^2}{T-B+1} \right\}^{1/2}, \quad (22)$$

where $\bar{\eta}_{OBS} = \sum_{b=1}^{T-B+1} \hat{\eta}_b / (T-B+1)$. According to Schmeiser et al. (1990), a reasonable choice of batch size B is $10 \leq B \leq 20$.

6.2 The Ordinal Probit Regression Model

We consider the model selection problem under the ordinal probit regression model, in which the prior distribution does not satisfy Condition II. Thus, the Savage-Dickey density ratio does not hold. Under this situation, we show how marginal posterior density estimation can still be used for estimating ratios of marginal likelihoods between each reduced model and the full model.

Let $\mathbf{y} = (y_1, y_2, \dots, y_n)'$ denote the vector of observed ordinal responses, each coded as one value from $0, 1, \dots, J-1$, and $\mathbf{X}$ denote the $n \times p$ covariate matrix with the i^{th} row equal to the covariates of the i^{th} subject $\mathbf{x}_i'$. Let $D = (\mathbf{y}, \mathbf{X}, n)$. Following Nandram and Chen (1996), the likelihood function is then given by

$$L(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\gamma}}, \sigma | D) = \prod_{i=1}^n \left[\Phi \left(\frac{\tilde{\gamma}_{y_i+1} - \mathbf{x}_i' \tilde{\boldsymbol{\beta}}}{\sigma} \right) - \Phi \left(\frac{\tilde{\gamma}_{y_i} - \mathbf{x}_i' \tilde{\boldsymbol{\beta}}}{\sigma} \right) \right], \quad (23)$$

where $\tilde{\boldsymbol{\gamma}} = (\tilde{\gamma}_2, \dots, \tilde{\gamma}_{J-2})'$, $-\infty = \tilde{\gamma}_0 < \tilde{\gamma}_1 = 0 \leq \tilde{\gamma}_2 \leq \dots \leq \tilde{\gamma}_{J-1} = 1 < \tilde{\gamma}_J = \infty$ are the cutoff points, $\tilde{\boldsymbol{\beta}}$ is a p -dimensional vector of the regression coefficients, and $\sigma > 0$ is a scale parameter. The likelihood function in (23) is derived based on the reparameterization of the ordinal regression model with latent variables proposed by Albert and Chib (1993) to accelerate convergence of the Gibbs sampling algorithm. Another attractive feature of (23) is that when $J=3$, there are no unknown cutoff points. The priors of $(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\gamma}}, \sigma^2)$ considered in Chen (2005b) are given by

$$\pi(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\gamma}}, \sigma^2) = \pi(\tilde{\boldsymbol{\beta}} | \sigma^2) \pi(\sigma^2) \pi(\tilde{\boldsymbol{\gamma}}), \quad (24)$$

where $\tilde{\boldsymbol{\beta}} | \sigma^2 \sim N(0, \tau_0^{-1} \sigma^2 \mathbf{I}_p)$, σ^2 follows an inverse gamma distribution $IG(b_{01}, b_{02})$, $\tilde{\boldsymbol{\gamma}}$ is assigned a uniform prior on the constrained space of $\tilde{\boldsymbol{\gamma}}$, $\mathbf{I}_p$ is the $p \times p$ identity matrix, and τ_0

> 0 , $b_{01} > 0$, and $b_{02} > 0$ are prespecified hyperparameters. Then, the resulting posterior distribution is given by

$$\pi(\widetilde{\boldsymbol{\beta}}, \widetilde{\boldsymbol{\gamma}}, \sigma^2 | D) \propto L(\widetilde{\boldsymbol{\beta}}, \widetilde{\boldsymbol{\gamma}}, \sigma | D) \pi(\widetilde{\boldsymbol{\beta}}, \widetilde{\boldsymbol{\gamma}}, \sigma^2). \quad (25)$$

To carry out Bayesian variable selection under the ordinal probit regression model, we let $(\mathbf{x}_i^{(m)}, \widetilde{\boldsymbol{\beta}}^{(m)})$ and $(\mathbf{x}_i^{(-m)}, \widetilde{\boldsymbol{\beta}}^{(-m)})$ denote the p_m and $p - p_m$ covariates and the corresponding regression coefficients included in and excluded from the m^{th} reduced model, respectively, such that $0 < p_m < p$, $\mathbf{x}_i = ((\mathbf{x}_i^{(m)})', (\mathbf{x}_i^{(-m)})')'$, and $\widetilde{\boldsymbol{\beta}} = ((\widetilde{\boldsymbol{\beta}}^{(m)})', (\widetilde{\boldsymbol{\beta}}^{(-m)})')'$. To ensure that the prior of $(\widetilde{\boldsymbol{\beta}}^{(m)}, \sigma^2, \boldsymbol{\gamma})$ under the reduced model m has the same structure as the one given in (24) under the full model, we take

$$\pi(\widetilde{\boldsymbol{\beta}}^{(m)}, \widetilde{\boldsymbol{\gamma}}, \sigma^2 | m) = \pi(\widetilde{\boldsymbol{\beta}}^{(m)} | \sigma^2) \pi(\sigma^2) \pi(\widetilde{\boldsymbol{\gamma}}), \quad (26)$$

where $\widetilde{\boldsymbol{\beta}}^{(m)} | \sigma^2 \sim N(0, \tau_0^{-1} \sigma^2 I_{p_m})$ and $\pi(\sigma^2)$ and $\pi(\widetilde{\boldsymbol{\gamma}})$ are defined in (24). The likelihood function $L(\widetilde{\boldsymbol{\beta}}^{(m)}, \widetilde{\boldsymbol{\gamma}}, \sigma | D, m)$ under the reduced model is given by (23) with $\mathbf{x}_i' \widetilde{\boldsymbol{\beta}}$ replaced by $(\mathbf{x}_i^{(m)})' \widetilde{\boldsymbol{\beta}}^{(m)}$. For this application, $\boldsymbol{\alpha} = (\widetilde{\boldsymbol{\gamma}}', \sigma^2)'$. Because $\mathbf{x}_i' \widetilde{\boldsymbol{\beta}} |_{\widetilde{\boldsymbol{\beta}}^{(-m)} = 0} = (\mathbf{x}_i^{(m)})' \widetilde{\boldsymbol{\beta}}^{(m)}$, Condition I holds. From (24), we have

$$\begin{aligned} \pi(\widetilde{\boldsymbol{\beta}}^{(m)}, \widetilde{\boldsymbol{\gamma}}, \sigma^2 | \boldsymbol{\beta}^{(-m)} = \mathbf{0}) &\propto (\sigma^2)^{-p/2} \\ &\exp\left[-\frac{\tau_0}{2\sigma^2} (\widetilde{\boldsymbol{\beta}}^{(m)})' (\widetilde{\boldsymbol{\beta}}^{(m)})\right] (\sigma^2)^{-(b_{01} + 1)} \exp(-b_{02}/\sigma^2), \end{aligned}$$

which is clearly not equal to $\pi(\widetilde{\boldsymbol{\beta}}^{(m)}, \widetilde{\boldsymbol{\gamma}}, \sigma^2 | m)$ given in (26) when $p_m < p$. Thus, Condition II is not satisfied. Under this setting, the Bayes factor of the reduced model m over the full model cannot be calculated using (20). To circumvent this problem, we take the following one-to-one transformations:

$$\boldsymbol{\beta} = \frac{\widetilde{\boldsymbol{\beta}}}{\sqrt{\sigma^2}}, \quad \gamma_j = \frac{\widetilde{\gamma}_j}{\sqrt{\sigma^2}}, \quad j = 2, \dots, J-2, \quad \text{and} \quad \gamma_{J-1} = \frac{1}{\sqrt{\sigma^2}}. \quad (27)$$

The absolute value of the determinant of the Jacobian of the transformations from $(\widetilde{\boldsymbol{\beta}}, \widetilde{\boldsymbol{\gamma}}, \sigma^2)$ to $(\boldsymbol{\beta}, \boldsymbol{\gamma})$ is $\gamma_{J-1}^{-(p+J)}$. Write $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_2, \dots, \boldsymbol{\gamma}_{J-1})'$. After the transformations, the likelihood function of $(\boldsymbol{\beta}, \boldsymbol{\gamma})$ is given by

$$L(\boldsymbol{\beta}, \boldsymbol{\gamma} | D) = \prod_{i=1}^n \left[\Phi(\gamma_{y_{i+1}} - \mathbf{x}'_i \boldsymbol{\beta}) - \Phi(\gamma_{y_i} - \mathbf{x}'_i \boldsymbol{\beta}) \right], \quad (28)$$

and the prior of $(\boldsymbol{\beta}, \boldsymbol{\gamma})$ is given by

$$\pi(\boldsymbol{\beta}, \boldsymbol{\gamma}) \propto \exp\left(-\frac{\tau_0}{2} \boldsymbol{\beta}' \boldsymbol{\beta}\right) \gamma_{J-1}^{-J+2b_{01}+2} \exp(-b_{02} \gamma_{J-1}^2), \quad (29)$$

where $\boldsymbol{\beta} \in R^p$ and $-\infty = \gamma_0 < \gamma_1 = 0 < \gamma_2 < \dots < \gamma_{J-1} < \gamma_J = \infty$. Now, applying the Bayesian variable selection procedure to $(\boldsymbol{\beta}, \boldsymbol{\gamma})$ with the likelihood function and the prior given by (28) and (29), we can show that both Conditions 1 and 2 are satisfied. Therefore, we can use the PWMDE via (20) to compute the Bayes factor.

Although the use of (28) and (29) leads to easy computation of the Bayes factor, sampling $(\boldsymbol{\beta}, \boldsymbol{\gamma})$ from the posterior distribution induced by (28) and (29) is not as efficient as sampling $(\widetilde{\boldsymbol{\beta}}, \widetilde{\boldsymbol{\gamma}}, \sigma^2)$ from (25) (e.g. Nandram and Chen, 1996). Because the transformations in (27) are one-to-one, MCMC samples of $(\widetilde{\boldsymbol{\beta}}, \widetilde{\boldsymbol{\gamma}}, \sigma^2)$ from (25), which can be generated by Nandram-Chen algorithm, can be directly used to obtain MCMC samples of $(\boldsymbol{\beta}, \boldsymbol{\gamma})$ via (27). Therefore, by combining these two settings of the ordinal probit regression model, we can achieve both convenient implementation of MCMC sampling and efficient computation of Bayes factors.

6.3 Analysis of the Prostate Cancer Data

We apply the ordinal probit regression model to the prostate cancer data ($n = 713$) as in Chen (2005b). We examine the relationships between a clinical ordinal response variable, Pathological Extracapsular Extension (PECE, y), and three covariates: Prostate Specific Antigen (PSA, x_1), Clinical Gleason Score (GLEAS, x_2, x_3), and Clinical Stage (CSTAGE, x_4). Here, PECE takes values of 0, 1, or 2, where 0 means that there is no cancer cell present in or near the capsule, 1 denotes that the cancer cells extend into but not through the capsule, and 2 indicates that cancer cells extend through the capsule. PSA is a continuous variable while CSTAGE is a binary outcome, which is assigned to 1 if the 1992 American Joint Commission on Cancer (AJCC) clinical stage T-category is 1, and assigned to 2 if the T-category is 2 or higher. As for GLEAS, considering similar biologic behaviors of tumors (well-differentiated, moderately-differentiated, or poorly-differentiated), it is trichotomized by two dummy variables: $x_2 = 1$ if GLEAS = 7, otherwise $x_2 = 0$; $x_3 = 1$ if GLEAS > 7, otherwise $x_3 = 0$. We note that in Chen (2005b), GLEAS was treated as a continuous variable. Because x_2 and x_3 are defined from the same variable GLEAS, we either include both x_2 and x_3 in the model or exclude them together from the model.

In this example, because $J = 3$, there are no unknown cutpoints in (23). In (24), we take $b_{01} = 1$ and $b_{02} = 0.1$. Due to lack of historical information about τ_0 and to avoid Bartlett's or Lindley's paradox (see Je reys, 1998; Lindley, 1957), we first use an empirical Bayes

method to find the best prior setting of τ_0 for the full model based on the marginal likelihood criterion. The PWK estimator (Wang et al., 2018) was used to calculate the marginal likelihoods under different τ_0 . Table 5 summarizes the marginal likelihood estimate and its MCSE for each τ_0 using an MCMC sample of size $T = 10,000$. The MCSE is calculated by (22) with the batch size $B = 1,000$.

From the results in Table 5, we choose $\tau_0 = 15$ as the best prior under the full model. On this basis, we use the Bayes factor for variable selection. Following (27), we first transform the regression parameters to satisfy the conditions for Savage-Dickey density ratio. Then, we apply PWMDE to estimate the marginal posterior density at zero values in (20) by constructing a ring partition of the working parameter space with $r = 4$ and $K = 1, 5$, or 10. Table 6 shows the 95% HPD interval for each β , Bayes factor estimates of the reduced models over the full model in log scale, their estimated MCSE's using (22) with the batch size $B = 1,000$, and their relative MCSEs denoted by rMCSEs, each of which is defined as MCSE divided by its $\log \widehat{BF}$. The results are obtained based on an MCMC sample of size 10,000. From Table 6, we see that the full model is preferred because all the values of $\log BF$ are negative. This result is also supported by the 95% HPD intervals, which do not contain 0. In addition, we observe that each MCSE dramatically drops when a reasonable number of subsets ($K = 5$ or 10) is used. The fact that each MCSE is relatively small compared to the magnitude of $\log BF$ empirically demonstrates that our proposed method is very accurate. Table 6 also shows the results for $\tau_0 = 0.1$, where the inconsistency between the Bayes factors and the 95% HPD intervals is evident. Even though the 95% HPD intervals suggest that all covariates are important for the case of $\tau_0 = 0.1$, the Bayes factor tends to favor the reduced model, which is known as the Bartlett's or Lindley's paradox.

We also use the KDE with several choices of the kernel function to compute $\pi(\beta^{-m}) = 0|D, M$ in (20). The KDE estimates and corresponding MCSEs of $\log BF$ for the model (x_1, x_2, x_3) when $\tau_0 = 0.1$ are 1.466 and 0.060 under the Gaussian kernel; 1.475 and 0.059 under the Epanechnikov kernel; 1.446 and 0.069 under the rectangular kernel; 1.472 and 0.059 under the triangular kernel; 1.473 and 0.058 under the biweight kernel; 1.472 and 0.058 under the cosine kernel; and 1.475 and 0.058 under the optcosine kernel. These MCSEs are larger than 0.010 shown in Table 6 using PWMDE with $K = 5$ or $K = 10$.

To examine empirical performance of the estimates (15) and (16) for $\pi(\theta_{01} = 0|D)$ and $\pi(\theta_{02} = 0|\theta_{01} = 0, D)$ using two MCMC samples, we consider models (x_1, x_4) and (x_2, x_3) when $\tau_0 = 0.1$. In order to make a fair comparison, we generate 5,000 MCMC sample points from both the full posterior distribution and the conditional posterior distribution. The resulting values of $\log BF$, MCSE, and rMCSE are 0.331, 0.021, and 6.42%, respectively, under model (x_1, x_4) , and -15.849 , 0.053, and 0.33%, respectively, under model (x_2, x_3) . In these computations, we fixed $r = 4$ and $K = 10$. In both cases, the estimates of $\log BF$ using (15) and (16) for $\pi(\theta_{01} = 0|D)$ and $\pi(\theta_{02} = 0|\theta_{01} = 0, D)$ have smaller MCSEs and rMCSE. These results empirically demonstrate that the approach based on the dimension reduction is quite promising in obtaining a more accurate estimate of $\log BF$.

7 Discussion

Marginal posterior density estimation provides complete information for the parameters of interest in Bayesian inference. While it is known that the conditional marginal density estimator is the best (minimizing the variance) in the class of importance weighted marginal density estimators (Chen, 1994), the method for realizing this estimator has not been accessible when the closed form of the conditional posterior density is not available. To circumvent it, we propose the partition weighted marginal density estimator (PWMDE) that requires only the unnormalized posterior density and an MCMC sample from the joint posterior distribution. As shown in our simulation studies (Section 4) and real data analyses (Section 5 and Section 6.3), the proposed PWMDE performs well when the dimension of the parameters of interest is not high. In order for our method to work well in general, we need the kernel function within the subset $A_k(\xi)$ takes on values close to the same constant. Therefore, when the dimension of the parameters of interest is high, PWMDE requires a large number of sets $A_k(\xi)$'s in order to achieve good performance, which may be computationally inhibitive. A strategy via expressing a high-dimensional density as a sequence of low-dimensional conditional densities discussed in Section 3.2 may provide a solution to the high-dimensional problem. However, this approach requires additional MCMC samples for these low-dimensional conditional distributions, which certainly increases computational cost.

As discussed in Section 3 and shown in Section 6, PWMDE is useful not only in computing marginal posterior densities but also in estimating marginal likelihoods through Chib's identity and Bayes factors via the Savage-Dickey density ratio. There is a rich literature on the Bayes factor estimates for Bayesian variable selection. When the number of variables is small to moderate (say, up to 20), the performance of the Bayes factor estimates using IWMDE has been examined in Chen et al. (1999, 2003), and Ibrahim et al. (1999, 2000). As shown in our simulation studies in Section 4, PWMDE is more efficient than the IWMDE. We note that the very same weight function specified in our simulation studies was also used in _S1_Reference9Chen et al. (1999, 2003) and Ibrahim et al. (1999, 2000). Therefore, the Bayes factor estimates with PWMDE are expected to perform well for a small to moderate number of variables. When the number of variables is large, it is computationally infeasible even to enumerate all possible models. In this regards, various computational algorithms have been developed, for example, including Hans et al. (2007), and Berger and Molina (2005). The PWMDE method in conjunction with these computational algorithms may be potentially useful for Bayesian computational problems with high-dimensional parameters.

Currently, KDE is a standard method for computing and displaying marginal posterior densities using MCMC samples in existing Bayesian software. As empirically shown in Sections 5 and 6, PWMDE can be much closer to CMDE and produce more accurate estimates of Bayes factors than KDE. Since the proposed method requires only the unnormalized posterior density and an MCMC sample from the joint posterior distribution, PWMDE has a potential to become a "black-box" algorithm, which can be implemented in existing software including SAS, OpenBugs (Thomas et al., 2006), as well as more specialized Bayesian software such as that used in phylogenetics: RevBayes (Höhna et al., 2016) and BEAST2 (Bouckaert et al., 2014).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors gratefully thank the editor, the associate editor, and the two referees for their constructive comments and suggestions that help improve the article. This material is based upon work supported by the National Science Foundation under Grant No. DEB-1354146. Dr. M.-H. Chen's research was also partially supported by NIH grants #GM70335 and #P01CA142538.

References

- Albert JH and Chib S (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association* 88, 669–679.
- Berger J and Molina G (2005). Posterior model probabilities via path-based pairwise priors. *Statistica Neerlandica* 59, 3–15.
- Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu C-H, Xie D, Suchard MA, Rambaut A, and Drummond AJ (2014, 4). BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Computational Biology* 10 (4), e1003537. [PubMed: 24722319]
- Chen M-H (1994). Importance-weighted marginal Bayesian posterior density estimation. *Journal of the American Statistical Association* 89 (427), 818–824.
- Chen M-H (2005a). Bayesian computation: from posterior densities to Bayes factors, marginal likelihoods, and posterior model probabilities In Dey D and Rao C (Eds.), *Handbook of Statistics* 25, *Bayesian Thinking: Modeling and Computation*, pp. 437–457. Elsevier.
- Chen M-H (2005b). Computing marginal likelihoods from a single MCMC output. *Statistica Neerlandica* 59, 16–29.
- Chen M-H and Ibrahim JG (2003). Conjugate priors for generalized linear models. *Statistica Sinica* 13 (2), 461–476.
- Chen M-H, Ibrahim JG, Shao Q-M, and Weiss RE (2003). Prior elicitation for model selection and estimation in generalized linear mixed models. *Journal of Statistical Planning and Inference* 111, 57–76.
- Chen M-H, Ibrahim JG, and Yiannoutsos C (1999). Prior elicitation, variable selection and Bayesian computation for logistic regression models. *Journal of the Royal Statistical Society. Series B, Statistical Methodology* 61, 223–242.
- Chen M-H and Kim S (2008). The Bayes factor versus other model selection criteria for the selection of constrained models In Hoijtink H, Klugkist I, and Boelen P (Eds.), *Bayesian Evaluation of Informative Hypotheses*, pp. 155–180. Springer.
- Chib S (1995). Marginal likelihood from the Gibbs output. *Journal of the American Statistical Association* 90 (432), 1313–1321.
- Chib S and Jeliazkov I (2001). Marginal likelihood from the Metropolis-Hastings output. *Journal of the American Statistical Association* 96, 270–281.
- Dickey JM (1971). The weighted likelihood ratio, linear hypotheses on normal location parameters. *The Annals of Mathematical Statistics* 42, 204–223.
- Fan Y, Wu R, Chen M-H, Kuo L, and Lewis PO (2011). Choosing among partition models in Bayesian phylogenetics. *Molecular Biology and Evolution* 28 (1), 523–532. [PubMed: 20801907]
- Friel N and Pettitt AN (2008). Marginal likelihood estimation via power posteriors. *Journal of the Royal Statistical Society. Series B, Statistical Methodology* 70 (3), 589–607.
- Gelfand AE, Smith AF, and Lee T-M (1992). Bayesian analysis of constrained parameter and truncated data problems using Gibbs sampling. *Journal of the American Statistical Association* 87 (418), 523–532.
- Goh G and Dey DK (2014). Bayesian model diagnostics using functional Bregman divergence. *Journal of Multivariate Analysis* 124, 371–383.

- Hans C, Dobra A, and West M (2007). Shotgun stochastic search for “large p” regression. *Journal of the American Statistical Association* 102, 507–516.
- Höhna S, Landis MJ, Heath TA, Boussau B, Lartillot N, Moore BR, Huelsenbeck JP, and Ronquist F (2016). RevBayes: Bayesian phylogenetic inference using graphical models and an interactive model-specification language. *Systematic Biology* 65 (4), 726–736. [PubMed: 27235697]
- Hoijtink H, Klugkist I, and Boelen P (2008). *Bayesian Evaluation of Informative Hypotheses*. Springer Science & Business Media.
- Ibrahim JG, Chen M-H, and MacEachern SN (1999). Bayesian variable selection for proportional hazards models. *The Canadian Journal of Statistics* 27 (4), 701–717.
- Ibrahim JG, Chen M-H, and Ryan LM (2000). Bayesian variable selection for time series count data. *Statistica Sinica* 10, 971–987.
- Jeffreys H (1998). *The Theory of Probability*. OUP Oxford.
- Lindley DV (1957). A statistical paradox. *Biometrika* 44, 187–192.
- Lui D (2014). Monte Carlo framework for prostate cancer correction and reconstruction in endorectal multi-parametric MRI. M.S. Thesis, Department of Systems Design Engineering, University of Waterloo, Canada.
- Lui D, Modhafar A, Haider MA, and Wong A (2015). Monte Carlo-based noise compensation in coil intensity corrected endorectal MRI. *BMC Medical Imaging* 15 (1), 1. [PubMed: 25645550]
- Meng X-L and Wong WH (1996). Simulating ratios of normalizing constants via a simple identity: a theoretical exploration. *Statistica Sinica* 6, 831–860.
- Nandram B and Chen M-H (1996). Reparameterizing the generalized linear model to accelerate Gibbs sampler convergence. *Journal of Statistical Computation and Simulation* 54, 129–144.
- Oh M-S (1999). Estimation of posterior density functions from a posterior sample. *Computational Statistics & Data Analysis* 29 (4), 411–427.
- Parzen E (1962). On estimation of a probability density function and mode. *The Annals of Mathematical Statistics* 33, 1065–1076.
- Rosenblatt M (1956). Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics* 27 (3), 832–837.
- Schmeiser BW, Avramidis TN, and Hashem S (1990). Overlapping batch statistics. In *Proceedings of the 22nd Conference on Winter Simulation*, pp. 395–398. IEEE Press.
- Stephens M and Donnelly P (2000). Inference in molecular population genetics. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 62 (4), 605–635.
- Thomas A, O’Hara B, Ligges U, and Sturtz S (2006). Making BUGS open. *R news* 6 (1), 12–17.
- Wang Y-B, Chen M-H, Kuo L, and Lewis PO (2018). A new Monte Carlo method for estimating marginal likelihoods. *Bayesian Analysis* 13 (2), 311–333. [PubMed: 29805725]
- Xie W, Lewis PO, Fan Y, Kuo L, and Chen M-H (2011). Improving marginal likelihood estimation for Bayesian phylogenetic model selection. *Systematic Biology* 60 (2), 150–160. [PubMed: 21187451]
- Yu JZ and Tanner MA (1999). An analytical study of several Markov chain Monte Carlo estimators of the marginal likelihood. *Journal of Computational and Graphical Statistics* 8 (4), 839–853.

Figure 1:

The estimated marginal posterior density curve for each μ_j with the fixed MCMC sample size $T = 1,000$ by (a) KDE, (b) PWMDE (K=20), and (c) PWMDE (K=50).

Figure 2:

The difference of KDE, PWMDE ($K=20$), or PWMDE ($K=50$) estimated curve from the CMDE for each μ_j with the fixed MCMC sample size $T = 1,000$.

Figure 3:

The estimated curves of joint marginal posterior density of μ_1 and μ_4 (left half) and their differences from the CMDE (right half).

Figure 4:
The chosen conditional working parameter space of μ_1 and μ_3 and its partition (small rectangles) for PWMDE.

Figure 5:

The estimated curves of joint marginal posterior density of μ_1 and μ_3 (left half) and their differences from the CMDE (right half). The red dashed line is the boundary $\mu_1 = \mu_3$ of the constraint $\mu_1 > \mu_3$.

Table 1:

Simulation results of PWMDE with $r = \sqrt{\chi_{3,0.95}^2}$ and IWMDE for estimating marginal t densities when $p = 3$ and $\nu - p = 2$.

	$\log \pi_{t_3}(0, 0, 0) = -3.911$			$\log \pi_{t_3}(1, 0, 0) = -4.326$		
	Mean	MCSE	RMSE	Mean	MCSE	RMSE
IWMDE	-3.910	0.018	0.018	-4.324	0.017	0.017
PWMDE						
K=5	-3.911	0.011	0.011	-4.325	0.010	0.010
K=10	-3.911	0.009	0.009	-4.325	0.008	0.008
K=15	-3.911	0.009	0.009	-4.325	0.008	0.008
K=20	-3.911	0.009	0.009	-4.325	0.008	0.008
	$\log \pi_{t_3}(0, 1, 0) = -4.421$			$\log \pi_{t_3}(0, 0, 1) = -4.326$		
	Mean	MCSE	RMSE	Mean	MCSE	RMSE
IWMDE	-4.419	0.016	0.017	-4.324	0.017	0.017
PWMDE						
K=5	-4.420	0.010	0.010	-4.325	0.011	0.011
K=10	-4.421	0.008	0.008	-4.325	0.010	0.010
K=15	-4.420	0.008	0.008	-4.325	0.010	0.010
K=20	-4.421	0.008	0.008	-4.325	0.009	0.009

Table 2:

Simulation results of PWMDE with $r = \sqrt{\chi_{3,0.95}^2}$ for estimating marginal t densities when $p = 3$ and $\nu - p$ increases to 10, 20, and 50, and $\log \pi_{\beta}(0, 0, 0) = -3.911$ is the true value.

	$\nu - p = 10$			$\nu - p = 20$			$\nu - p = 50$		
	Mean	MCSE	RMSE	Mean	MCSE	RMSE	Mean	MCSE	RMSE
IWMDE	-3.919	0.109	0.109	-3.942	0.246	0.247	-4.033	0.289	0.313
PWMDE									
K=5	-3.910	0.017	0.018	-3.910	0.020	0.020	-3.910	0.031	0.031
K=10	-3.910	0.014	0.014	-3.911	0.016	0.016	-3.911	0.022	0.022
K=15	-3.910	0.013	0.013	-3.911	0.015	0.015	-3.910	0.019	0.019
K=20	-3.910	0.013	0.013	-3.911	0.014	0.014	-3.910	0.018	0.018

Table 3:

Simulation results of PWMDE with $r = \sqrt{\chi^2_{10,0.95}}$ and IWMDE for estimating marginal t densities when p increases to 10 and $v-p = 10, 20$, and 50, $\log \pi_B(\theta_0) = -9.632$ is the true value, and θ_0 is the 10-dimensional vector of zeros.

	$v - p = 10$				$v - p = 20$				$v - p = 50$			
	Mean	MCSE	RMSE	Time	Mean	MCSE	RMSE	Time	Mean	MCSE	RMSE	Time
IWMDE	-9.837	0.565	0.601	8.9	-10.506	0.766	1.162	12.3	-11.342	0.784	1.881	14.5
PWMDE												
K=5	-9.635	0.082	0.082	15.9	-9.641	0.127	0.127	18.5	-9.658	0.199	0.200	25.1
K=10	-9.634	0.063	0.063	26.7	-9.638	0.105	0.105	31.1	-9.652	0.175	0.176	42.4
K=15	-9.632	0.055	0.055	37.4	-9.636	0.093	0.094	43.6	-9.654	0.162	0.164	59.6
K=20	-9.632	0.054	0.054	48.1	-9.634	0.086	0.086	56.1	-9.647	0.136	0.137	76.8
K=25	-9.632	0.049	0.049	61.3	-9.634	0.079	0.079	66.3	-9.645	0.133	0.134	93.9
K=30	-9.632	0.048	0.048	72.1	-9.632	0.082	0.082	78.4	-9.645	0.138	0.139	111.0
K=35	-9.633	0.049	0.049	82.9	-9.631	0.081	0.081	90.4	-9.640	0.127	0.127	128.3
K=40	-9.632	0.048	0.048	94.0	-9.633	0.078	0.078	102.1	-9.641	0.133	0.134	145.4

* Time is the average CPU time (sec.) for each simulation replicate.

Table 4:

Piecewise L_1 -distance for each density curve

Figure 2	KDE	PWMDE (K=20)	PWMDE (K=50)
μ_1	0.065	0.014	0.001
μ_2	0.069	0.002	0.006
μ_3	0.055	0.000	0.000
μ_4	0.082	0.029	0.001
Figure 3	KDE	PWMDE (K=400)	PWMDE (K=2500)
(μ_1, μ_4)	0.174	0.016	0.000
Figure 5	KDE	PWMDE (K=590)	PWMDE (K=3725)
(μ_1, μ_3)	0.141	0.003	0.002

Table 5:Marginal likelihood estimate under different precision τ_0

τ_0	$c(\widehat{D M})$	MCSE	τ_0	$c(\widehat{D M})$	MCSE
0.1	-766.398	0.009	20	-756.641	0.009
1	-760.791	0.007	30	-757.240	0.005
10	-756.690	0.011	40	-758.065	0.009
15	-756.516	0.011	50	-758.928	0.011

Table 6: HPD interval and Bayes Factor in log scale when $\tau_0 = 15$ and $\tau_0 = 0.1$

	$\tau_0 = 15$			$\tau_0 = 0.1$		
Variable	Mean	SD	95% HPD	Mean	SD	95% HPD
x_0	0.573	0.071	(0.434, 0.716)	0.581	0.070	(0.446, 0.718)
x_1	0.494	0.103	(0.303, 0.701)	0.512	0.103	(0.322, 0.723)
x_2	0.129	0.063	(0.013, 0.257)	0.128	0.062	(0.008, 0.253)
x_3	0.241	0.070	(0.108, 0.378)	0.239	0.069	(0.103, 0.373)
x_4	0.146	0.064	(0.021, 0.270)	0.145	0.063	(0.027, 0.273)
Model	$\log BF$	MCSE	rMCSE	$\log BF$	MCSE	rMCSE
PWMDE (K=5)			PWMDE (K=5)			
(x_1, x_2, x_3)	-1.033	0.006	0.59%	1.471	0.010	0.70%
(x_2, x_3, x_4)	-17.086	0.039	0.23%	-15.425	0.060	0.39%
(x_1, x_4)	-4.645	0.024	0.52%	0.330	0.034	10.23%
(x_2, x_3)	-19.848	0.061	0.31%	-15.791	0.069	0.44%
(x_1)	-6.263	0.035	0.56%	1.213	0.053	4.33%
(x_4)	-27.993	0.047	0.17%	-21.902	0.060	0.28%
PWMDE (K=10)			PWMDE (K=10)			
(x_1, x_2, x_3)	-1.034	0.006	0.60%	1.471	0.010	0.69%
(x_2, x_3, x_4)	-17.088	0.039	0.23%	-15.424	0.057	0.37%
(x_1, x_4)	-4.649	0.023	0.49%	0.327	0.031	9.51%
(x_2, x_3)	-19.844	0.059	0.30%	-15.783	0.065	0.41%
(x_1)	-6.268	0.032	0.51%	1.200	0.048	3.96%
(x_4)	-27.989	0.046	0.16%	-21.893	0.058	0.26%