Marginal Likelihood Estimation from a Single Run Metropolis-Hastings Output for Models with Local Independence

Silia Vitoratou, Ioannis Ntzoufras and Irini Moustaki

First Draft: 3 August 2011, This version: 24/10/2011

Marginal Likelihood Estimation from a Single Run Metropolis-Hastings Output for Models with Local Independence

Silia Vitoratou, Ioannis Ntzoufras and Irini Moustaki

Abstract

Chib's method for estimating the marginal likelihood required for model evaluation and comparison within the Bayesian paradigm, makes use of Gibbs sampling outputs from reduced Markov chain Monte Carlo (MCMC) runs for each parameter separately. More recently, the Chib-Jeliazkov method extended the application of the original approach to cases where the full conditional probabilities are not analytically available, estimating the marginal likelihood from multiple Metropolis-Hastings runs. This paper shows that a marginal likelihood estimation from a Metropolis-Hastings algorithm output can be computed from a single Markov chain Monte Carlo (MCMC) run of the full model under mild independence conditions. Marginal likelihood estimations are notoriously difficult to compute, and particularly so in high dimensional problems, such as those involving latent variable and/or random effect models. Thus, the applicability of the proposed estimation method is demonstrated by its application to simulated and real datasets, and we suggest that our findings can be further extended to data augmentation schemes leading to local (conditional) independence.

KEYWORDS: Generalised linear latent variable models, mixed effects models, bridge sampling, Monte Carlo error, Laplace-Metropolis estimator

1 INTRODUCTION

Bayesian model comparison via the Bayes factors, posterior model probabilities and odds (Kass and Raftery, 1995) requires the computation of the marginal likelihood given by:

$$f(\mathbf{Y}|m) = \int f(\mathbf{Y}|\boldsymbol{\theta}, m)\pi(\boldsymbol{\theta}|m)d\boldsymbol{\theta}, \tag{1}$$

where m stands for the hypothesized model, and $\pi(\boldsymbol{\theta}|m)$ is the density of the model specific parameter vector $\boldsymbol{\theta} \in \Omega \subset \mathbb{R}^p$ (m will be dropped hereafter for simplicity). The marginal likelihood often involves high dimensional integrals making the analytic computation infeasible except in some special cases. Several approximating methods have been proposed in the literature for estimating the marginal likelihood, including the importance sampling estimator (Newton and Raftery, 1994), the harmonic mean estimator (Kass and Raftery 1995, Raftery et al. 2007), Chib's estimator (Chib, 1995), the Bridge sampling estimator (Meng and Wong, 1996), the Metropolised Laplace estimator (Lewis and Raftery, 1997), Chib and Jeliazkov estimator (Chib and Jeliazkov, 2001), and, lately, the power posterior estimator (Friel and Pettit, 2008).

This paper focuses on the estimator proposed by Chib and Jeliazkov (2001), which is an extension of the marginal likelihood estimator of Chib (1995). Both methods are based on

the *candidate's identity* (Besag, 1989) explained below.

$$f(\mathbf{Y}) = \frac{f(\mathbf{Y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\pi(\boldsymbol{\theta}|\mathbf{Y})} \Leftrightarrow \log f(\mathbf{Y}) = \log f(\mathbf{Y}|\boldsymbol{\theta}) + \log \pi(\boldsymbol{\theta}) - \log \pi(\boldsymbol{\theta}|\mathbf{Y}). \tag{2}$$

From equation (2), the marginal likelihood depends on the posterior density of the model parameters $\pi(\boldsymbol{\theta}|\mathbf{Y})$. Since (2) holds for every point $\boldsymbol{\theta}$ of the parameter space, the posterior density can be approximated using a specific point $\boldsymbol{\theta}^*$. Following Chib (1995), let us suppose that the parameter space is split into p blocks. Then the posterior can be decomposed as:

$$\pi(\boldsymbol{\theta^*}|\mathbf{Y}) = \pi(\theta_1^*, \theta_2^*, \cdots, \theta_P^*|\mathbf{Y}) = \pi(\theta_1^*|\mathbf{Y})\pi(\theta_2^*|\mathbf{Y}, \theta_1^*) \cdots \pi(\theta_p^*|\mathbf{Y}, \theta_1^*, \theta_2^*, \cdots, \theta_{p-1}^*), \tag{3}$$

where each conditional density is called an ordinate. The integrated likelihood is calculated in a straightforward manner when (3) is analytically available. In the case when the posterior ordinates, $\pi(\theta_j^*|\mathbf{Y}, \theta_1^*, \theta_2^*, \cdots, \theta_{j-1}^*)$, are not available in a closed form, Chib (1995) presented an algorithm that uses the output from the Gibbs sampler to estimate them. In addition, Chib and Jeliazkov (2001) extended the method to deal with cases where the full conditional posterior distributions are not available and, therefore, a Metropolis–Hastings (MH) algorithm is used to generate posterior samples. Such an algorithm can be summarized by the following simple steps:

At every simulation step and for $j = 1, \ldots, p$

1. When θ_j is the current parameter value, propose θ_j' from a proposal centered at θ_j with density $q(\theta_j, \theta_j'|\boldsymbol{Y}, \theta_{\backslash j})$; where $\theta_{\backslash j}$ is the parameter vector $\boldsymbol{\theta}$ excluding θ_j.
2. Accept the proposed move with probability

$$a\big(\theta_j, \theta_j'|\boldsymbol{Y}, \theta_{\backslash j}\big) = \min\left\{1, \frac{f\big(\boldsymbol{Y}|\theta_j', \theta_{\backslash j}\big)\pi\big(\theta_j', \theta_{\backslash j}\big)q\big(\theta_j', \theta_j|\boldsymbol{Y}, \theta_{\backslash j}\big)}{f\big(\boldsymbol{Y}|\theta_j, \theta_{\backslash j}\big)\pi\big(\theta_j, \theta_{\backslash j}\big)q\big(\theta_j, \theta_j'|\boldsymbol{Y}, \theta_{\backslash j}\big)}\right\}. \tag{4}$$

Chib and Jeliazkov (2001) exploited the reversibility condition which assumes that the probability of sampling any pair (θ_j, θ_j') is the same regardless of the sequence by which the values are generated. Hence, the reversibility condition (with respect to θ_j^*) is written as:

$$K(\theta_j, \theta_j^*|\mathbf{Y}, \theta_{\backslash j})\pi(\theta_j|\mathbf{Y}, \theta_{\backslash j}) = K(\theta_j^*, \theta_j|\mathbf{Y}, \theta_{\backslash j})\pi(\theta_j^*|\mathbf{Y}, \theta_{\backslash j}). \tag{5}$$

with

$$K(\theta_j, \theta_j^*|\mathbf{Y}, \theta_{\backslash j}) = a(\theta_j, \theta_j^*|\mathbf{Y}, \theta_{\backslash j})q(\theta_j, \theta_j^*|\mathbf{Y}, \theta_{\backslash j}), \quad j = 1, \cdots, p, \tag{6}$$

denoting the transition probability of sampling θ_j^* given that θ_j has been already generated (that is, the MH sub-kernel). From (5), they obtained that each of the posterior ordinate $\pi(\theta_j^*|\mathbf{Y}, \theta_1^*, \cdots, \theta_{j-1}^*)$ appearing in (3) is equal to

$$CJ_j = \frac{E_1\Big\{a\big(\theta_j, \theta_j^*|\mathbf{Y}, \psi_{j-1}^*, \psi^{j+1}\big)\, q\big(\theta_j, \theta_j^*|\mathbf{Y}, \psi_{j-1}^*, \psi^{j+1}\big)\Big\}}{E_2\Big\{a\big(\theta_j^*, \theta_j|\mathbf{Y}, \psi_{j-1}^*, \psi^{j+1}\big)\Big\}}, \tag{7}$$

where $\psi_{j-1} = (\theta_1, \cdots, \theta_{j-1})$ and $\psi^{j+1} = (\theta_{j+1}, \cdots, \theta_p)$ for $j = 1, \ldots, p$ with ψ_0 and ψ^{p+1} referring to the empty sets. The expectations in the numerator and the denominator are with respect to $\pi\big(\theta_j, \psi^{j+1}|\mathbf{Y}, \psi^*_{j-1}\big)$ and $\pi\big(\psi^{j+1}|\mathbf{Y}, \psi^*_j\big)q\big(\theta_j, \theta^*_j|\psi^*_{j-1}, \psi^{j+1}\big)$.

A Monte Carlo estimator $\widehat{CJ}_j$ for each CJ_j can be obtained by replacing the expectations in (7) with their corresponding sample means from simulated samples. The final posterior estimator $(\widehat{CJ})$ is given by multiplying the estimators for each block.

Since the expectations in (7) are conditional on specific parameter points $\psi^*_{j-1} = (\theta^*_1, \cdots, \theta^*_{j-1})$ and $\psi^*_j = (\theta^*_1, \cdots, \theta^*_j)$, the corresponding Monte Carlo estimates cannot be obtained by the initial (full) MCMC run. Specifically, for a p dimensional parameter space, $p-1$ reduced runs are needed. Therefore, in high dimensional problems, the computational time can be frustratingly long. As an example, Chib and Jeliazkov (2001) address the issue of multiple latent variable blocks, where the latent vector $\mathbf{Z}$ may easily "*run into hundreds if not thousands*". For such cases, the authors provide a solution when at least one posterior ordinate $\pi(\theta_j|\mathbf{Y}, \psi_{j-1})$ is analytically available. To be more specific, they rewrite the corresponding ordinate $\pi(\theta^*_j|\mathbf{Y}, \psi^*_{j-1})$ as an integral involving the latent variable quantities $\mathbf{Z}$, that is:

$$\begin{aligned} \pi(\theta^*_j|\mathbf{Y}, \psi^*_{j-1}) &= \int \pi(\theta^*_j, \mathbf{Z}|\mathbf{Y}, \psi^*_{j-1}) d\mathbf{Z} \\ &= \int \pi(\theta^*_j|\mathbf{Y}, \psi^*_{j-1}, \mathbf{Z}) \pi(\mathbf{Z}|\mathbf{Y}, \psi^*_{j-1}) d\mathbf{Z}. \end{aligned} \tag{8}$$

Therefore $p-1$ ordinates are approximated via (7) as usual, while the integral (8) is approximated as the average of $\pi(\theta^*_j|\mathbf{Y}, \psi^*_{j-1})$ with respect to $\pi(\mathbf{Z}|\mathbf{Y}, \psi^*_{j-1})$. This approach is actually equivalent to marginalizing out the latent variables as nuisance parameters.

In practice, even one posterior ordinate of the above type is rarely available in closed form. On the other hand, the dimensionality of $\mathbf{Z}$ makes the implementation of the Chib and Jeliaskov (CJ) estimator unattractive in its original form since it requires a large number of reduced (or nested) MCMC runs. In this article, we propose a simplification of the CJ estimator for computing the marginal likelihood from a *single* Metropolis-Hastings run for models with conditional (local) independence such as the generalised linear latent variable models (GLLVM) and mixed effect models. The proposed modification avoids the multiple runs, making the Chib and Jeliazkov (2001) method applicable in such cases.

The rest of the article is organized as follows. In Section 2, we describe the proposed method after introducing some preliminary definitions and results. Section 3 describes the relation of our proposed modified estimator with the one obtained using bridge sampling. Section 4.1 describes the implementation in generalised latent variable models including illustrations on simulated and real datasets. Concluding remarks are provided in the closing section of this article.

2 ESTIMATING THE MARGINAL LIKELIHOOD IN MODELS WITH LOCAL INDEPENDENCE

2.1 Definitions and Properties

In this section, we discuss two necessary model properties, that lead to the simplification of the CJ estimator. Let us first define our model structure and the corresponding notation.

We focus on models which can be defined with a likelihood of the following structure:

$$f\big(\boldsymbol{Y}|\boldsymbol{\Theta}=(\theta_0,\theta_1,\ldots,\theta_p),\boldsymbol{L}\big)=f\big(\boldsymbol{Y}|\boldsymbol{\theta}=(\theta_1,\ldots,\theta_p),\mathbf{Z}=(\theta_0,\boldsymbol{L})\big)\ \text{with}\ \boldsymbol{Y}=(\boldsymbol{Y}_1,\ldots,\boldsymbol{Y}_p), \tag{9}$$

where

- $\boldsymbol{Y}$ is a $n \times p$ data array of n observations and p observed items (or variables),
- $\mathbf{Y}_j$ is the $n \times 1$ vector with the data values for item j,
- $\boldsymbol{L}$ is the $k \times n$ matrix of the latent variables,
- $\boldsymbol{\Theta}$ is the whole parameters $(k+1) \times p$ vector,
- θ_0 is the set of parameters which is common across different variables,
- θ_j for $j = 1, \ldots, p$ are the variable specific parameters (linked to $\mathbf{Y}_j$ only)

In the above model formulation, the pair of parameters and the latent variables $(\boldsymbol{\Theta}, \boldsymbol{L})$ are substituted by a new pair $(\boldsymbol{\theta}, \mathbf{Z})$ with $\boldsymbol{\theta}$ being the variable specific parameters (that is decomposed later to parameters which are related to one variable at a time) and $\mathbf{Z}$ being the set of parameters and/or latent variables which are common and shared across different variables or items. In latent variable models, parameters shared across different items do not exist unless equality constraints are imposed. Hence $\mathbf{Z}$ solely refers to latent variables $\boldsymbol{L}$. The local independence assumption is common in such models and it is described in Definition 2.1 which follows.

Definition 2.1 *The model presented in (9) embodies the* **local independence** *assumption if the observed response variables (or items) are independent of each other conditional on the latent variable(s):*

$$f(\boldsymbol{Y}|\boldsymbol{\theta},\boldsymbol{Z})=\prod_{i=1}^{p} f(\boldsymbol{Y}_j|\theta_j,\boldsymbol{Z})\,. \tag{10}$$

The local independence assumption implies that correlations among the items are induced solely by $\mathbf{Z}$, that is, the latent variables and the common parameters. We now show that this property can be extended to the posterior distributions when *prior local independence* exists. This notion is introduced in Definition 2.2 which follows.

Definition 2.2 *For the model presented in (9), a set of variable specific parameters* $\boldsymbol{\theta}$ *is defined to be* **a-priori locally independent** *if they are a-priori independent conditionally on* $\boldsymbol{Z}$*. Therefore the prior structure will satisfy the following equation:*

$$f(\boldsymbol{\theta}|\boldsymbol{Z}) = \prod_{j=1}^{p} f(\theta_j|\boldsymbol{Z})\,. \tag{11}$$

Similarly we can introduce the *posterior local independence* by the following definition.

Definition 2.3 *For the model presented in (9), a set of variable specific parameters* $\boldsymbol{\theta}$ *is defined to be* **a-posteriori locally independent** *if they are a-posteriori independent conditionally upon* $\boldsymbol{Z}$*. Therefore the posterior structure will satisfy the following equation*

$$f(\boldsymbol{\theta}|\boldsymbol{Y},\boldsymbol{Z}) = \prod_{j=1}^{p} f(\theta_j|\boldsymbol{Y}_j,\boldsymbol{Z})\,. \tag{12}$$

Lemma 2.1 *For any model with structure (9) for which the assumptions of*

a) local independence and

b) prior local independence

hold, then the property of posterior local independence also holds.

Proof

$$\begin{aligned}\pi(\boldsymbol{\theta}|\mathbf{Y},\mathbf{Z}) &= \frac{f(\mathbf{Y}|\boldsymbol{\theta},\mathbf{Z})\pi(\boldsymbol{\theta}|\mathbf{Z})}{\int f(\mathbf{Y}|\boldsymbol{\theta},\mathbf{Z})\pi(\boldsymbol{\theta}|\mathbf{Z})d\boldsymbol{\theta}} = \frac{\prod_{j=1}^{p} f(\mathbf{Y}_j|\theta_j,\mathbf{Z})\prod_{j=1}^{p}\pi(\theta_j|\mathbf{Z})}{\int\cdots\int\prod_{j=1}^{p} f(\mathbf{Y}_j|\theta_j,\mathbf{Z})\prod_{j=1}^{p}\pi(\theta_j|\mathbf{Z})d\theta_1\cdots d\theta_p} \\ &= \prod_{j=1}^{p}\frac{f(\mathbf{Y}_j|\theta_j,\mathbf{Z})\pi(\theta_j|\mathbf{Z})}{\int f(\mathbf{Y}_j|\theta_j,\mathbf{Z})\pi(\theta_j|\mathbf{Z})d\theta_j} = \prod_{j=1}^{p}\pi(\theta_j|\mathbf{Y}_j,\mathbf{Z})\,. \end{aligned} \tag{13}$$

□

The second necessary condition is related to the acceptance probability of the MH algorithm which, for any model with structure (9), is described by the following steps:

1. for $j = 1,\ldots,p$
   (a) When θ_j is the current parameter value, propose θ_j' from a proposal with density $q(\theta_j,\theta_j'|\boldsymbol{Y},\theta_{\setminus j},\mathbf{Z})$; where $\theta_{\setminus j}$ is $\boldsymbol{\theta}$ excluding θ_j.
   (b) Accept the proposed move with probability

$$a\big(\theta_j,\theta_j'|\boldsymbol{Y},\theta_{\setminus j},\mathbf{Z}\big) = \min\left\{1, \frac{\pi\big(\theta_j'|\boldsymbol{Y},\theta_{\setminus j},\mathbf{Z}\big)q\big(\theta_j',\theta_j|\boldsymbol{Y},\theta_{\setminus j},\mathbf{Z}\big)}{\pi\big(\theta_j|\boldsymbol{Y},\theta_{\setminus j},\mathbf{Z}\big)q\big(\theta_j,\theta_j'|\boldsymbol{Y},\theta_{\setminus j},\mathbf{Z}\big)}\right\}. \tag{14}$$

2. Sample $\mathbf{Z}$ from $f(\mathbf{Z}|\boldsymbol{Y}, \boldsymbol{\theta})$ using any sampling scheme.

Lemma 2.2 *For any model with structure (9), the acceptance probability (14) of the MH algorithm described above simplifies to:*

$$a(\theta_j, \theta_j'|\,\boldsymbol{Y}, \theta_{\setminus j}, \boldsymbol{Z}) = a(\theta_j, \theta_j'|\,\boldsymbol{Y}, \boldsymbol{Z}) = \min\left\{1, \frac{f(\,\boldsymbol{Y}_j|\theta_j', \boldsymbol{Z})\pi(\theta_j'|\,\boldsymbol{Z})\, q(\theta_j', \theta_j|\,\boldsymbol{Y}, \boldsymbol{Z})}{f(\,\boldsymbol{Y}_j|\theta_j, \boldsymbol{Z})\pi(\theta_j|\,\boldsymbol{Z})\, q(\theta_j, \theta_j'|\,\boldsymbol{Y}, \boldsymbol{Z})}\right\}. \qquad (15)$$

when

a) the posterior local independence property holds and

b) each θ_j is proposed independently from $\theta_{\setminus j}$.

The above lemma is straightforward to prove by implementing conditions (a) and (b) on the acceptance probability (14).

Lemmas 2.1 and 2.2 also hold in the special case of prior independence between θ_j and $\mathbf{Z}$, that is $\pi(\boldsymbol{\theta}, \mathbf{Z}) = \prod_{j=1}^{p} \pi(\theta_j)\pi(\mathbf{Z})$, and/or when common parameters across different variables (that is θ_0) do not exist which refers to the usual factor models. In Section 2.2, we illustrate the simplification of the CJ estimator of the marginal likelihood using the lemmas of this section.

2.2 CJ Estimator in a Single Run for Models with Conditional Independence

Following similar arguments as in Chib and Jeliazkov (2001), we start from the reversibility of the sub-kernel at any point θ_j^*. If we multiply both sides of (5) with $\pi(\theta_{\setminus j}|\mathbf{Y}, \mathbf{Z})$ then,

$$K(\theta_j, \theta_j^*|\mathbf{Y})\,\pi(\theta_j, \theta_{\setminus j}|\mathbf{Y}, \mathbf{Z}) = K(\theta_j^*, \theta_j|\mathbf{Y})\,\pi(\theta_j^*, \theta_{\setminus j}|\mathbf{Y}, \mathbf{Z}),$$

which results in

$$K(\theta_j, \theta_j^*|\mathbf{Y})\,\pi(\theta_j|\mathbf{Y}, \mathbf{Z}) \prod_{j'=1, j'\neq j}^{p} \pi(\theta_{j'}|\mathbf{Z}, \mathbf{Y}) = K(\theta_j^*, \theta_j|\mathbf{Y})\,\pi(\theta_j^*|\mathbf{Y}, \mathbf{Z}) \prod_{j'=1, j'\neq j}^{p} \pi(\theta_{j'}|\mathbf{Z}, \mathbf{Y})$$

due to the posterior local independence assumption.

By integrating over θ_j,

$$\int K(\theta_j, \theta_j^*|\mathbf{Y})\,\pi(\theta_j|\mathbf{Y}, \mathbf{Z})d\theta_j = \int K(\theta_j^*, \theta_j|\mathbf{Y})\,\pi(\theta_j^*|\mathbf{Y}, \mathbf{Z})d\theta_j,$$

and solving with respect to $\pi(\theta_j^*|\mathbf{Y}, \mathbf{Z})$ we get:

$$CJ_j^I = \pi(\theta_j^*|\mathbf{Y}, \mathbf{Z}) = \frac{\int K(\theta_j, \theta_j^*|\mathbf{Y})\,\pi(\theta_j|\mathbf{Y}, \mathbf{Z})d\theta_j}{\int K(\theta_j^*, \theta_j|\mathbf{Y})d\theta_j}\ . \qquad (16)$$

The modified expression for the posterior $\pi(\boldsymbol{\theta}|\mathbf{Y})$ is then given by multiplying CJ_j^I over all p blocks and integrate out the latent vector:

$$\begin{aligned} CJ^I = \int \prod_{j=1}^{p} \pi(\theta_j|\mathbf{Y}_j,\mathbf{Z})\pi(\mathbf{Z}|\mathbf{Y})d\mathbf{Z} &= \int \prod_{j=1}^{p} \left[\frac{\int K(\theta_j,\theta_j^*|\mathbf{Y})\,\pi(\theta_j|\mathbf{Y},\mathbf{Z})d\theta_j}{\int K(\theta_j^*,\theta_j|\mathbf{Y})d\theta_j}\right] \pi(\mathbf{Z}|\mathbf{Y})\,d\mathbf{Z} \\ &= \int \left[\frac{\prod_{j=1}^{p} K(\theta_j,\theta_j^*|\mathbf{Y})}{\int \prod_{j=1}^{p} K(\theta_j^*,\theta_j|\mathbf{Y})d\theta_j}\right] \pi(\boldsymbol{\theta},\mathbf{Z}|\mathbf{Y})\,d(\boldsymbol{\theta},\mathbf{Z}) \\ &= E_{\boldsymbol{\theta},\mathbf{Z}|\mathbf{Y}}\left[\frac{\prod_{j=1}^{p} a\left(\theta_j,\theta_j^*|\mathbf{Y},\mathbf{Z}\right) q\left(\theta_j,\theta_j^*|\mathbf{Y},\mathbf{Z}\right)}{E_q\left[\prod_{j=1}^{p} a\left(\theta_j^*,\theta_j|\mathbf{Y},\mathbf{Z}\right)\right]}\right], \end{aligned} \tag{17}$$

where E_q is the expectation with respect to the proposal density $q(\boldsymbol{\theta}^*,\boldsymbol{\theta}|\mathbf{Y},\mathbf{Z}) = \prod_{j=1}^{p} q\big(\theta_j^*,\theta_j|\mathbf{Y},\mathbf{Z}\big)$. Due to Lemma 2.2, the expectation in the denominator of (17) can be replaced by the product of p expectations that involve densities of much lower dimension and thus can be estimated with a reduced Monte Carlo error as illustrated by Vitoratou et al. (2011). Thus (17) can be estimated from:

$$\widehat{CJ}^{\,I} = \frac{1}{R}\sum_{r=1}^{R}\left[\frac{\prod_{j=1}^{p} a\big(\theta_j^{(r)},\theta_j^*|\mathbf{Y},\mathbf{Z}^{(r)}\big) q\big(\theta_j^{(r)},\theta_j^*|\mathbf{Y},\mathbf{Z}\big)}{\prod_{j=1}^{p}\left[\frac{1}{M}\sum_{m=1}^{M} a\big(\theta_j^{*(m)},\theta_j|\mathbf{Y},\mathbf{Z}^{(r)}\big)\right]}\right]. \tag{18}$$

The sample $\big\{\theta_1^{(r)},\theta_2^{(r)},\cdots,\theta_p^{(r)},\mathbf{Z}^{(r)}\big\}_{r=1}^{R}$ comes from the joint posterior of $(\boldsymbol{\theta},\mathbf{Z})$ which is available from a full MCMC run. For each sampled set of latent and parameter values $\big(\boldsymbol{\theta}^{(r)},\mathbf{Z}^{(r)}\big)$, $r = 1,...,R$, points $\{\theta_j^{(m)}\}_{m=1}^{M}$ are generated from the proposal density $q(\theta_j,\theta_j^*|\mathbf{Y},\mathbf{Z})$. These values are used to compute the expectation in the denominator of (17). From (18), it is straightforward to see that a single MCMC run from the posterior of the model under study is required to compute the *independence* estimator $\widehat{CJ}^{\mathrm{I}}$.

To sum up, in this section we have introduced a modification of the original CJ estimator of the marginal likelihood, that is, the independence CJ estimator, $\widehat{CJ}^{\mathrm{I}}$. This estimator can be easily implemented in high dimensional models involving latent variables and random effects even when none of the posterior ordinates is available analytically as it is required in Chib and Jeliazkov (2001). The proposed estimator is based on Lemma 2.2 that requires the existence of posterior local independence and the generation of a posterior sample using an MH algorithm with independent proposals. The assumption of posterior local independence is mild since, due to Lemma 2.1, it is true for any model which assumes local independence when a prior distribution with independence structure (11) is adopted. The local independence assumption is fundamental for a wide class of models such as latent variable and random effect models. Finally, the prior local independence induced by (11) seems to be a reasonable assumption since such models require local independence for the parameters.

3 CJ Estimator and Bridge Sampling

The CJ estimator can be viewed as a special case of the bridge sampling estimator (Meng and Wong, 1996). It was originally addressed for problems that require the approximation of the ratio r of two normalizing constants c_1 and c_2 for any two densities $p_j(\boldsymbol{\theta}) = g_j(\boldsymbol{\theta})/c_j$, $(i = 1, 2)$ with supports Ω_1 and Ω_2 respectively. For any arbitrary function $h(\boldsymbol{\theta})$ defined in $\Omega_1 \cap \Omega_2$ such as that:

$$0 < \left| \int_{\Omega_1 \cap \Omega_2} h(\boldsymbol{\theta}) p_1(\boldsymbol{\theta}) p_2(\boldsymbol{\theta}) d\boldsymbol{\theta} \right| < \infty, \tag{19}$$

the ratio $r = \frac{c_1}{c_2}$ is given by the identity:

$$r = \frac{\int g_1(\boldsymbol{\theta}) h(\boldsymbol{\theta}) p_2(\boldsymbol{\theta}) d\boldsymbol{\theta}}{\int g_2(\boldsymbol{\theta}) h(\boldsymbol{\theta}) p_1(\boldsymbol{\theta}) d\boldsymbol{\theta}} = \frac{E_{p_2}\left[g_1(\boldsymbol{\theta}) h(\boldsymbol{\theta})\right]}{E_{p_1}\left[g_2(\boldsymbol{\theta}) h(\boldsymbol{\theta})\right]}. \tag{20}$$

The above identity can be used to construct a Monte Carlo estimate of r using two random samples from p_1 and p_2.

Mira and Nicholls (2004) showed that the CJ estimator can be seen as a bridge sampling estimator, by appropriately defining the densities p_1, p_2 and h, for each block j. Similarly, the independence estimator $\widehat{CJ}^{\mathrm{I}}$ can be written as a bridge sampling estimator when:

$$\begin{array}{ll} c_1 = 1, & c_2 = 1/\pi(\theta_j^*|\mathbf{Y}, \mathbf{Z}) \\ g_1 = \pi(\theta_j|\mathbf{Y}, \mathbf{Z}), & g_2 = q(\theta_j^*, \theta_j)/\pi(\theta_j^*|\mathbf{Y}, \mathbf{Z}) \\ p_1 = \pi(\theta_j|\mathbf{Y}, \mathbf{Z}), & p_2 = q(\theta_j^*, \theta_j), \qquad h = \dfrac{a(\theta_j^*, \theta_j|\mathbf{Y}, \mathbf{Z})}{\pi(\theta_j|\mathbf{Y}, \mathbf{Z})} \end{array}.$$

By substituting the above quantities in (20) and employing the *detailed balance relation* for q, a and π (see equation 5) in the nominator we obtain:

$$BS_j = \frac{\int \pi(\theta_j^*|\mathbf{Y}, \mathbf{Z}) q(\theta_j^*, \theta_j) \dfrac{a(\theta_j^*, \theta_j|\mathbf{Y}, \mathbf{Z})}{\pi(\theta_j|\mathbf{Y}, \mathbf{Z})} \pi(\theta_j|\mathbf{Y}, \mathbf{Z}) d\theta_j}{\int \pi(\theta_j|\mathbf{Y}, \mathbf{Z}) \dfrac{a(\theta_j^*, \theta_j|\mathbf{Y}, \mathbf{Z})}{\pi(\theta_j|\mathbf{Y}, \mathbf{Z})} q(\theta_j^*, \theta_j) d\theta_j} \tag{21}$$

$$= \frac{\int a(\theta_j, \theta_j^*|\mathbf{Y}, \mathbf{Z}) q(\theta_j, \theta_j^*) \pi(\theta_j|\mathbf{Y}, \mathbf{Z}) d\theta_j}{\int a(\theta_j^*, \theta_j|\mathbf{Y}, \mathbf{Z}) q(\theta_j^*, \theta_j) d\theta_j} = CJ_j^I \tag{22}$$

By marginalizing out the latent vector as in (17), we obtain the modified CJ^{I} expression.

4 IMPLEMENTATION OF THE CJ ESTIMATOR IN GLLVM

4.1 Model formulation

The Generalised Linear Latent Variable Model (GLLVM) (Moustaki and Knott, 2000) consists of three components: (a) the multivariate random component $\widetilde{Y} = (Y_1, Y_2, \ldots, Y_p)$ of

the response variables, (b) the linear predictor denoted by η_j and (c) the link function $\upsilon(\cdot)$, which connects the previous two components. Hence, a GLLVM can be summarized as:

$$Y_j|\mathbf{Z} \sim ExpF, \quad \eta_j = \alpha_j + \sum_{\ell=1}^{k} \beta_{j\ell} Z_\ell, \quad \text{and} \quad \upsilon_j\Big(\mu_j(\mathbf{Z})\Big) = \eta_j \tag{23}$$

for $j = 1, \ldots p$; where $ExpF$ is a member of the exponential family and $\mu_j(\mathbf{Z}) = \mathrm{E}(\mathrm{Y}_j|\mathbf{Z})$. Finally, a multivariate distribution $\pi(\mathbf{Z})$ needs to be specified for the latent variables, which is usually assumed to be a *standard normal* distribution.

It is further assumed that, given $\mathbf{Z}$ and $\boldsymbol{\theta}$, the responses to the observed variables are independent (*local independence* assumption) resulting in

$$f(\boldsymbol{Y}|\mathbf{Z}, \boldsymbol{\theta}) = \prod_{i=1}^{n} \prod_{j=1}^{p} f(Y_{ij}|\theta_j, \mathbf{Z}_i)\,, \tag{24}$$

where $\boldsymbol{Y}$ is a $n \times p$ observed data matrix with elements Y_{ij} denoting the response of subject i to item j and $\mathbf{Z}_i$ are the subject specific values of the latent variables $\mathbf{Z}$, $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \boldsymbol{\beta})$, $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_p)$, $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_p)$, $\boldsymbol{\beta}_1 = (\beta_{11}, \ldots, \beta_{1k})$ and $\theta_j = (\alpha_j, \boldsymbol{\beta}_j)$.

The model likelihood is obtained by marginalizing out the latent variables, and is given by

$$f(\boldsymbol{Y}|\boldsymbol{\theta}) = \prod_{i=1}^{n} f(\mathbf{Y}_i|\boldsymbol{\theta}) = \prod_{i=1}^{n} \int f(\mathbf{Y}_i|\boldsymbol{\theta}, \mathbf{Z}_i)\, \pi(\mathbf{Z}_i)\, d\mathbf{Z}_i\,, \tag{25}$$

where $\mathbf{Y}_i$ is the vector of responses for subject i (i.e. the i-th row of matrix $\boldsymbol{Y}$). The integrals with respect to the subject specific latent variables $\mathbf{Z}_i$ in (25) can be approximated with fixed Gauss-Hermite quadrature points (used to calculate each $f(\mathbf{Y}_i|\boldsymbol{\theta})$ in equation 25). Other more accurate approximations can be also used, such as the adaptive quadrature points (Rabe-Hesketh et al. 2005, Schilling and Bock 2005) or Laplace approximations (Huber et al., 2004).

The corresponding marginal likelihood $f(\mathbf{Y}) = \int f(\boldsymbol{Y}|\boldsymbol{\theta}, \mathbf{Z})\, \pi(\boldsymbol{\theta}, \mathbf{Z})\, d(\boldsymbol{\theta}, \mathbf{Z})$ needed for computing the Bayes factor is a highly dimensional integral, which is not available analytically. The results of Section 2.1 can be implemented here since the local independence assumption is a fundamental component of the model formulation, as described in (23).

In our simulations we used binary items (latent trait models) that are a special case of the GLLVM discussed in Moustaki and Knott (2000). The logit link is used for the response probabilities, giving:

$$\mathrm{logit}\Big[E\big(Y_j|\mathbf{Z}\big)\Big] = \log \frac{P(Y_j = 1 \mid \mathbf{Z})}{(1 - P(Y_j = 1 \mid \mathbf{Z})} = \alpha_j + \sum_{\ell=1}^{k} \beta_{j\ell} Z_\ell, \quad j = 1, \ldots, p. \tag{26}$$

4.2 Parametrization and Prior Specification

Patz and Junker (1999) implemented Bayesian inference using the full model likelihood (24) for latent variable models with binary items. Model identification is crucial under the presence

of k latent variables. Here, we implement the Cholesky decomposition on the loadings $\mathbf{B} = \left(\beta_{j\ell}\right)$ with $j = 1, \ldots, p, \ell = 1, \ldots, k$ as described by Lopes and West (2004) and Dunson (2006). Therefore, we set $\beta_{j\ell} = 0$ for all $j < \ell$ and $\beta_{jj} > 0$. Moreover, we use a prior based on the ideas presented by Ntzoufras et al. (2000) and further explored in the context of generalised linear models by Fouskakis et al. (2009, equation 6). For GLLVMs with binary responses, this prior corresponds to a $N(0, 4)$ for all non-constrained loadings and for all α_j. For the diagonal elements of $\boldsymbol{B}$, we use a standardized normal distribution as a prior for each $\log \beta_{jj}$ inducing prior a standard deviation for β_{jj} approximately equal to 2, in analogy with the rest non-zero elements of $\boldsymbol{B}$. Hence, the structure of $\boldsymbol{B}$ can be summarized as follows

$$\beta_{j\ell} = \begin{cases} 0 & \text{if} \quad j < \ell \\ LN(0,1) & \text{if} \quad j = \ell \\ N(0,4) & \text{if} \quad j > \ell \end{cases}$$

where $Y \sim LN(\mu, \sigma^2)$ is the log-normal distribution with the mean and the variance of $\log Y$ being equal to μ and σ^2, respectively. Finally, latent variables are assumed to be a-priori distributed as standard normal distributions i.e. $Z_\ell \sim N(0, 1)$ for all subjects.

4.3 Monte Carlo Error and Laplace-Metropolis Estimator

The Monte Carlo error (MCE) of the $\widehat{CJ^{\mathrm{I}}}$ was estimated using the method of batch means (Schmeiser 1982, Bratley et al. 1987). The simulated sample was divided into 30 batches and the marginal log-likelihood was approximated via $\widehat{CJ^{\mathrm{I}}}$ at each batch. The mean over all batches, denoted by $\widehat{CJ^{\mathrm{I}}_{bm}}$, is referred to as the batch mean estimator, while the the standard deviation of the log-marginal likelihood estimator over the different batches is considered as its MCE estimate. The same procedure was repeated for the $\widehat{CJ^{\mathrm{I}}}$, which were estimated using three alternative measures of central location of the posterior distribution (the componentwise posterior mean, median and mode) as θ^*.

Moreover, the Laplace-Metropolis estimator ($\widehat{LM}$) proposed by Lewis and Raftery (1997) was used as benchmark method. The Laplace-Metropolis method was implemented on the posterior $\pi(\boldsymbol{\theta}|\boldsymbol{Y})$, therefore, the vector of the latent variables $\mathbf{Z}$ were marginalized out. The normal approximation used in the Laplace method was applied to the original parameters for all α_j and $\beta_{j\ell}$, with $j < \ell$, and on the $\log \beta_{jj}$ for $j = 1, \ldots, k$ for the diagonal loadings. For the latter, we have used the logarithms instead of the original parameters in order to avoid asymmetries caused by their positivity constraint and, by this way, to achieve a well behaved approximation of the marginal likelihood.

4.4 Tuning M and R

We initially use a dataset generated from a one-factor model with 4 binary items and 400 individuals ($p = 4$, $N = 400$ and $k = 1$ respectively, that is 408 unknown parameters). We use this rather restricted example in order to examine the convergence of the estimator as a function of the number of M and R values generated from the proposal and the posterior densities, respectively. Specifically, 300,000 posterior observations were generated after discarding additional 10,000 iterations as a burn in period from a Metropolis-Hastings, within

a Gibbs, algorithm. A thinning interval of 10 iterations was additionally considered in order to diminish autocorrelations, leaving a total of 30,000 values available for posterior analysis. All simulations were conducted using R version 2.12 on a quad core i5 Central Processor Unit (CPU), at 3.2GHz and with 4GB of RAM.

Before dividing the simulated sample into batches, we have graphically examined the convergence of the estimator by changing

a) M, that is, the number of points generated from the proposal density $q(\boldsymbol{\theta}, \boldsymbol{\theta}^*|\mathbf{Y}, \mathbf{Z})$ used for the estimation of the denominator in (18),

b) R, that is, the number of points generated from the posterior $\pi(\boldsymbol{\theta}, \mathbf{Z}|Y)$ that are required for the estimation of $\widehat{CJ}^{\mathrm{I}}$ within each batch.

We initially focused on (a), with M ranging from 100 to 2000, and kept R fixed at 1000 iterations. Figure 1(a) illustrates that all versions of $\widehat{CJ}^{\mathrm{I}}$ were stabilized up to a decimal point, even for $M \geq 40$. Time increased linearly, with M varying from 0.5 to 4.7 mins, which is approximately one minute increment per 25 generated values.

Regarding (b), the ergodic estimator was computed with R taking values from 100 to 2000 and $M = 50$, which seem more than sufficient according to Figure 1(a). The ergodic estimators of all versions of $\widehat{CJ}^{\mathrm{I}}$ for each selected R are depicted in Figure 1(b). The estimates were close and stable for $R \geq 500$. The CPU time was also increased linearly from 0.5 to 9 mins at the cost of half a minute per 100 additional iterations.

Based on Figure 1, we proceeded with thirty batches of size $R = 1000$ and $M = 50$ to ensure convergence of the estimates. Figure 2 presents the marginal likelihood estimates based on CJ and LM using the posterior mean, median and mode as points of central location. When using the posterior mean, $\widehat{LM}$ was found to be equal to -977.76, while $\widehat{CJ}^{\mathrm{I}}_{bm}$ was equal to -977.73, with the estimated MCE being equal to 0.026. The estimators are quite robust, regardless of the choice of the posterior point of central location. Specifically, the $\widehat{LM}$ was -977.65 at the median and -977.71 at the mode. Similarly, the $\widehat{CJ}^{\mathrm{I}}_{bm}$ was -977.77 at the median and -977.75 at the mode, with equivalent MCEs (0.020 and 0.022 respectively).

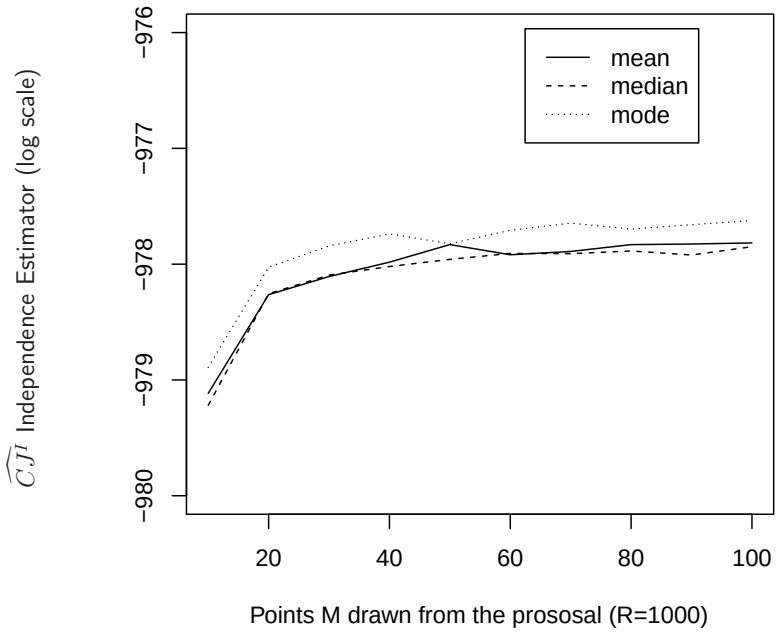
In the next section we proceed with more realistic illustrations, using both simulated and real data sets. In all the examples which follow, the same tuning procedure was followed but it is not reported for brevity.

4.5 Computation of Bayes Factors: Simulated Examples

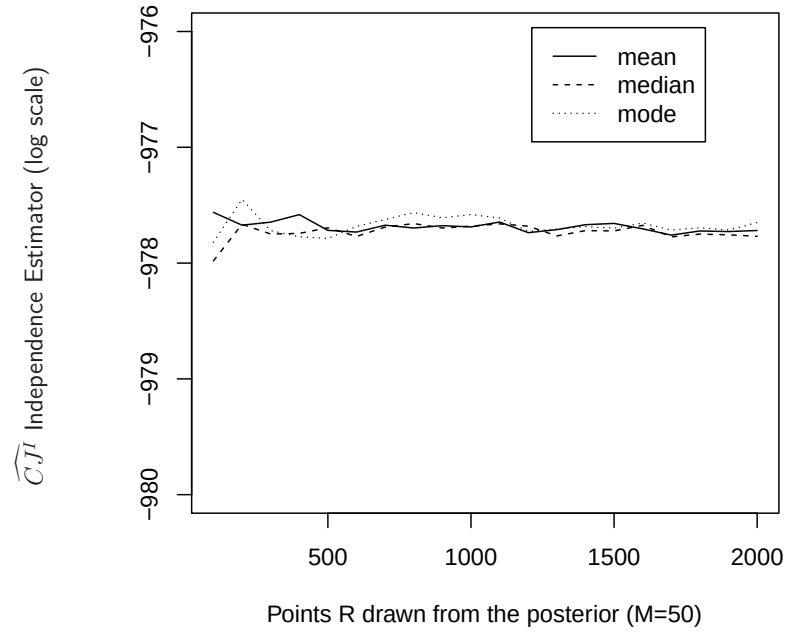
Here we demonstrate the performance of the CJ estimator using the output from a single run of a Metropolis-Hastings algorithm in three simulated datasets of larger size, allowing, in addition, for the models to be fitted with multiple factors of higher dimension. We consider the datasets with the following setups:

a) $N = 600$ observations with $p = 6$ items generated from a $k = 1$ factor model

b) $N = 600$ observations with $p = 6$ items generated from a $k = 2$ factor model

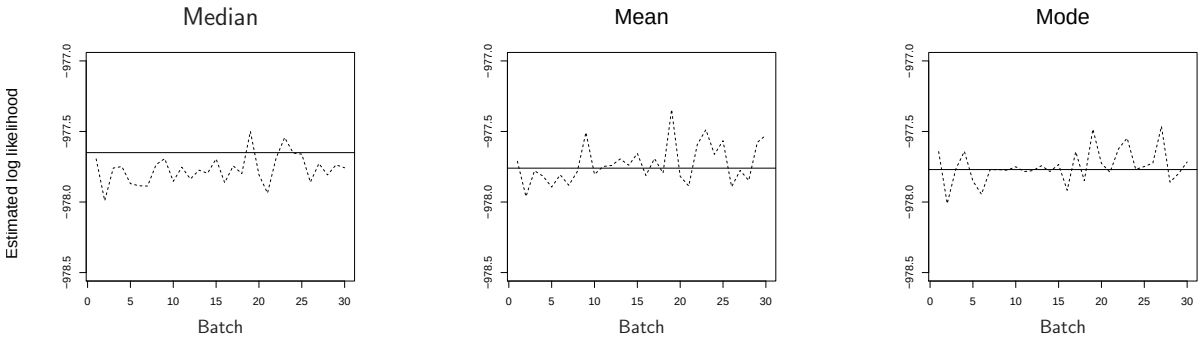


(a) Sensitivity of $\widehat{CJ^I}$ on different M with $R = 1000$.



(b) Sensitivity of $\widehat{CJ^I}$ on different R with $M = 50$.

Figure 1: Ergodic $\widehat{CJ^I}$ using three posterior measures of central location (mean, median and mode) for different M (number of values generated from the proposal) and for different R (number of MCMC iterations); p=4 items, $N = 400$ individuals and k=1 latent factor.



(a) (b) (c)

Figure 2: $\widehat{CJ^I}$ (dotted line) over 30 batches of size R=1000 compared with $\widehat{LP}$ (solid line) estimated from an MCMC output of 30,000 iterations using the posterior median, mean or mode as measures of central location; p=4 items, $N = 400$ individuals and $k = 1$ factor.

c) $N = 800$ observations with $p = 7$ items generated from a $k = 3$ factor model

All model parameters were selected randomly from a uniform distribution, $U(-2, 2)$. The number of unknown parameters for the posterior ordinate in (18) is equal to $k(p + N) + p$, corresponding to 606, 1218 and 2428 parameters, respectively, for each of the three situations described above. Models that either overestimate or underestimate k were also considered, this time evaluating the Bayes factor in favour of the true generating model. Using the same procedure as in Section 4.4, we have concluded that it is sufficient to select 30 batches of 1000, 2000 and 3000 iterations for the one, two and three-factor models, respectively. All estimators were evaluated at the componentwise posterior median (that is, θ^*=posterior median).

The $\widehat{LM}$ is also reported in Table 1 as a gold standard. It is computed over 30,000 iterations, while $\widehat{CJ}^{\mathrm{I}}$ refers to the first batch. The batch mean estimator and the corresponding error were calculated as described in Section 4.3. Under all scenarios, the two methods provide close estimates of the marginal likelihood. The Monte Carlo error of the $\widehat{CJ}^{\mathrm{I}}$ estimator is fairly small but naturally gets higher as the number of unknown parameters in the posterior ordinate increase for a fixed number of iterations. Nevertheless, this Monte Carlo error can be efficiently reduced by increasing the number of MCMC iterations.

The BF estimates (in log scale) reported in Table 2 are based on the marginal likelihood estimates presented in Table 1. In all three simulated datasets, the estimated Bayes factors $\widehat{BF}$ indicated the true model. Moreover, when the $\widehat{CJ}^{\mathrm{I}}$ was used, the true model was suggested by the BF estimator at every batch. Bayes factors for the second and the third dataset clearly indicate the true model, with values ranging from e^{33} to e^{116}. Only in the first dataset is the Bayes factor much lower and equal to $e^3 \approx 20$. In the latter case, or in more extreme cases where two competing models have Bayes factors close to one, the Monte Carlo error should be small enough in order to be able to identify which model is a-posteriori supported. Here

Table 1: Simulated Example Results: Marginal Likelihood Estimates in Section 4.5.

Dataset	p	N	k_{true}	k_{model}	$\log \widehat{LM}$	$\log \widehat{CJ}^{\mathrm{I}}$	$\widehat{\log CJ}^{\mathrm{I}}_{bm}$	$MCE(\log \widehat{CJ}^{\mathrm{I}})$
1	6	600	1	1	-2175.3	-2175.2	-2175.1	0.016
				2	-2178.2	-2178.2	-2178.2	0.253
2	6	600	2	1	-2187.2	-2187.6	-2187.5	0.033
				2	-2070.8	-2071.3	-2071.2	0.066
3	7	800	3	1	-3422.4	-3422.3	-3422.5	0.029
				2	-3374.4	-3374.1	-3375.2	0.133
				3	-3341.3	-3339.1	-3339.3	0.332

p: number of items; N: number of individuals; k_{true} and k_{model}: number of factors in the true and evaluated model, respectively; $\widehat{LM}$ and $\widehat{CJ}^{\,I}$: Laplace-Metropolis and Chib and Jeliaskov estimates of the marginal likelihood; $\widehat{\log CJ}^{\,I}_{bm}$: Batch mean estimator of the log-marginal likelihood; $MCE(\log \widehat{CJ}^{\,I})$: Monte Carlo error of the $\log \widehat{CJ}^{\,I}$ estimator obtained as the standard deviation of 30 batches of equal size as the estimate reported in 7th column of the table.

we estimated an error equal to 0.25, with 95% of the estimates ranging between $e^{2.5} = 12.2$ and $e^{3.3} = 27.1$, in all cases safely inferring in favour of the true generating mechanism.

Table 2: Simulated Example Results: Bayes Factor Estimates in Section 4.5.

Dataset details				Comparison			Batch summaries of $\log \widehat{CJ}^{\mathrm{I}}$			
#	p	N	k_{true}	k_1 vs. k_2	$\log \widehat{BF}^{\,(LM)}$	$\log \widehat{BF}^{\,(CJ)}$	Mean	S.D.	1^{st}Q	3^{rd}Q
1	6	600	1	$1-2$	3.1	3.0	3.1	0.25	2.5	3.3
2	6	600	2	$2-1$	116.3	116.3	116.3	0.08	116	116.5
3	7	800	3	$3-1$	81.1	83.3	83.2	0.33	81.5	84.5
				$3-2$	33.3	35.0	35.9	0.35	34.3	37.7

p: number of items; N: number of individuals; k_{true}: number of factors in the true model; k_1 vs. k_2: the Bayes factor comparing the k_1 versus the k_2-factor model is estimated; $\widehat{BF}^{\,(LM)}$ and $\widehat{BF}^{\,(CJ)}$: Estimated Bayes factors based on Laplace-Metropolis and Chib and Jeliaskov estimates of the marginal likelihood; Batch summaries of $\log \widehat{CJ}^{\,I}$: Summaries based on 30 batches of $\log \widehat{CJ}^{\,I}$ (mean=Batch mean estimate, S.D.= standard deviation - provides an estimate for the Monte Carlo Error, 1^{st}Q and 3^{rd}Q: first and third quartiles).

4.6 Implementation on real datasets

We proceed with two real-data examples also analyzed in Bartholomew et al. (2008, chapter 8). In all examples the marginal likelihood was estimated via CJ^{I} and LP methods at the median point, over samples of 10 thousand iterations (after discarding 1000 iterations as a burn in period and keeping 1 every 10 iterations to reduce autocorrelations).

The first data set is originally provided by Bock and Lieberman (1970) and is part of the Law School Admission Test (LSAT) completed by $N = 1005$ individuals. The test consists of five items and was originally designed to measure one latent factor which is also supported by the computed Bayes factor (≈ 0.22 and 0.24 for LM and CJ^I, respectively; posterior weight of one-factor model 0.802 and 0.817 respectively) reported in the first row of Table 3.

The second data set is part of the 1990 Workplace Industrial Relations Survey (WIRS, Airey et al. 1992). The Bayes factor of the two versus the one-factor model clearly supports the latter ($\log BF_{21} \approx 69$); see second line of Table 3. Bartholomew et al. (2008) suggest to omit the first item of the scale in order to improve the model fit. The analysis was replicated for the 5 items version implying again two factor model as the best model but the BF_{21} is now much lower and of magnitude approximately equal to 40.

Table 3: Marginal Likelihood and Bayes Factors for the Real Datasets in Section 4.6

	$\log \widehat{LM}$			$\log \widehat{CJ}^{\mathrm{I}}$		
Dataset	1-factor	2-factor	$\log \widehat{BF}_{21}^{(LM)}$	1-factor	2-factor	$\log \widehat{BF}_{21}^{(CJ)}$
1. LSAT	-2494.8	-2496.2	-1.4	-2495.1	-2496.6	-1.5
2. WIRS-6 items	-3456.1	-3387.1	69.0	3456.2	-3387.3	68.9
3. WIRS-5 items	-2786.6	-2782.8	3.8	-2786.8	-2783.1	3.7

$\widehat{LM}$ and $\widehat{CJ}^{\,I}$: Laplace-Metropolis and Chib and Jeliaskov estimates of the marginal likelihood; 1-factor and 2-factor columns: estimates of the log-marginal likelihood for the 1-factor and 2-factor models, respectively $\widehat{BF}_{21}^{(LM)}$ and $\widehat{BF}_{21}^{(CJ)}$: Estimated Bayes factors of 2-factor versus 1-factor model based on $\widehat{LM}$ and $\widehat{CJ}^{\,I}$, respectively.

5 CONCLUSION

The paper introduces a simplification of the Chib and Jeliazkov (2001) marginal likelihood estimator for models with conditional independence. This approach drastically reduces the computational effort required for the marginal likelihood estimate, since only a single MCMC run is required for the computation of the posterior ordinate. For this reason, this result is quite appealing in high dimensional models that include either random effects or latent variables.

This strategy can be easily implemented even for models with no latent variables when the posterior distribution can be augmented using auxiliary variables which introduce local independence between parameters (Tanner and Wong 1987,van Dyk and Meng 2001).

Acknowledgements

This research was funded by the Research Centre of the Athens University of Economics and Business.

References

Airey, C., N. Tremlett, and R. Hamilton (1992). The workplace industrial relations survey 1990. *Technical Report (Main and Panel Surveys)*, Social and Community Planning Research.

Bartholomew, D. J., F. Steele, I. Moustaki, and J. Galbraith (2008). *Analysis of Multivariate Social Science Data* (2nd ed.). Chapman & Hall/CRC.

Besag, J. (1989). A candidate's formula: A curious result in Bayesian prediction. *Biometrika 76*, 183.

Bock, R. D. and M. Lieberman (1970). Fitting a response model for n dichotomously scored items. *Psychometrika 35*, 179–197.

Bratley, P., B. L. Fox, and L. E. Schrage (1987). *A Guide to Simulation*. New York: Springer-Verlag.

Chib, S. (1995). Marginal likelihood from the Gibbs output. *Journal of the American Statistical Association 90*, 1313–1321.

Chib, S. and I. Jeliazkov (2001). Marginal likelihood from the Metropolis-Hastings output. *Journal of the American Statistical Association 96*, 270–281.

Dunson, D. B. (2006). Efficient Bayesian model averaging in factor analysis. Technical report, Biostatistics Branch, National Institute of Environmental Health Sciences, Duke University.

Fouskakis, D., I. Ntzoufras, and D. Draper (2009). Bayesian variable selection using cost-adjusted BIC, with application to cost-effective measurement of quality of health care. *Annals of Applied Statistics 3*, 663–690.

Friel, N. and A. Pettit (2008). Marginal likelihood estimation via power posteriors. *Journal of Royal Statistical Society 770*, 589–607.

Huber, P., E. Ronchetti, and M.-P. Victoria-Feser (2004). Estimation of generalized linear latent variable models. *Journal of the Royal Statistical Society, Series B 66*, 893–908.

Kass, R. and A. Raftery (1995). Bayes factors. *Journal of the American Statistical Association 90*, 773–795.

Lewis, S. and A. Raftery (1997). Estimating Bayes factors via posterior simulation with the Laplace Metropolis estimator. *Journal of the American Statistical Association 92*, 648–655.

Lopes, H. F. and M. West (2004). Bayesian model assessment in factor analysis. *Statistica Sinica 14*, 4167.

Meng, X. and W. Wong (1996). Simulating ratios of normalizing constants via a simple identity: A theoretical exploration. *Statistica Sinica 6*, 831–860.

Mira, A. and G. Nicholls (2004). Bridge estimation of the probability density at a point. *Statistica Sinica 14*, 603–612.

Moustaki, I. and M. Knott (2000). Generalized latent trait models. *Psychometrika 65*, 391–411.

Newton, M. and A. Raftery (1994). Approximate Bayesian inference with the weighted likelihood bootstrap. *Journal of the Royal Statistical Society 56*, 3–48.

Ntzoufras, I., P. Dellaportas, and J. Forster (2000). Bayesian variable and link determination for generalised linear models. *Journal of Statistical Planning and Inference 111*, 165–180.

Patz, R. and B. A. Junker (1999). A straightforward approach to Markov chain Monte Carlo methods for item response models. *Journal of Educational and Behavioral Statistics 24*, 146–178.

Rabe-Hesketh, S., A. Skrondal, and A. Pickles (2005). Maximum likelihood estimation of limited and discrete dependent variable models with nested random effects. *Journal of Econometrics 128*, 301–323.

Raftery, A., M. Newton, J. Satagopan, and P. Krivitsky (2007). Estimating the integrated likelihood via posterior simulation using the harmonic mean identity (with discussion). J. Bernardo, J. Bayarri, and J. Berger, eds., *Bayesian Statistics*, Vol. 8, pp. 371–416. Oxford University Press.

Schilling, S. and R. Bock (2005). High-dimensional maximum marginal likelihood item factor analysis by adaptive quadrature. *Psychometrika 70*, 533–555.

Schmeiser, B. W. (1982). Batch size effects in the analysis of simulation output. *Operations Research 30*, 556–568.

Tanner, M. A. and W. H. Wong (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association 82*, 528–540.

van Dyk, D. A. and X.-L. Meng (2001). The art of data augmentation. *Journal of Computational and Graphical Statistics 10*, 1–50.

Vitoratou, S., I. Moustaki, and I. Ntzoufras (2011). Using the independence assumption to reduce the Monte Carlo error in latent variable models. (work in progress), Department of Statistics, Athens University of Economics and Business.