國 立 政 治 大 學 統 計 學 系

博 士 學 位 論 文

指 導 教 授: 翁久幸 博士

馬可夫鏈蒙地卡羅收斂的研究與貝氏漸進的表現

A Study of MCMC Convergence and Performance Evaluation of Bayesian Asymptotics

研究生: 許正宏 撰

中 華 民 國 一 百 年 六 月

謝 辭

本論文能順利完成, 幸蒙吾師翁久幸教授的悉心指導, 舉凡研究方向, 觀念的啓迪, 架構的匡正, 求學態度的斧正, 乃至於本論文內容的逐字批閱與修飾, 特此獻上最深的敬意與感謝。同時, 也謝謝口試委員曹振海老師、蕭朱杏老師、徐南蓉老師、蔡紋琦老師與黃子銘老師於論文口試期間, 對我的鼓勵以及論文的指正, 遂使本論文得以臻於完備。

修業期間, 感謝系上諸位老師在課業知識的傳授與助教在行政事務的協助。同時, 也謝謝同窗好友漢葳、金田、學弟宇翔、泰期、建郎的切磋討論與激勵, 獲益匪淺。

隨著論文的付梓, 博士班的修業即將完成, 這段日子裏, 有挫折, 也有收穫和成長。經過這一段歷練, 不僅淬勵我的人生, 也讓自己的視野更宏觀。

最後, 特別感謝我敬愛的雙親對我的養育與支持; 感謝內人愼如的包容與犧牲, 在她細心照顧小孩與操持家務下, 讓我能全心投入並順利完成論文。

僅以本論文獻給我最摯愛的家人。

許正宏 謹誌

中華民國一百年六月

中文摘要

本論文主要討論貝氏漸近的比較, 推導出參數的聯合後驗分配與利用圖形來診斷馬可夫鏈蒙地卡羅的收斂。Johnson (1970) 利用泰勒展開式得到個別後驗分配的展開式, 此展開式是根據概似函數與先驗分配。Weng (2010b) 和 Weng and Hsu (2011) 利用 Stein's 等式且由概似函數與先驗分配估計後驗動差; 將這些後驗動差代入 Edgeworth 展開式得到近似後驗分配, 此近似分配的誤差可精確到大 O 的負 3/2 次方與 Johnson's 相同。另外 Weng and Hsu (2011) 發現 Weng (2010b) 和 Johnson (1970) 的近似展開式各別項誤差到大 O 的負 1 次方不一致, 由模擬結果得到 Weng's 在此項表現比 Johnson's 好。另外由 Weng (2010b) 得到一維參數 的 Edgeworth 近似後驗分配延伸到二維參數的聯合後驗分配; 並應用二維參數的聯合後驗分配於多階段資料。本論文我們提出利用圖形來診斷馬可夫鏈蒙地卡羅收斂的方法, 並且應用一般化線性模型與混合常態模型做爲模擬。

關鍵字: Edgeworth 展開式; 馬可夫鏈蒙地卡羅; 個別後驗分配; Stein's 等式。

Abstract

Johnson (1970) obtained expansions for marginal posterior distributions through Taylor expansions. The expansion in Johnson (1970) is expressed in terms of the likelihood and the prior. Weng (2010b) and Weng and Hsu (2011) showed that by using Stein's identity we can approximate the posterior moments in terms of the likelihood and the prior; then substituting these approximations into an Edgeworth series one can obtain an expansion which is correct to $O(t^{-3/2})$, similar to Johnson's. Weng and Hsu (2011) found that the $O(t^{-1})$ terms in Weng (2010b) and Johnson (1970) do not agree and further compared these two expansions by simulation study. The simulations confirmed this finding and revealed that our $O(t^{-1})$ term gives better performance than Johnson's. In addition to the comparison of Bayesian asymptotics, we try to extend Weng (2010a)'s Edgeworth series for the distribution of a single parameter to the joint distribution of all parameters. Since the calculation is quite complicated, we only derive expansions for the two-parameter case and apply it to the experiment of multi-stage data. Markov Chain Monte Carlo (MCMC) is a popular method for making Bayesian inference. However, convergence of the chain is always an issue. Most of convergence diagnosis in the literature is based solely on the simulation output. In this dissertation, we proposed a graphical method for convergence diagnosis of the MCMC sequence. We used some generalized linear models and mixture normal models for simulation study. In summary, the goals of this dissertation are threefold: to compare some results in Bayesian asymptotics, to study the expansion for the joint distribution of the parameters and its applications, and to propose a method for convergence diagnosis of the MCMC sequence.

Key words: Edgeworth expansion; Markov Chain Monte Carlo; marginal posterior distribution; Stein's identity.

Contents

List of Figures

1 Introduction

Let $g(\theta)$ be a smooth function on the parameter space Θ. The calculation of the posterior mean of $g(\theta)$, given a sample of observations x_t, requires integration over Θ of the form

$$E_\xi^t[g(\theta)] = E_\xi[g(\theta)|x_t] = \frac{\int_\Theta g(\theta)\exp(\ell_t(\theta))\xi(\theta)d\theta}{\int_\Theta \exp(\ell_t(\theta))\xi(\theta)d\theta}, \tag{1}$$

where ℓ_t is the log-likelihood function and ξ the prior. Quantities about the posterior distribution such as moments, quantiles, cumulative distribution function, etc can be expressed in the form of (1) by some functions g.

The integrations involved in (1) are usually intractable. The approximation techniques can be divided into deterministic and nondeterministic methods. The nondeterministic method refers to the Monte Carlo integration such as Markov Chain Monte Carlo (MCMC) methods, which draw samples approximately from the desired distribution and forms sample averages to estimate the expectation. The deterministic approaches include Laplace's method, variational Bayes, among others. The Laplace's method approximates integrals by Taylor expansions and properties of Gaussian distribution (see Section 2.3 for details); the variational Bayes method constructs a lower bound on marginal likelihood of the data and then tries to optimize this bound.

Both deterministic and nondeterministic methods have their advantages. Computing techniques like Markov chain Monte Carlo and importance sampling have made many computations possible. Still, analytic approximations are simpler to compute for some models, and are useful as a starting point for more exact methods. The study of these analytic approximations are often referred to as Bayesian asymptotics. Moreover, when it comes to sequential updating with new and huge data, the MCMC methods may not be computationally feasible, the reason being that it does not use of the analysis from the previous data; see, for example, Section 2.8 in Glickman (1993).

A conventional deterministic approach to the problem (1) starts from a Taylor series expansion at the maximum likelihood estimator (or at the modes of the integrands), proceeds from there to develop expansions on both the numerator and denominator, and then obtains approximations by formal division of the two series. For example, Johnson (1967, 1970) derived expansions, correct to order $O(t^{-3/2})$, for the posterior distribution associated with some pivotal quantity. Tierney and Kadane (1986) renewed interest in Laplace's method by assuming that g is positive and expanding the integrand of the numerator in (1) at the mode of the integrand itself, rather than at the posterior mode. They also derived an expression for marginal posterior density of the parameter and argued that by numerical renormalization the approximate density has relative errors of order $O(t^{-3/2})$ in neighborhoods of the mode.

Recently Weng (2010a) uses a version of Stein's identity to derive an Edgeworth series for the posterior distribution of a normalized quantity. Note that an Edgeworth series expands a probability distribution in terms of its moments; in contrast, the expansion in Johnson (1970) is expressed in terms of the likelihood and the prior. Weng (2010b) and Weng and Hsu (2011) further showed that by using that Stein's identity we can approximate the posterior moments in terms of the likelihood and the prior; then substituting these approximations into the Edgeworth series one can obtain an expansion which is correct to $O(t^{-3/2})$, similar to Johnson's. We shall compare these three $O(t^{-3/2})$ results. Especially, Johnson (1970) formulas have existed for about four decades, but to the best of our knowledge there seems to be no related simulation studies in the literature. So, our study filled a gap in this area.

In addition to the comparison of Bayesian asymptotics, we try to extend Weng (2010a)'s Edgeworth series for the distribution of a single parameter to the joint distribution of all parameters. Since the calculation is quite complicated, we only derive expansions for the two-parameter case (see Section 3.2). Using such expansions we conduct experiments to study whether analytic approximations perform well when data arrive in different stages.

The present thesis also studies how to use the Edgeworth expansion for posterior distribution to validate convergence of MCMC simulation results. We use some generalized linear models and mixture normal models for simulation study.

In summary, the goals of this dissertation are threefold: to compare some results in Bayesian asymptotics, to study the expansion for the joint distribution of the parameters and its applications, and to propose a method for convergence diagnosis of the MCMC sequence.

The next chapter introduces the model and reviews some existing results. Chapter 3 presents theoretical results. Some simulation studies in Weng and Hsu (2011) are given in Subsection 4.1.1. Subsection 4.1.2 gives comparisons with Tierney and Kadane (1986). Section 4.2 considers approximations when data arrives in two stages. The rest of Chapter 4 contains simulation studies for diagnosing MCMC series. Chapter 5 gives some remarks and directions for future research.

2 Preliminaries

2.1 The Model

Let X_t be a random vector distributed according to a family of probability densities $p_t(x_t|\theta)$, where t is a discrete or continuous parameter and $\theta \in \Theta$, an open subset in $\Re^p$. Consider a Bayesian model in which θ has a prior density ξ which is twice differentiable on $\Re^p$ and vanishes off of Θ. Assume that the log-likelihood function $\ell_t(\theta)$ is twice continuously differentiable with respect to θ. Assume also that the maximum likelihood estimator $\hat{\theta}_t$ exists and satisfies $\nabla \ell_t(\hat{\theta}_t) = 0$ and $-\nabla^2 \ell_t(\hat{\theta}_t)$ being positive definite, where ∇ indicates differentiation with respect to θ. Define Σ_t and Z_t as

$$\Sigma_t^T \Sigma_t = -\nabla^2 \ell_t(\hat{\theta}_t), \tag{2}$$

$$Z_t = \Sigma_t(\theta - \hat{\theta}_t). \tag{3}$$

Then the posterior density of θ given data x_t is $\xi_t(\theta) \propto \exp(\ell_t(\theta))\xi(\theta)$, and the posterior density of Z_t is

$$\zeta_t(z) \propto \xi_t(\theta(z)) \propto \exp[\ell_t(\theta) - \ell_t(\hat{\theta}_t)]\xi(\theta), \tag{4}$$

where the relation of θ and z is given in (3). Now define

$$u_t(\theta) = \ell_t(\theta) - \ell_t(\hat{\theta}_t) + \frac{1}{2}\|z_t\|^2. \tag{5}$$

So, (4) can be rewritten as

$$\zeta_t(z) \propto \phi_p(z) f_t(z), \tag{6}$$

where $f_t(z) = \xi(\theta(z))\exp[u_t(\theta)]$ and ϕ_p denotes the standard p-variate normal density and ϕ is the abbreviation of ϕ_1.

2.2 Stein's Identity and Bayesian Edgeworth Expansion

The Stein's Identity here was developed by Woodroofe (1989). Let Φ_p denote the standard p-variate normal distribution and let Φ be the abbreviation of Φ_1. Write

$$\Phi_p h = \int h d\Phi_p$$

for functions h for which the integral is finite.

For $r > 0$, denote $H_r^{(p)}$ as the collection of all measurable functions $h : \Re^p \to \Re$ for which $|h(z)|/b \leq 1 + \|z\|^r$ for some $b > 0$. Given $h \in H_r^{(p)}$, let $h_0 = \Phi_p h$, $h_p = h$,

$$\begin{aligned} h_k(y_1, ..., y_k) &= \int_{\Re^{p-k}} h(y_1, ..., y_k, w)\Phi_{p-k}(dw), & (7) \\ g_k(y_1, ..., y_p) &= e^{\frac{1}{2}y_k^2} \int_{y_k}^{\infty} [h_k(y_1, ..., y_{k-1}, w) - h_{k-1}(y_1, ..., y_{k-1})]e^{-\frac{1}{2}w^2} dw, & (8) \end{aligned}$$

for $-\infty < y_1, ..., y_p < \infty$ and $k = 1, ..., p$. Then let

$$Uh = (g_1, ..., g_p)^T \text{ and } Vh = (U^2 h + U^2 h^T)/2, \qquad (9)$$

where $U^2 h$ is the $p \times p$ matrix whose kth column is Ug_k and g_k is as in (8).

We will call a function $f : \Re^k \to \Re$ almost differentiable if there exists a function $\nabla f : \Re^k \to \Re^k$ such that

$$f(x + y) - f(x) = \int_0^1 y^T \nabla f(x + ty) dt$$

for $x, y \in \Re^k$. Of course, a continuously differentiable function f is almost differentiable with ∇f equal to the gradient.

Lemma 1 *(Stein's Identity) Let r be a nonnegative integer. Suppose that f is an almost differentiable function on $\Re^p$ and*

$$\int_{\Re^p} |f| d\Phi_p + \int_{\Re^p} (1 + \|z\|^r)\|\nabla f(z)\|\Phi_p(dz) < \infty.$$

Then

$$\Phi_p(fh) = \Phi_p(f)\Phi_p(h) + \int_{\Re^p} (Uh(z))^T \nabla f(z)\Phi_p(dz), \qquad (10)$$

for all $h \in H_r^{(p)}$. If $\partial f/\partial z_j, j = 1, ..., p$, are almost differentiable, and

$$\int_{\Re^p} (1 + \|z\|^r)\|\nabla^2 f(z)\|\Phi_p(dz) < \infty,$$

then

$$\Phi_p(fh) = \Phi_p(f)\Phi_p(h) + (\Phi_p Uh)^T \int_{\Re^p} \nabla f(z)\Phi_p(dz) + \int_{\Re^p} tr[(Vh(z))\nabla^2 f(z)]\Phi_p(dz), \quad (11)$$

for all $h \in H_r^{(p)}$.

Here tr denotes the trace of a matrix. Note that if A and B are $p \times p$ matrices, then $\mathrm{tr}(AB) = \sum_{i,j} a_{ij}b_{ji} = \sum_{i=1}^{p} a_i^T b_i$. Simple calculations by taking $f(z)$ in (10) as z_i and $f(z)$ in (11) as $z_i z_j$ yield

$$\begin{aligned} \Phi_p(Uh) &= \int_{\Re^p} zh(z)\Phi_p(dz), & (12)\\ \Phi_p(Vh) &= \int_{\Re^p} \frac{1}{2}(zz^T - I_p)h(z)\Phi_p(dz). & (13) \end{aligned}$$

From (6), the posterior distribution is of a form appropriate for Stein's Identity. So, by Lemma 1,

$$E_\xi^t\{h(Z_t)\} = \Phi_p h + E_\xi^t\left\{[Uh(Z_t)]^T \frac{\nabla f_t(Z_t)}{f_t(Z_t)}\right\}, \quad (14)$$

$$E_\xi^t\{h(Z_t)\} = \Phi_p h + (\Phi_p Uh)^T E_\xi^t\left[\frac{\nabla f_t(Z_t)}{f_t(Z_t)}\right] + E_\xi^t\left\{\mathrm{tr}\left[Vh(Z_t)\frac{\nabla^2 f_t(Z_t)}{f_t(Z_t)}\right]\right\}. \quad (15)$$

In particular, if $h(z) = z_i$, then by (8) it follows $Uh(z) = e_i$; and if $h(z) = z_i z_j$ and $i < j$, then $Uh(z) = z_i e_j$, where $\{e_1, \cdots, e_k\}$ denote the standard basis for $\Re^k$. For example, if $k = 3$ and $h(z) = z_1 z_2$, then

$$Uh(z) = \begin{pmatrix} 0 \\ z_1 \\ 0 \end{pmatrix} = \begin{pmatrix} g_1(z) \\ g_2(z) \\ g_3(z) \end{pmatrix}$$

$$\text{and } U^2h(z) = U(Uh) = U\big(g_1, g_2, g_3\big)^T = \big(Ug_1, Ug_2, Ug_3\big) = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

where Ug_i is obtained by applying the operator U on g_i. With these special h functions, (14) and (15) became

$$\begin{aligned} E_\xi^t Z_t &= E_\xi^t\left[\frac{\nabla f_t(Z_t)}{f_t(Z_t)}\right], & (16)\\ E_\xi^t(Z_{ti}Z_{tq}) &= \delta_{iq} + E_\xi^t\left[\frac{\nabla^2 f_t(Z_t)}{f_t(Z_t)}\right]_{iq}, \quad \text{for } i, q = 1, \ldots, k, & (17) \end{aligned}$$

where $\delta_{iq} = 1$ if $i = q$ and 0 otherwise, and $[\cdot]_{iq}$ indicates the (i, q) component of a matrix.

Weng (2010a) obtained an Edgeworth expansion for the posterior distribution of Z_{tj}. Since some arguments below are needed in Section 3.2, we give detailed descriptions. Let q_k denote Hermite polynomials, given by $q_k(z)\phi(z) = (-d/dz)^k\phi(z)$. For instance, for $k = 1, ..., 4$ the Hermite polynomials are $q_1(z) = z$, $q_2(z) = z^2 - 1$, $q_3(z) = z^3 - 3z$, and $q_4(z) = z^4 - 6z^2 + 3$. Then, for $j = 1, ..., p$,

$$\sup_{h^* \in H_r^{(1)}} \left| E_\xi^t(h^*(Z_{tj})) - \Phi h^* - \sum_{\substack{k \in \{1,\ldots,3s\} \\ k \neq 3s-1}} (\Phi U^k h^*) E_\xi^t \Big[\frac{\partial^k f_t / \partial z_{tj}^k}{f_t}(Z_t) \Big] \right| = O(t^{-\frac{3s+1}{2}+s}), \quad (18)$$

where $h^* : \Re \to \Re$ is a measurable function and $H_r^{(p)}$ is defined at the beginning of this section. Together with the facts that

$$E_\xi^t\Big(\frac{\partial^k f_t / \partial z_{tj}^k}{f_t}(Z_t) \Big) = E_\xi^t(q_k(Z_{tj})) \quad (19)$$

and

$$\Phi(U^k h^*) = \frac{1}{k!} \int_\Re q_k(z) h^*(z) \Phi(dz), \quad (20)$$

the posterior distribution of Z_{tj} can be expressed in terms of its moments.

In (18), taking $h^*(z_j) = 1_{\{z_j \le z_j^*\}}$ leads to an expansion for $P_\xi^t(Z_{tj} \le z_j^*)$; and collecting terms in this expansion according to their orders yields

$$P_\xi^t(Z_{tj} \le z_j^*) = \Phi(z_j^*) - \sum_{i=1}^m R_{it}(z_j^*)\phi(z_j^*) + O(t^{-\frac{m+1}{2}}), \quad (21)$$

where

$$R_{it}(z_j^*) = \sum_{v \in J_i} \frac{1}{v!} q_{v-1}(z_j^*) E_\xi^t(q_v(Z_{tj})) = O(t^{-\frac{i}{2}}).$$

Here $J_1 = \{1, 3\}$ and $J_i = \{3i - 4, 3i - 2, 3i\}$ for $i > 1$; for example, $J_2 = \{2, 4, 6\}$, $J_3 = \{5, 7, 9\}$. Equation (21) is referred to as a Bayesian Edgeworth expansion.

If Σ_t in (2) is obtained by Cholesky decomposition, then it is an upper triangular and by (3) we obtain

$$Z_{tp} = [\Sigma_t]_{pp}(\theta_p - \hat{\theta}_{tp}). \quad (22)$$

From this, we have $P_\xi^t(\theta_p \le a_p) = P_\xi^t(Z_{tp} \le z_p^*)$, where $z_p^* = [\Sigma_t]_{pp}(a_p - \hat{\theta}_{tp})$. By (18)-(20) and the relation $\int_{-\infty}^w q_k(z)\phi(z)dz = -q_{k-1}(w)\phi(w)$ it follows that

$$\xi_p^t(a_p) = [\Sigma_t]_{pp}\{\phi(z_p^*) + \sum_{\substack{k \in \{1,\ldots,3s\} \\ k \neq 3s-1}} \frac{1}{k!} q_k(z_p^*)\phi(z_p^*) E_\xi^t(q_k(Z_{tp})) + O(t^{-\frac{3s+1}{2}+s})\}. \quad (23)$$

Equation (23) is a marginal posterior density.

2.3 The Laplace Method

In this section we describe the approximate marginal posterior density proposed by Tierney and Kadane (1986). A sequence of observations $y = (y_1, \dots, y_t)$ is assumed to be drawn independently from density $f(y|\theta)$; so the posterior distribution of θ is

$$\pi(\theta|y) = \frac{\pi(\theta)\prod_{i=1}^{t} f(y_i|\theta)}{\int \pi(\theta)\prod_{i=1}^{t} f(y_i|\theta)d\theta} \tag{24}$$

where $\pi(\theta)$ is the prior of θ.

Laplace's method provides an approximation for integrals of the form $\int e^{tL(\theta)}d\theta$ when t is large. If L has a unique maximum at $\hat{\theta}$ and we set $\Sigma = -(\nabla^2 L(\hat{\theta}))^{-1}$, then we can approximate $L(\theta)$ by $L(\hat{\theta}) - (1/2)(\theta - \hat{\theta})^T \Sigma^{-1}(\theta - \hat{\theta})$. This produces the approximation

$$\int e^{tL(\theta)}d\theta = \det(2\pi\Sigma/t)^{1/2}\exp\{tL(\hat{\theta})\}.$$

To obtain an approximation for (24), set $\theta = (\theta_1, \theta_2, .., \theta_p) = (\theta_1, \theta_{-1})$. Suppose $\hat{\theta}$ is the posterior mode, and let Σ be minus the inverse of the Hessian of $\log[\pi(\theta)\prod_{i=1}^{t} f(y_i|\theta)]$ at $\hat{\theta}$; thus Σ is a $p \times p$ matrix. For a given θ_1, let the $(p-1)$ vector $\hat{\theta}^*_{-1} = \hat{\theta}^*_{-1}(\theta_1)$ maximize the function $\pi(\theta_1, .)e^{\ell(\theta_1, .)}$, the function πe^{ℓ} with θ_1 held fixed, and let $\Sigma^* = \Sigma^*(\theta_1)$ be minus the inverse of the Hessian of $\log[\pi(\theta_1, .)\prod_{i=1}^{t} f(y_i|\theta_1, .)]$, a $(p-1)\times(p-1)$ matrix.

Tierney and Kadane (1986) applied Laplace's method to the integrals in the numerator and denominator of the expression

$$\pi_1(\theta_1|y) = \frac{\int \pi(\theta_1, \theta_{-1})e^{\ell(\theta_1, \theta_{-1})}d\theta_{-1}}{\int \pi(\theta)e^{\ell(\theta)}d\theta},$$

for the marginal posterior density of θ_1 and obtained the approximation

$$\hat{\pi}_1(\theta_1|y) = \left(\frac{\det\Sigma^*(\theta_1)}{2\pi t \det\Sigma}\right)^{1/2} \frac{\pi(\theta_1, \hat{\theta}^*_{-1})e^{\ell(\theta_1, \hat{\theta}^*_{-1})}}{\pi(\hat{\theta})e^{\ell(\hat{\theta})}}. \tag{25}$$

By (25), one only needs to be able to maximize slightly modified likelihood functions and to evaluate the observed information at the maxima. The principal regularity condition required is that the likelihood times prior be unimodal.

2.4 The Gibbs Sampling and MCMC Convergence Diagnostic

The Gibbs sampler, proposed by Geman and Geman (1984), is a special case of single-component Metropolis-Hastings. For details, see Gilks et al. (1995, chapter 5) and Tanner (1993, chapter 6). It is widely used in statistical applications. The Gibbs sampler is to obtain samples from a joint distribution, via iterated sampling from full conditional distributions. Let $\theta = (\theta_1, \theta_2, \ldots, \theta_p)$, the full conditional distributions are

$$\pi(\theta_i|\theta_1, \theta_2, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_p, y) = \frac{\pi(\theta, y)}{\int \pi(\theta, y) d\theta_i} \tag{26}$$

for $i = 1, 2, \ldots, p$, where $\pi(\theta, y) = \pi(\theta) \prod_{i=1}^{t} f(y_i|\theta)$. The Gibbs sampling algorithm is described below. Given the starting point

$$\theta^{(0)} = (\theta_1^{(0)}, \theta_2^{(0)}, \ldots, \theta_p^{(0)}),$$

this algorithm iterates the following loop:

$$\text{Sample } \theta_1^{(i+1)} \text{ from } \pi(\theta_1|\theta_2^{(i)}, \ldots, \theta_p^{(i)}, y)$$

$$\text{Sample } \theta_2^{(i+1)} \text{ from } \pi(\theta_2|\theta_1^{(i+1)}, \theta_3^{(i)}, \ldots, \theta_p^{(i)}, y)$$

$$\vdots$$

$$\text{Sample } \theta_p^{(i+1)} \text{ from } \pi(\theta_p|\theta_1^{(i+1)}, \ldots, \theta_{p-1}^{(i+1)}, y).$$

The vectors $\theta^{(1)}, \theta^{(2)}, \ldots, \theta^{(t)}, \ldots$ are a realization of a Markov chain.

MCMC Convergence Diagnostic By Gelman et al. (1995, chapter 11), for each scalar estimand θ, we label the draws J parallel sequences of length t as θ_{ij} ($i = 1, ..., t$; $j = 1, ..., J$), and we compute B and W, the between- and within-sequence variances:

$$B = \frac{t}{J-1} \sum_{j=1}^{J} (\overline{\theta_{.j}} - \overline{\theta_{..}})^2, \text{ where } \overline{\theta_{.j}} = \frac{1}{t} \sum_{i=1}^{t} \theta_{ij}, \ \ \overline{\theta_{..}} = \frac{1}{J} \sum_{j=1}^{J} \overline{\theta_{.j}}$$

$$W = \frac{1}{J} \sum_{j=1}^{J} s_j^2, \text{ where } s_j^2 = \frac{1}{t-1} \sum_{i=1}^{t} (\theta_{ij} - \overline{\theta_{.j}})^2.$$

The between-sequence variances, B , contains a factor of t because it is based on the variance of the within-sequence means, $\overline{\theta_{.j}}$, each of which is an average of t values θ_{ij}.

We can estimate $\text{var}(\theta|y)$, the marginal posterior variance of the estimand, by a weighted average of W and B, namely

$$\widehat{\text{var}}^{+}(\theta|y) = \frac{t-1}{t}W + \frac{1}{t}B,$$

which overestimates the marginal posterior variance assuming the starting distribution is approximately overdispersed, but is unbiased under stationarity, or in the limit $t \to \infty$. We monitor convergence of the iterative simulation by estimating the factor by which the scale of the current distribution for θ might be reduced if the simulations were continued in the limit $t \to \infty$. This potential scale reduction is estimated by

$$\widehat{R} = \sqrt{\frac{\widehat{\text{var}}^{+}(\theta|y)}{W}},$$

which declines to 1 as $t \to \infty$.

2.5 The Generalized Linear Model

We briefly review the generalized linear model. For details, see McCullagh and Nelder (1989). Generalized linear models are an extension of classical linear models. We may demonstrate the classical linear model in the form: The components of y are independent normal variables with constant variance σ^2 and

$$E(y) = \mu, \text{ where } \mu = X\beta. \tag{27}$$

For generalized linear models, we shall rearrange (27) to produce the following three-part specification:

1. The random component: the components of y have independent normal distributions with $E(y) = \mu$ and constant variance σ^2;

2. The systematic component: covariates $x_1, x_2, \ldots, x_p$ produce a linear predictor η given by

$$\eta = \sum_{1}^{p} x_j \theta_j;$$

3. The link between the random and systematic components:

$$\mu = \eta.$$

If we write

$$\eta_i = \mathrm{g}(\mu_i), \tag{28}$$

then $g(.)$ will be called the link function. In this formulation, generalized linear models allow two extensions; first the distribution in component 1 may come from an exponential family other than the normal, and secondly the link function in component 3 may become any monotonic differentiable function. We shall consider several principal functions in subsequent chapters, namely:

$$\text{logit} : \eta = \log\{\mu/(1-\mu)\},$$

$$\text{probit} : \eta = \Phi^{-1}(\mu),$$

$$\text{poisson} : \eta = \log\mu,$$

$$\text{gamma} : \eta = -1/\mu.$$

3 Theoretical Results

3.1 Validation of Simulation Results

The expansion (23) can be used to validate simulation results. If the posterior sample is given, then using standard techniques we can obtain an approximate density based on this sample. On the other hand, we can calculate the empirical moments of Z_{tp} based on this sample and plug these moment approximations into (23). This gives another density approximation. We shall call this approximation the *"implied density"* from the sample, and denote it as $\widehat{\xi}_p^{t}{}^{(imp)}$:

$$\widehat{\xi}_p^{t}{}^{(imp)}(a_p) = [\Sigma_t]_{pp}\left\{\phi(z_p^*) + \sum_{\substack{k\in\{1,\dots,3s\}\\ k\neq 3s-1}} \frac{1}{k!} q_k(z_p^*)\phi(z_p^*) E_\xi^t\widetilde{(q_k(Z_{tp}))}^{(mcmc)} + O(t^{-\frac{3s+1}{2}+s})\right\}. \tag{29}$$

If the posterior sample has converged to true distribution, then theoretically the posterior density and $\widehat{\xi}_p^{t}{}^{(imp)}$ should be close. This is the basic idea we use to validate convergence.

For inference on θ_p we use Z_{tp} in (22), which requires the MLE of θ_p and $[\Sigma_t]_{pp}$, the (p,p) component of the Cholesky decomposition of $-\nabla^2\hat{\ell}_t$. Note that it is important that

Z_{tp} involves only θ_p so that $P^t_\xi(\theta_p \le a_p) = P^t_\xi(Z_{tp} \le z^*_p)$ for $z^*_p = [\Sigma_t]_{pp}(a_p - \hat{\theta}_{tp})$. The inference on θ_1 can be achieved by taking Σ_t in (2) to be lower triangular so that Z_{t1} in (3) can be expressed as $Z_{t1} = [\Sigma_t]_{11}(\theta_1 - \hat{\theta}_{t1})$; and the inference of other θ_i can be done by the following: exchange the ith and pth columns in the design matrix and obtain the corresponding (p,p) component of the Cholesky decomposition.

If for each θ_i we need to obtain a new design matrix and re-do model fitting etc, it may not be efficient. In the rest of this section we propose a simple method for this problem.

In (3), if $\Sigma_t = [\sigma_{ij}]$ has only one nonzero element σ_{ii} in the ith row, then

$$Z_{ti} = \sigma_{ii}(\theta_i - \hat{\theta}_{ti}) \tag{30}$$

and it follows that $\{Z_{ti} \le z^*_i\} = \{\theta_i \le z^*_i/\sigma_{ii} + \hat{\theta}_{ti}\}$ and $P^t_\xi(\theta_i \le a_i) = P^t_\xi(Z_{ti} \le z^*_i)$, where $z^*_i = \sigma_{ii}(a_i - \hat{\theta}_{ti})$. Therefore, we obtain the implied density similar to (29) for all θ_i. That is,

$$\widehat{\xi^t_i}^{(imp)}(a_i) = \sigma_{ii}\left\{\phi(z^*_i) + \sum_{\substack{k\in\{1,\ldots,3s\}\\ k\neq 3s-1}} \frac{1}{k!} q_k(z^*_i)\phi(z^*_i)\widetilde{E^t_\xi(q_k(Z_{ti}))}^{(mcmc)} + O(t^{-\frac{3s+1}{2}+s})\right\},$$

where Z_{ti} is defined in (30).

The lemma below shows how to obtain such Σ_t.

Lemma 2 *Let A be a $p\times p$ symmetric and positive definite matrix. Let E_{ip} be the $p \times p$ matrix obtained by exchanging the ith and pth row (or column) of I_p for $i = 1, 2, ..., p$. Let $C = E_{ip}AE_{ip}$, $D = \mathrm{chol}(C)$ so that $C = D^TD$ and $B = E_{ip}DE_{ip}$. Then, for each $i = 1, 2, ..., p$,*
(a) $E^T_{ip} = E_{ip}$ and $E_{ip}E_{ip} = I_p$
(b) $B = [b_{jl}]\,, 1 \le j, l \le p$ has only one nonzero element b_{ii} in the ith row, and $B^TB = A$

Proof. First consider (a). For a fixed $i \in \{1, 2, ..., p\}$, write $E_{ip} = [e_{jl}]\,, 1 \le j, l \le p$. Then,

$$e_{ip} = e_{pi} = 1,$$
$$e_{kk} = 1 \;\; for \;\; k \ne i, p,$$

and the other elements are all zero. So $E_{ip}^T = E_{ip}$. Write $F = E_{ip}E_{ip} = [f_{jl}]$, $1 \leq j, l \leq p$. Then

$$f_{ii} = e_{ip}e_{pi} = 1 \;\; and \;\; f_{pp} = e_{pi}e_{ip} = 1,$$
$$f_{kk} = e_{kk}e_{kk} = 1 \;\; for \;\; 1 \leq k \leq p \;\; and \;\; k \neq i, p,$$

and the other elements are all zero. So $E_{ip}E_{ip} = I_p$.

Next consider (b). Since A is symmetric and positive definite matrix, $C = E_{ip}AE_{ip}$ is also symmetric and positive definite matrix. Since $D = \text{chol}(C)$, $C = D^T D$ and D is an upper triangular matrix. Note that $E_{ip}D$ is the matrix obtained by exchanging the ith and pth rows of D. So, the ith row of $E_{ip}D$ has only one nonzero element, and it is in the (i, p) entry. Moreover, this nonzero element is exactly d_{pp}. Note that $B = (E_{ip}D)E_{ip}$ is the matrix obtained by exchanging the ith and pth columns of $E_{ip}D$. So, the ith row of B has only one nonzero element, which is in the (i, i) entry, and $b_{ii} = d_{pp}$. For example, let $p = 3$

$$E_{23} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \text{ and } D = \begin{pmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 6 \end{pmatrix}.$$

Then,

$$E_{23}D = \begin{pmatrix} 1 & 2 & 3 \\ 0 & 0 & 6 \\ 0 & 4 & 5 \end{pmatrix} \text{ and } B = E_{23}DE_{23} = \begin{pmatrix} 1 & 3 & 2 \\ 0 & 6 & 0 \\ 0 & 5 & 4 \end{pmatrix}.$$

Then,

$$\begin{aligned} B^T B &= E_{ip}^T D^T E_{ip}^T E_{ip} D E_{ip} \\ &= E_{ip} D^T E_{ip} E_{ip} D E_{ip} \\ &= E_{ip} D^T D E_{ip} \\ &= E_{ip} C E_{ip} \\ &= E_{ip}(E_{ip} A E_{ip}) E_{ip} \\ &= A, \end{aligned}$$

where the second, the third and the sixth equalities follows from part (a). □

We summarize the steps for obtaining σ_{ii} in (30). Let $-\nabla^2\hat{\ell}_t$ be minus the observed information matrix. For $i = 1, 2, \ldots, p$,

1. let $C = E_{ip}(-\nabla^2\hat{\ell}_t)E_{ip}$;
2. let $D = \text{chol}(C)$ so that D is upper triangular and $C = D^T D$;
3. let $B = E_{ip} D E_{ip}$;
4. then $B^T B = -\nabla^2\hat{\ell}_t$ and the (i, i) entry in B is the σ_{ii} in (30).

3.2 Joint Posterior Distributions

In this section we try to extend the marginal posterior density (23) to the joint posterior density. Some notations are needed. Let

$$\begin{aligned} h^{(0)} = h, h^{(1)} = Uh = \{h_i^{(1)}, i = 1, \cdots, p\}, \\ h^{(2)} = U^2 h = U(Uh) = \{h_{i_1 i_2}^{(2)}, i_1, i_2 = 1, \cdots, p\}, \\ h^{(k)} = U^k h = U(U^{k-1}h) = \{h_{i_1 i_2 \cdots i_k}^{(k)}, i_1, i_2, \cdots, i_k = 1, \cdots, p\}, \end{aligned} \tag{31}$$

where Uh is defined in (8). So, $h^{(1)}$ is a $p \times 1$ array of functions, $h^{(2)}$ is a $p \times p$ array, and $h^{(k)}$ is a $p \times p \times \ldots \times p$($k$ terms) array. For any function $f : R^p \to R$, denote by $f_{i_1,\ldots,i_s}^{(s)}(z)$ the sth derivative of f with respect to $z_{i_1}, \ldots, z_{i_s}$. One can rewrite (14) and (15) as

$$\begin{aligned} E_\xi^t\{h(Z_t)\} &= \Phi_p h + \sum_i E_\xi^t\left[h_i^{(1)} \frac{f_{t,i}^{(1)}(Z_t)}{f_t(Z_t)}\right] \\ E_\xi^t\{h(Z_t)\} &= \Phi_p h + \sum_i \Phi_p h_i^{(1)} E_\xi^t\left[\frac{f_{t,i}^{(1)}(Z_t)}{f_t(Z_t)}\right] + \sum_{i,j} E_\xi^t\left[h_{ij}^{(2)} \frac{f_{t,ij}^{(2)}(Z_t)}{f_t(Z_t)}\right]; \end{aligned}$$

and we can obtain a more general form:

$$\begin{aligned} E_\xi^t\{h(Z_t)\} &= \Phi_p h + \sum_{i_1} \Phi_p h_{i_1}^{(1)} E_\xi^t\left[\frac{f_{t,i_1}^{(1)}(Z_t)}{f_t(Z_t)}\right] + \sum_{i_1,i_2} \Phi_p h_{i_1 i_2}^{(2)} E_\xi^t\left[\frac{f_{t,i_1 i_2}^{(2)}(Z_t)}{f_t(Z_t)}\right] \\ &\quad + \sum_{i_1,i_2,i_3} (\Phi_p h_{i_1 i_2 i_3}^{(3)}) E_\xi^t\left[\frac{f_{t,i_1 i_2 i_3}^{(3)}(Z_t)}{f_t(Z_t)}\right] + \cdots + \sum_{i_1,\ldots,i_s} (\Phi_p h_{i_1 \ldots i_s}^{(s)}) E_\xi^t\left[\frac{f_{t,i_1 \ldots i_s}^{(s)}(Z_t)}{f_t(Z_t)}\right] \\ &\quad + \sum_{i_1,\ldots,i_{s+1}} E_\xi^t\left[h_{i_1 \ldots i_{s+1}}^{(s+1)}(Z_t) \frac{f_{t,i_1 \ldots i_{s+1}}^{(s+1)}(Z_t)}{f_t(Z_t)}\right], \end{aligned} \tag{32}$$

provided all the expectations exist. Take $h(z) = 1_{\{Z_i \le z_i^*,\, i=1,\cdots,p\}}$, where $z_i^* \in R$. Then the left hand side of (32) becomes the joint cumulative distribution of Z_t, and the right hand side gives an expansion for the joint distribution. To obtain an expansion similar to (23), we need to extend the results in (19) and (20).

Q1: How does $E_\xi^t\left[f_{t,i_1\ldots i_s}^{(s)}(Z_t)/f_t(Z_t)\right]$ relate to the moments of Z_t?

Q2: How to calculate $\Phi_p h_{i_1\ldots i_s}^{(s)}$ and $\partial^p \Phi_p h_{i_1\ldots i_s}^{(s)}/\partial a_1 \cdots \partial a_p$?

Since the generalization requires heavy calculations, in this thesis we only look at 2-dimensional cases. For Q1, from (16) and (17) we know that

$$E_\xi^t\left[\frac{\nabla f_t(Z_t)}{f_t(Z_t)}\right] \text{ and } E_\xi^t\left[\frac{\nabla^2 f_t(Z_t)}{f_t(Z_t)}\right]$$

involve the first and the second posterior moments of Z_t; and (16) and (17) are derived by taking $h(z)$ in (14) and (15) as z_i and $z_i z_j$. To generalize these results, note that if $h(z) = z_i^3$, then by (8) it follows $Uh(z) = (z_i^2+2)e_i$; and if $h(z) = z_i^2 z_j$ with $i<j$, then $Uh(z) = z_i^2 e_j$ where $\{e_1, \cdots, e_k\}$ denote the standard basis for $\Re^k$. If $h(z) = z_1^3$, then by (9) it follows that

$$Uh(z) = \begin{pmatrix} z_1^2+2 \\ 0 \end{pmatrix} \text{ and } U^2h(z) = U\big(g_1, g_2\big) = \begin{pmatrix} z_1 & 0 \\ 0 & 0 \end{pmatrix};$$

and by (31), U^3h is a $2\times2\times2$ array with $h_{111}^{(3)} = 1$ and $h_{ijk}^{(3)} = 0$ for $(i,j,k) \ne (1,1,1)$. By (12), (13) and (32) with $s=3$, we have

$$\begin{aligned}
E_\xi^t Z_{t1}^3 &= 3E_\xi^t Z_{t1} + E_\xi^t\left\{\frac{f_{111}^{(3)}}{f}(Z_t)\right\} \\
E_\xi^t(Z_{t1}^2 Z_{t2}) &= E_\xi^t Z_{t2} + E_\xi^t\left\{\frac{f_{112}^{(3)}}{f}(Z_t)\right\} \\
E_\xi^t(Z_{t1} Z_{t3}^2) &= E_\xi^t Z_{t1} + E_\xi^t\left\{\frac{f_{133}^{(3)}}{f}(Z_t)\right\} \\
E_\xi^t(Z_{t1} Z_{t2} Z_{t3}) &= E_\xi^t\left\{\frac{f_{123}^{(3)}}{f}(Z_t)\right\}.
\end{aligned}$$

For Q2, let

$$A = \{\theta : \theta_i \le a_i, i=1,2\},\ B = \{z = \Sigma_t(\theta - \hat\theta_t) : \theta \in A\}, \text{ and } h(z) = 1_{\{z\in B\}}. \tag{33}$$

We have

$$
\begin{aligned}
\Phi_2 h &= \int_B \phi_2(z)dz \\
&= \int_A \phi_2(z)|\Sigma_t|d\theta \\
&= |\Sigma_t| \int_{-\infty}^{a_1}\int_{-\infty}^{a_2} \phi_2(z)d\theta_2 d\theta_1.
\end{aligned}
$$

So,

$$
\begin{aligned}
\frac{\partial^2 \Phi_2 h}{\partial a_1 \partial a_2} &= |\Sigma_t| \frac{\partial^2}{\partial a_1 \partial a_2}\left\{\int_{-\infty}^{a_1}\int_{-\infty}^{a_2} \phi_2(z)d\theta_2 d\theta_1\right\} \\
&= |\Sigma_t|\phi_2(z^*),
\end{aligned}
$$

where

$$
z^* = \Sigma_t(a - \hat{\theta}_t). \tag{34}
$$

By (12)

$$
\begin{aligned}
\Phi_2 U h &= \int z h(z)\Phi_2(dz) \\
&= \int_A z\phi_2(z)|\Sigma_t|d\theta \\
&= |\Sigma_t| \int_{-\infty}^{a_1}\int_{-\infty}^{a_2} z\phi_2(z)d\theta_2 d\theta_1,
\end{aligned}
$$

and it follows that

$$
\begin{aligned}
\frac{\partial^2 \Phi_2 U h}{\partial a_1 \partial a_2} &= |\Sigma_t| \frac{\partial^2}{\partial a_1 \partial a_2}\left\{\int_{-\infty}^{a_1}\int_{-\infty}^{a_2} z\phi_2(z)d\theta_2 d\theta_1\right\} \\
&= |\Sigma_t| z^*\phi_2(z^*),
\end{aligned}
$$

which is a 2×1 vector; moreover, taking $f(z)$ in (32) as $z_i z_j$ gives $\Phi_2 h_{ij}^{(2)}$, where $i,\ j = 1,\ 2$. First, let $f(z) = z_1^2$. Then

$$
\begin{pmatrix} f_1^{(1)}(z) \\ f_2^{(1)}(z) \end{pmatrix} = \begin{pmatrix} 2z_1 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} f_{11}^{(2)}(z) & f_{12}^{(2)}(z) \\ f_{21}^{(2)}(z) & f_{22}^{(2)}(z) \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}.
$$

So, by (32),

$$
\begin{aligned}
\Phi_2 h_{11}^{(2)} &= \int \frac{1}{2}(z_1^2 - 1)h(z)\Phi_2(dz) \\
&= \int_A \frac{1}{2}(z_1^2 - 1)\phi_2(z)|\Sigma_t|d\theta \\
&= |\Sigma_t| \int_{-\infty}^{a_1}\int_{-\infty}^{a_2} \frac{1}{2}(z_1^2 - 1)\phi_2(z)d\theta_2 d\theta_1;
\end{aligned}
$$

and hence,

$$\begin{aligned}\frac{\partial^2 \Phi_2 h_{11}^{(2)}}{\partial a_1 \partial a_2} &= |\Sigma_t| \frac{\partial^2}{\partial a_1 \partial a_2} \left\{ \int_{-\infty}^{a_1} \int_{-\infty}^{a_2} \frac{1}{2}(z_1^2 - 1)\phi_2(z) d\theta_2 d\theta_1 \right\} \\ &= \frac{1}{2}|\Sigma_t|(z_1^{*2} - 1)\phi_2(z^*);\end{aligned}$$

The two terms $\Phi_2 h_{12}^{(2)}$ and $\Phi_2 h_{21}^{(2)}$ may be different, but can be combined. So we don't need to solve out the two terms individually. To see how, let $f(z) = z_1 z_2$. Then,

$$\begin{pmatrix} f_1^{(1)}(z) \\ f_2^{(1)}(z) \end{pmatrix} = \begin{pmatrix} z_2 \\ z_1 \end{pmatrix}, \quad \begin{pmatrix} f_{11}^{(2)}(z) & f_{12}^{(2)}(z) \\ f_{21}^{(2)}(z) & f_{22}^{(2)}(z) \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

By (16) and (17) it follows that

$$\begin{aligned}\Phi_2 h_{12}^{(2)} + \Phi_2 h_{21}^{(2)} &= \int z_1 z_2 h(z) \Phi_2(dz) \\ &= \int_A z_1 z_2 \phi_2(z) |\Sigma_t| d\theta \\ &= |\Sigma_t| \int_{-\infty}^{a_1} \int_{-\infty}^{a_2} z_1 z_2 \phi_2(z) d\theta_2 d\theta_1.\end{aligned}$$

So,

$$\begin{aligned}\frac{\partial^2 (\Phi_2 h_{12}^{(2)} + \Phi_2 h_{21}^{(2)})}{\partial a_1 \partial a_2} &= |\Sigma_t| \frac{\partial^2}{\partial a_1 \partial a_2} \left\{ \int_{-\infty}^{a_1} \int_{-\infty}^{a_2} z_1 z_2 \phi_2(z) d\theta_2 d\theta_1 \right\} \\ &= |\Sigma_t| z_1^* z_2^* \phi_2(z^*);\end{aligned}$$

$$\begin{aligned}\Phi_2 h_{22}^{(2)} &= \int \frac{1}{2}(z_2^2 - 1) h(z) \Phi_2(dz) \\ &= \int_A \frac{1}{2}(z_2^2 - 1)\phi_2(z) |\Sigma_t| d\theta \\ &= |\Sigma_t| \int_{-\infty}^{a_1} \int_{-\infty}^{a_2} \frac{1}{2}(z_2^2 - 1)\phi_2(z) d\theta_2 d\theta_1;\end{aligned}$$

and hence,

$$\begin{aligned}\frac{\partial^2 \Phi_2 h_{22}^{(2)}}{\partial a_1 \partial a_2} &= |\Sigma_t| \frac{\partial^2}{\partial a_1 \partial a_2} \left\{ \int_{-\infty}^{a_1} \int_{-\infty}^{a_2} \frac{1}{2}(z_2^2 - 1)\phi_2(z) d\theta_2 d\theta_1 \right\} \\ &= \frac{1}{2}|\Sigma_t|(z_2^{*2} - 1)\phi_2(z^*);\end{aligned}$$

Now, taking $f(z)$ in (32) as $z_i z_j z_k$ yields $\Phi_2 h_{ijk}^{(3)}$, where $i,\ j,\ k = 1,\ 2$. First, let $f(z) = z_1^3$. Then

$$\begin{pmatrix} f_1^{(1)}(z) \\ f_2^{(1)}(z) \end{pmatrix} = \begin{pmatrix} 3z_1^2 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} f_{11}^{(2)}(z) & f_{12}^{(2)}(z) \\ f_{21}^{(2)}(z) & f_{22}^{(2)}(z) \end{pmatrix} = \begin{pmatrix} 6z_1 & 0 \\ 0 & 0 \end{pmatrix},$$

$f_{111}^{(3)}(z) = 6$ and $f_{ijk}^{(3)}(z) = 0$ for $(i, j, k) \neq (1, 1, 1)$. So, by (32),

$$\begin{aligned}
\Phi_2 h_{111}^{(3)} &= \int \frac{1}{6}(z_1^3 - 3z_1) h(z) \Phi_2(dz) \\
&= \int_A \frac{1}{6}(z_1^3 - 3z_1)\phi_2(z)|\Sigma_t| d\theta \\
&= |\Sigma_t| \int_{-\infty}^{a_1} \int_{-\infty}^{a_2} \frac{1}{6}(z_1^3 - 3z_1)\phi_2(z) d\theta_2 d\theta_1;
\end{aligned}$$

and hence,

$$\begin{aligned}
\frac{\partial^2 \Phi_2 h_{111}^{(3)}}{\partial a_1 \partial a_2} &= |\Sigma_t| \frac{\partial^2}{\partial a_1 \partial a_2} \left\{ \int_{-\infty}^{a_1} \int_{-\infty}^{a_2} \frac{1}{6}(z_1^3 - 3z_1)\phi_2(z) d\theta_2 d\theta_1 \right\} \\
&= \frac{1}{6}|\Sigma_t|(z_1^{*3} - 3z_1^*)\phi_2(z^*);
\end{aligned}$$

Next let $f(z) = z_1^2 z_2$. Then,

$$\begin{pmatrix} f_1^{(1)}(z) \\ f_2^{(1)}(z) \end{pmatrix} = \begin{pmatrix} 2z_1 z_2 \\ z_1^2 \end{pmatrix}, \quad \begin{pmatrix} f_{11}^{(2)}(z) & f_{12}^{(2)}(z) \\ f_{21}^{(2)}(z) & f_{22}^{(2)}(z) \end{pmatrix} = \begin{pmatrix} 2z_2 & 2z_1 \\ 2z_1 & 0 \end{pmatrix},$$

$f_{ijk}^{(3)}(z) = 0$ except the three terms $f_{112}^{(3)}(z) = f_{121}^{(3)}(z) = f_{211}^{(3)}(z) = 2$, and $f_{t,i_1,i_2,i_3,i_4}^{(4)}(z) = 0$. By (32),

$$\begin{aligned}
E_\xi^t\{h(Z_t)\} = \Phi_2 h &+ \sum_{i_1} \Phi_p h_{i_1}^{(1)} E_\xi^t \left[\frac{f_{t,i_1}^{(1)}(Z_t)}{f_t(Z_t)} \right] + \sum_{i_1,i_2} \Phi_2 h_{i_1 i_2}^{(2)} E_\xi^t \left[\frac{f_{t,i_1 i_2}^{(2)}(Z_t)}{f_t(Z_t)} \right] \\
&+ \sum_{i_1,i_2,i_3} \Phi_2 h_{i_1 i_2 i_3}^{(3)} E_\xi^t \left[\frac{f_{t,i_1 i_2 i_3}^{(3)}(Z_t)}{f_t(Z_t)} \right] + \sum_{i_1,\ldots,i_4} E_\xi^t \left[h_{i_1\ldots i_4}^{(4)} \frac{f_{t,i_1\ldots i_4}^{(4)}(Z_t)}{f_t(Z_t)} \right];
\end{aligned}$$

together with (16) and (17) it follows that

$$\begin{aligned}
\Phi_2 h_{112}^{(3)} + \Phi_2 h_{121}^{(3)} + \Phi_2 h_{211}^{(3)} &= \int \frac{1}{2}(z_1^2 z_2 - z_2) h(z) \Phi_2(dz) \\
&= \int_A \frac{1}{2}(z_1^2 z_2 - z_2)\phi_2(z)|\Sigma_t| d\theta \\
&= |\Sigma_t| \int_{-\infty}^{a_1} \int_{-\infty}^{a_2} \frac{1}{2}(z_1^2 z_2 - z_2)\phi_2(z) d\theta_2 d\theta_1.
\end{aligned}$$

So,

$$\begin{aligned}
\frac{\partial^2 (\Phi_2 h_{112}^{(3)} + \Phi_2 h_{121}^{(3)} + \Phi_2 h_{211}^{(3)})}{\partial a_1 \partial a_2} &= |\Sigma_t| \frac{\partial^2}{\partial a_1 \partial a_2} \left\{ \int_{-\infty}^{a_1} \int_{-\infty}^{a_2} \frac{1}{2}(z_1^2 z_2 - z_2)\phi_2(z) d\theta_2 d\theta_1 \right\} \\
&= \frac{1}{2}|\Sigma_t|(z_1^{*2} z_2^* - z_2^*)\phi_2(z^*).
\end{aligned}$$

Though the three terms $\Phi_2 h^{(3)}_{112}$, $\Phi_2 h^{(3)}_{121}$ and $\Phi_2 h^{(3)}_{211}$ may be different, we do not need to write them out separately. Similarly, we can obtain

$$\begin{aligned}
\Phi_2 h^{(3)}_{122} + \Phi_2 h^{(3)}_{212} + \Phi_2 h^{(3)}_{221} &= \int \frac{1}{2}(z_1 z_2^2 - z_1) h(z) \Phi_2(dz) \\
&= \int_A \frac{1}{2}(z_1 z_2^2 - z_1)\phi_2(z)|\Sigma_t| d\theta \\
&= |\Sigma_t| \int_{-\infty}^{a_1}\int_{-\infty}^{a_2} \frac{1}{2}(z_1 z_2^2 - z_1)\phi_2(z) d\theta_2 d\theta_1,
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2(\Phi_2 h^{(3)}_{122} + \Phi_2 h^{(3)}_{212} + \Phi_2 h^{(3)}_{221})}{\partial a_1 \partial a_2} &= |\Sigma_t| \frac{\partial^2}{\partial a_1 \partial a_2}\left\{\int_{-\infty}^{a_1}\int_{-\infty}^{a_2} \frac{1}{2}(z_1 z_2^2 - z_1)\phi_2(z) d\theta_2 d\theta_1\right\} \\
&= \frac{1}{2}|\Sigma_t|(z_1^* z_2^{*2} - z_1^*)\phi_2(z^*);
\end{aligned}$$

$$\begin{aligned}
\Phi_2 h^{(3)}_{222} &= \int \frac{1}{6}(z_2^3 - 3z_2) h(z) \Phi_2(dz) \\
&= \int_A \frac{1}{6}(z_2^3 - 3z_2)\phi_2(z)|\Sigma_t| d\theta \\
&= |\Sigma_t| \int_{-\infty}^{a_1}\int_{-\infty}^{a_2} \frac{1}{6}(z_2^3 - 3z_2)\phi_2(z) d\theta_2 d\theta_1,
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 \Phi_2 h^{(3)}_{222}}{\partial a_1 \partial a_2} &= |\Sigma_t| \frac{\partial^2}{\partial a_1 \partial a_2}\left\{\int_{-\infty}^{a_1}\int_{-\infty}^{a_2} \frac{1}{6}(z_2^3 - 3z_2)\phi_2(z) d\theta_2 d\theta_1\right\} \\
&= \frac{1}{6}|\Sigma_t|(z_2^{*3} - 3z_2^*)\phi_2(z^*).
\end{aligned}$$

In the following we shall write down the joint posterior density of θ.

Let A, B, and $h(z)$ be as in (33). So, $\theta \in A$ if and only if $z \in B$; and the joint probability and joint density of θ are

$$P^t_\xi(\theta_i \leq a_i, i = 1, 2) = P^t_\xi(\theta \in A) = P^t_\xi(Z_t \in B),$$

and

$$\xi^t(a_1, a_2) = \frac{\partial^2 P^t_\xi(\theta \in A)}{\partial a_1 \partial a_2} = \frac{\partial^2 E^t_\xi[h(Z_t)]}{\partial a_1 \partial a_2}.$$

With the results in previous paragraphs, we have

$$\begin{aligned}
\xi^t(a_1,a_2) &= \frac{\partial^2 P_\xi^t(\theta\in A)}{\partial a_1\partial a_2} = \frac{\partial^2 E_\xi^t[h(Z_t)]}{\partial a_1\partial a_2} \\
&= \frac{\partial^2}{\partial a_1\partial a_2}\Phi_2 h + \sum_{i_1}\frac{\partial^2}{\partial a_1\partial a_2}\Phi_2 h_{i_1}^{(1)} E_\xi^t\left[\frac{f_{t,i_1}^{(1)}(Z_t)}{f_t(Z_t)}\right] \\
&+ \sum_{i_1,i_2}\frac{\partial^2}{\partial a_1\partial a_2}\Phi_2 h_{i_1 i_2}^{(2)} E_\xi^t\left[\frac{f_{t,i_1 i_2}^{(2)}(Z_t)}{f_t(Z_t)}\right] \\
&+ \sum_{i_1,i_2,i_3}\frac{\partial^2}{\partial a_1\partial a_2}\Phi_2 h_{i_1 i_2 i_3}^{(3)} E_\xi^t\left[\frac{f_{t,i_1 i_2 i_3}^{(3)}(Z_t)}{f_t(Z_t)}\right] \\
&+ \sum_{i_1,\ldots,i_4} E_\xi^t\left\{h_{i_1\ldots i_4}^{(4)}(Z_t)\frac{f_{t,i_1\ldots i_4}^{(4)}(Z_t)}{f_t(Z_t)}\right\}
\end{aligned}$$

$$\begin{aligned}
= \quad & |\Sigma_t|\phi_2(z^*) + |\Sigma_t|\phi_2(z^*)\sum_{i_1} z_{i_1}^* E_\xi^t Z_{ti_1} \\
& + \frac{1}{2}|\Sigma_t|\phi_2(z^*)\sum_{i_1 i_2}[(z_{i_1}^* z_{i_2}^* - \delta_{i_1 i_2})(E_\xi^t Z_{ti_1} Z_{ti_2} - \delta_{i_1 i_2})] \\
& + \frac{\partial^2}{\partial a_1\partial a_2}\Phi_2 h_{111}^{(3)} E_\xi^t\left[\frac{f_{t,111}^{(3)}(Z_t)}{f_t(Z_t)}\right] \\
& + \frac{\partial^2}{\partial a_1\partial a_2}(\Phi_2 h_{112}^{(3)} + \Phi_2 h_{121}^{(3)} + \Phi_2 h_{211}^{(3)}) E_\xi^t\left[\frac{f_{t,112}^{(3)}(Z_t)}{f_t(Z_t)}\right] \\
& + \frac{\partial^2}{\partial a_1\partial a_2}(\Phi_2 h_{122}^{(3)} + \Phi_2 h_{212}^{(3)} + \Phi_2 h_{221}^{(3)}) E_\xi^t\left[\frac{f_{t,122}^{(3)}(Z_t)}{f_t(Z_t)}\right] \\
& + \frac{\partial^2}{\partial a_1\partial a_2}\Phi_2 h_{222}^{(3)} E_\xi^t\left[\frac{f_{t,222}^{(3)}(Z_t)}{f_t(Z_t)}\right] \\
& + \sum_{i_1,\ldots,i_4} E_\xi^t\left\{h_{i_1\ldots i_4}^{(4)}(Z_t)\frac{f_{t,i_1\ldots i_4}^{(4)}(Z_t)}{f_t(Z_t)}\right\}
\end{aligned}$$

$$
\begin{aligned}
= \quad & |\Sigma_t|\phi_2(z^*)[1 + q_1(z_1^*)E_\xi^t Z_{t_1} + q_1(z_2^*)E_\xi^t Z_{t_2} + \frac{1}{2}q_2(z_1^*)(E_\xi^t Z_{t1}^2 - 1) \\
& + q_1(z_1^*)q_1(z_2^*)(E_\xi^t Z_{t1}Z_{t2}) + \frac{1}{2}q_2(z_2^*)(E_\xi^t Z_{t2}^2 - 1) \\
& + \frac{1}{6}q_3(z_1^*)(E_\xi^t Z_{t1}^3 - 3E_\xi^t Z_{t1}) + \frac{1}{2}q_2(z_1^*)q_1(z_2^*)(E_\xi^t Z_{t1}^2 Z_{t2} - E_\xi^t Z_{t2}) \\
& + \frac{1}{2}q_1(z_1^*)q_2(z_2^*)(E_\xi^t Z_{t1}Z_{t2}^2 - E_\xi^t Z_{t1}) + \frac{1}{6}q_3(z_2^*)(E_\xi^t Z_{t2}^3 - 3E_\xi^t Z_{t2})] \\
& + \sum_{i_1,\dots,i_4} E_\xi^t \left\{ h_{i_1\dots i_4}^{(4)}(Z_t)\frac{f_{t,i_1\dots i_4}^{(4)}(Z_t)}{f_t(Z_t)} \right\}.
\end{aligned} \tag{35}
$$

4 Experimental Results

In Section 4.1 we compare the second order approximations for posterior densities by Johnson (1970), Tierney and Kadane (1986), and Weng (2010a). For Section 4.2, we evaluate the performance of (37) when data comes in two stages. In Section 4.3-4.5, we use some GLM examples to study the use of implied density (see (29) in Section 3.1) to diagnose convergence of simulation series. In Section 4.6, we use the expansion (23) to diagnose the posterior density with multi modes, but the result is not well. All computations here are done in R (2010).

4.1 Comparison of Second Order Approximations

4.1.1 Comparison with Johnson (1970)

Johnson (1970) obtained expansions for marginal posterior distributions through Taylor expansions. Weng (2010a) showed that the marginal posterior distribution of Z_{tp} can be expanded as

$$
P_\xi^t(Z_{tp} \le z_p^*) = \Phi(z_p^*) - \sum_{i=1}^{m} R_{it}(z_p^*)\phi(z_p^*) + O(t^{-\frac{m+1}{2}}), \tag{36}
$$

where

$$
R_{it}(z_p^*) = \sum_{v \in J_i} \frac{1}{v!} q_{v-1}(z_p^*) E_\xi^t(q_v(Z_{tp})) = O(t^{-\frac{i}{2}}).
$$

Here q_v are Hermite polynomials. Weng (2010b) applied a version of Stein's identity to approximate the posterior moments in (23) and derived the marginal posterior density for

θ_p:

$$\xi_p^t(a_p) = [\Sigma_t]_{pp}\{\phi(z_p^*) + \sum_{i=1}^{m} \hat{Q}_{it}(z_p^*)\phi(z_p^*) + O(t^{-\frac{m+1}{2}})\}, \tag{37}$$

where $m = 1, 2$, which gives approximations accurate to $O(t^{-1})$ and $O(t^{-3/2})$, respectively. Here

$$Q_{it}(z_p^*) = \sum_{j \in J_i} \frac{1}{j!} q_j(z_p^*) E_\xi^t(q_j(Z_{tp})) = O(t^{-\frac{i}{2}}),$$

and $\hat{Q}_{it}(z_p^*)$ is obtained by replacing $E_\xi^t(q_j(Z_{tp}))$ with analytic approximations. These analytic approximations involve the loglikelihood and the prior and their derivatives. See Weng (2010b) and Weng and Hsu (2011). Note that $\hat{Q}_{1t}$ is of order $O(t^{-1/2})$ and $\hat{Q}_{2t}$ is $O(t^{-1})$. Weng and Hsu (2011) found that the $O(t^{-1})$ terms in Weng (2010b) and Johnson (1970) do not agree and further compared these two expansions by simulation study. The simulations confirmed this finding and revealed that our $O(t^{-1})$ term gives better performance than Johnson's. The materials below are taken from Weng (2010b) and Weng and Hsu (2011).

Johnson (1970) considered the posterior distribution of a centered and scaled variable (see his Eq. (2.1), p. 853) in 1-dimensional case:

$$\psi = (\theta - \hat{\theta}_t) b(\hat{\theta}_t), \tag{38}$$

where t is the sample size and

$$b(\hat{\theta}_t) = \Big[-\frac{1}{t} \sum_{i=1}^{t} \frac{\partial^2}{\partial \theta^2} \log f(x_i, \theta)|_{\theta = \hat{\theta}_t} \Big]^{1/2}.$$

Denote the posterior cdf of $t^{1/2}\psi$ by F_t. He showed that the posterior distribution of F_t possesses an asymptotic expansion in powers of $t^{-1/2}$ (see his Theorem 2.1):

$$|F_t(w) - \Phi(w) - \sum_{j=1}^{K} \gamma_j(w, x) t^{-j/2}| \leq D_1 t^{-\frac{1}{2}(K+1)}, \tag{39}$$

and his Proposition 2.1 shows that each $\gamma_j(w, x)$ is a polynomial in w having coefficients bounded in x multiplied by the standard normal density. Here we use two examples for simulation comparison. The first example is a Binomial-Beta model. Suppose that $X \sim$ Bin(t, θ), where the prior of θ is assumed to be Beta(a, b). Take $a = 0.5, b = 4, t = 5, x = 2$ and $a = 0.5, b = 4, t = 30, x = 12$. Thus, the posterior distributions of θ are Beta(2.5,7) and Beta(12.5,22) respectively, which are shown in Figure 1.

We compare the approximate posterior density of θ by Johnson's formulas and (37) to orders $O(t^{-1})$ and $O(t^{-3/2})$. Here Johnson's approximation to $O(t^{-1})$ is obtained by taking $K=1$ in (39):

$$p_t(w) \equiv \frac{dF_t(w)}{dw} = \phi(w) + \frac{d\gamma_j(w,x)}{dw} t^{-1/2} + O(t^{-1});$$

and the approximation to $O(t^{-3/2})$ is by taking $K=2$ in (39). Figures 1(a1) and (a2) give the true density and (37) to $O(t^{-1})$ and $O(t^{-3/2})$; Figures 1(b1) and (b2) give the true density and Johnson's approximations to $O(t^{-1})$ and $O(t^{-3/2})$; and Figures 1(c1) and (c2) contain the two $O(t^{-1})$ approximations. We have some observations. First, Figures 1(c1) and (c2) show that the two $O(t^{-1})$ approximations are quite close. Secondly, Figure 1(a1) shows that our approximation to $O(t^{-3/2})$ is closer to the true density than approximation to $O(t^{-1})$, but Figure 1(b1) reveals that Johnson's formula to $O(t^{-3/2})$ does not improve upon $O(t^{-1})$. Thirdly, Figures 1(a2) and (b2) show that the negative value of the two approximations has improved for θ (ranges between 0.5 and 1).

Next we consider a Poisson-Gamma example. Let $y_1, ..., y_t$ be an i.i.d. sample from Poisson(θ), where the prior of θ is assumed to be Gamma(a,b). Suppose that $(y_1,y_2,y_3,y_4,y_5) = (3,5,7,10,3)$ and that $(a,b)=(30,5)$. Thus, the MLE of θ is 5.6, the prior mean of θ is 6 and the posterior distribution of θ follows Gamma$(a+\sum_{i=1}^t y_i, b+t)$=Gamma(58,10). We have some observations from Figure 2(a1) to Figure 2(c1). First, Figure 2(c1) indicates that the two $O(t^{-1})$ approximations are fairly close; secondly, Figure 2(a1) shows that (37) to $O(t^{-3/2})$ improves upon $O(t^{-1})$, but Figure 2(b1) shows that Johnson's does not.

Now we try different prior distributions to see its effect on the approximations. Suppose that $(a,b)=(15,5)$. So, the prior mean of θ is 3 and $\theta|_y \sim$ Gamma(43,10). The results are in Figures 3(a1) $\sim$ 3(c1). As before, Figure 3(c1) indicates that the two $O(t^{-1})$ approximations are close. However, due to the fact that the prior mean of θ may be farther from the MLE, from Figures 3(a1) and 3(b1) we found that both $O(t^{-3/2})$ approximations are worse than $O(t^{-1})$. A closer look at these two $O(t^{-3/2})$ curves show that Johnson's approximation (ranges between -2 and 1.5) fluctuates more widely than ours (ranges between -1 and 1).

Next we try different sample sizes for Poisson-Gamma example. We do $(y_1,y_2,y_3,y_4,y_5) = (3,5,7,10,3)$ for three times to obtain a sample size 15 with $(a,b)=(30,5)$ and $(15,5)$. When $(a,b)=(30,5)$, the MLE of θ is 5.6, the prior mean of θ is 6 and $\theta|_y$ follows

Gamma(114,20). Figure 2(c2) indicates that the two $O(t^{-1})$ approximations are close. Figure 2(a2) shows that the negative value of our $O(t^{-3/2})$ approximation has improved for θ ranging between 2 and 4 while Figure 2(b2) shows that the negative value of Johnson's $O(t^{-3/2})$ approximation has also improved for θ ranging between 7 and 9. When $(a, b) = (15, 5)$, the prior mean of θ is 3 and $\theta|_y$ follows Gamma(99,20). Figure 3(c2) indicates that the two $O(t^{-1})$ approximations are close. Figure 3(a2) shows that (37) to $O(t^{-3/2})$ shortens the range of negative value according to Figure 3(a1) while Figure 3(b2) shows that Johnson's also shortens the range of negative value according to Figure 3(b1). In Figure 3, the prior mean of θ is farther from the MLE; the larger the sample size is, the better the result will be.

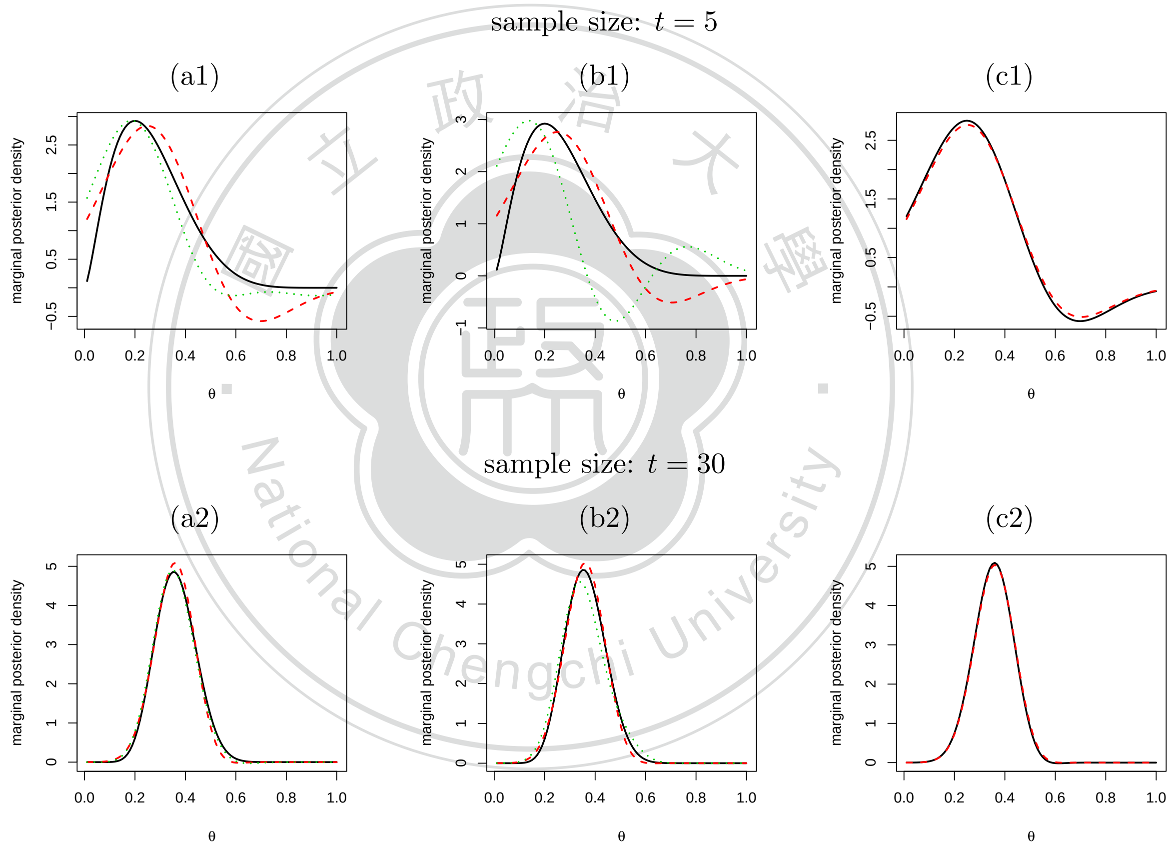


Figure 1: Marginal posterior pdf of θ. Beta-Binomial model.
(a1) and (a2) Solid: Exact; Dashed: Eq (37) to $O(t^{-1})$; Dotted: Eq (37) to $O(t^{-3/2})$.
(b1) and (b2) Solid: Exact; Dashed: Johnson's $O(t^{-1})$; Dotted: Johnson's $O(t^{-3/2})$.
(c1) and (c2) Solid: Eq (37) to $O(t^{-1})$; Dashed: Johnson's $O(t^{-1})$.

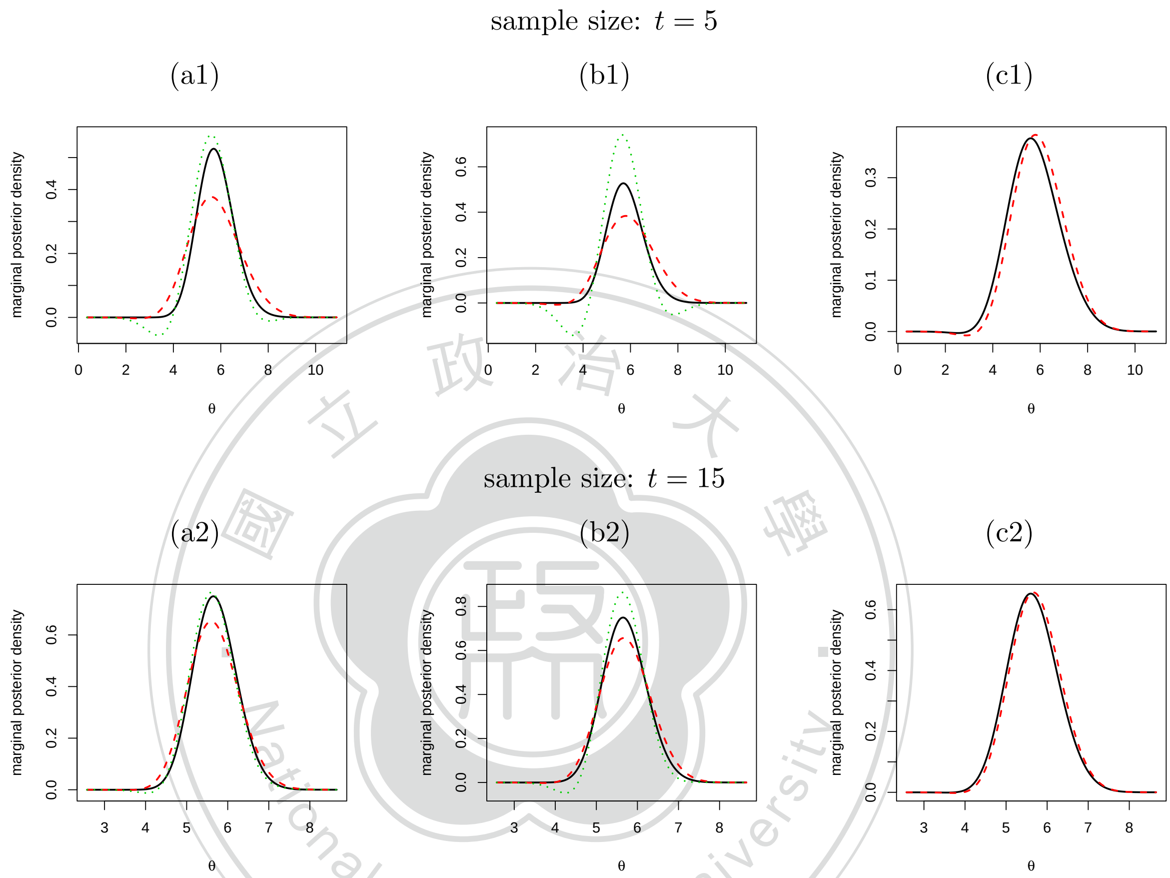


Figure 2: Marginal posterior pdf of θ. Poisson model with prior Gamma(30,5).
(a1) and (a2) Solid: Exact; Dashed: Eq (37) to $O(t^{-1})$; Dotted: Eq (37) to $O(t^{-3/2})$.
(b1) and (b2) Solid: Exact; Dashed: Johnson's $O(t^{-1})$; Dotted: Johnson's $O(t^{-3/2})$.
(c1) and (c2) Solid: Eq (37) to $O(t^{-1})$; Dashed: Johnson's $O(t^{-1})$.

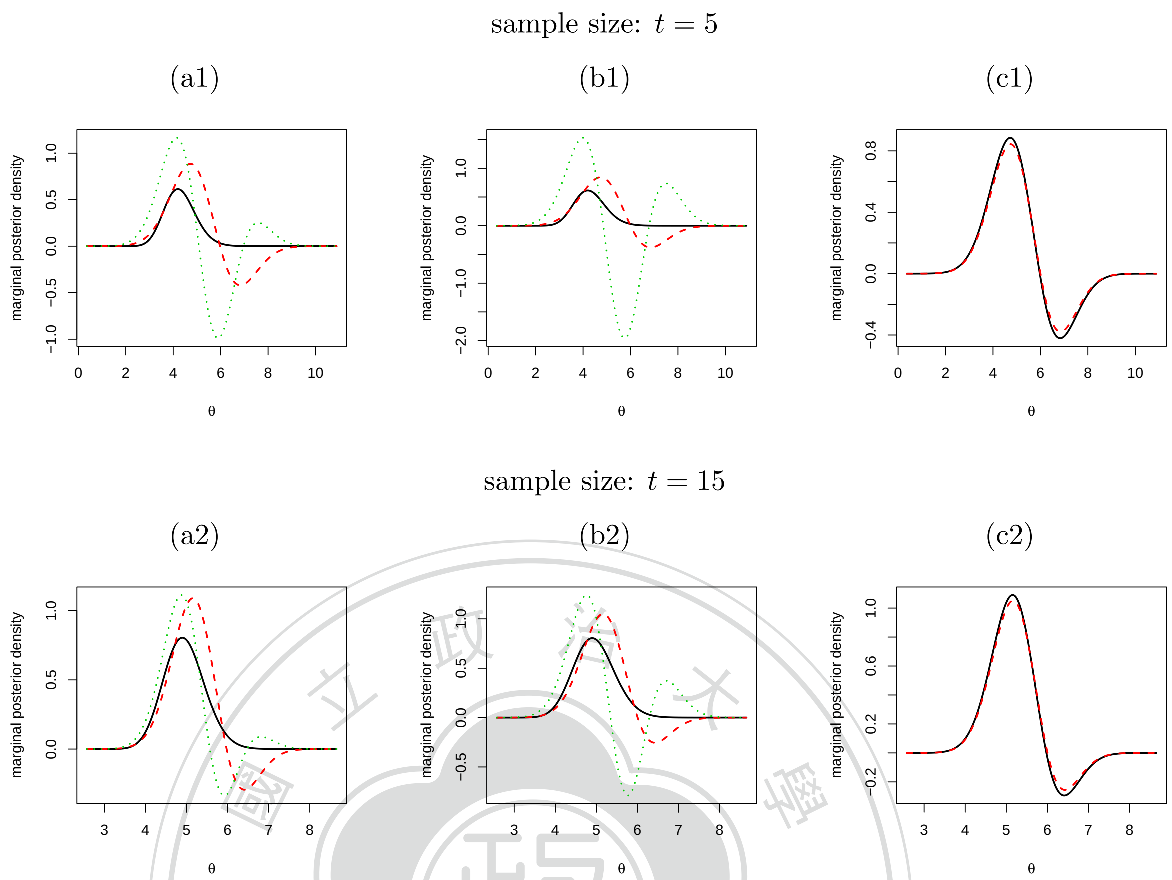


Figure 3: Marginal posterior pdf of θ. Poisson model with prior Gamma(15,5). (a1) and (a2) Solid: Exact; Dashed: Eq (37) to $O(t^{-1})$; Dotted: Eq (37) to $O(t^{-3/2})$. (b1) and (b2) Solid: Exact; Dashed: Johnson's $O(t^{-1})$; Dotted: Johnson's $O(t^{-3/2})$. (c1) and (c2) Solid: Eq (37) to $O(t^{-1})$; Dashed: Johnson's $O(t^{-1})$.

4.1.2 Comparison with Tierney and Kadane (1986)

In this section we compare (37) with (24) by Tierney and Kadane. We consider a data taken from Mendenhall et al. (1989); see also Tanner (1993). The explanatory variable is the number of days of radiotherapy received by each of 24 patients, and the response variable is the absence (1) and presence (0) of disease at a site three years after treatment. A problem of interest is to use the covariate (days) to predict outcome.

We fit the data using the logistic regression model

$$\log\Big(\frac{p_i}{1-p_i}\Big) = \theta_1 + \theta_2 x_i,$$

where x_i is the covariate for patient i and p_i is the probability of success (no disease). So, $p_i = \exp(\theta_1+\theta_2 x_i)/(1+\exp(\theta_1+\theta_2 x_i))$. The intercept θ_1 represents the log-odds of success for zero days, while the slope θ_2 represents the change in the log-odds of success for every unit increase in the covariate. The loglikelihood and the second partial derivatives are

$$\ell_t(\theta) = \sum_{i=1}^{t}[y_i \log p_i + (1-y_i)\log(1-p_i)] = \sum_{i=1}^{t}[y_i(\theta_1+\theta_2 x_i) - \log(1+\exp(\theta_1+\theta_2 x_i))],$$

$$\ell_{11}^{(2)} = -\sum_i p_i(1-p_i), \ell_{12}^{(2)} = -\sum_i x_i p_i(1-p_i), \ell_{22}^{(2)} = -\sum_i x_i^2 p_i(1-p_i). \tag{40}$$

Now we take flat priors on both θ_1 and θ_2. The comparison of the density in (25) and (37) with $m = 2$ and moments replaced by approximations are in Figure 4(a); the two methods and the true density are very close. Next we take the standard normal density priors on both θ_1 and θ_2. The results are in Figure 4(b); the result of (25) is close to the exact density, but (37) performs poorly. Here the posterior means of $q_i(Z_{t2})$, $i = 1, ..., 4$ are $(1.485, 1.460, -0.045, -3.333)$ and $(3.134, 7.207, -0.336, -3.874)$, respectively. The reason why analytic approximation (37) fails may be that the prior mean is farther from the MLE. If we change priors to $N(0, 2)$, a less informative prior than $N(0, 1)$, then the result of (37) improves a bit; see Figure 4(c). Since the posterior standard error of θ_2 is around $1/23.25 = 0.043$, the prior mean 0 is about two standard errors away from $\hat{\theta}_{t2}$. Figure 4 showed that the less informative the prior is (i.e. larger variance), the more accurate the approximate density is.

Next we consider a data first analyzed by Finney (1947); see also Albert and Chib (1993) for illustrating a sampling method for marginal posterior densities, and Myers (1990)[p.330-332]. A probit model was fit to the data. The model is

$$p_i = \Phi(\theta_1 + \theta_2 c_{1i} + \theta_3 c_{2i}),\ i = 1, ..., 39,$$

where Φ is the cdf of the standard normal density, c_{1i} is the volume of air inspired, c_{2i} is the rate of air inspired, and the binary outcome is the occurrence or non-occurrence on a transient vasorestriction on the skin of the digits. Now we take flat priors on both θ_1 and

θ_2 and obtain the results are in Figure 5, where (37) (with $m = 2$) is very close to the exact density, but (25) shifts to left slightly.

(a) (b) (c)

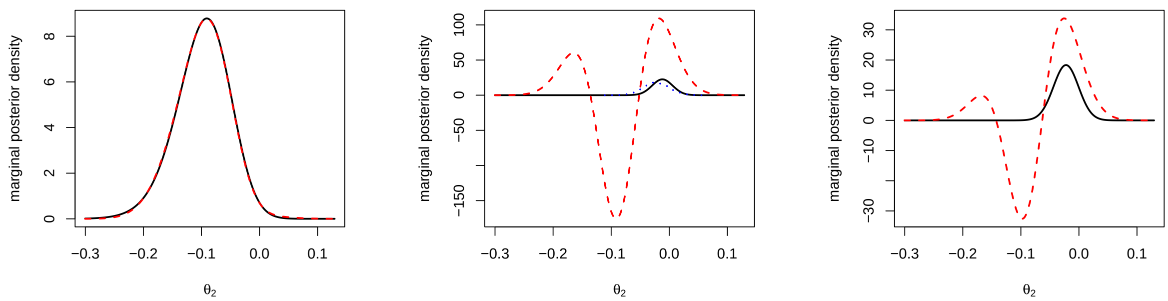


Figure 4: Marginal posterior pdf of θ_2. Logit2p model.

Solid: Equation (25); Dashed: Equation (37); Dotted: Exact distribution.

(a) flat-prior. (b) N(0,1) prior. (c) N(0,2) prior.

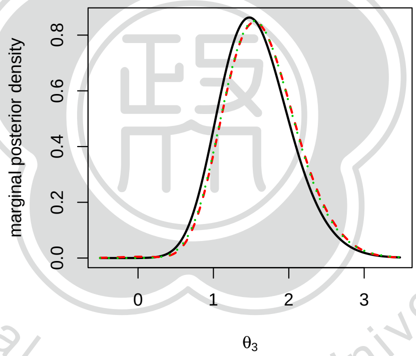


Figure 5: Marginal posterior pdf of θ_3. Probit model with flat prior. Solid: Equation (25); Dashed: Equation (37); Dotted: Exact distribution by numerical integration.

4.2 Multi-stage Data

The lemma below is well-known. It says that if data $D = (D_1, D_2)$, then the posterior density of θ given D is the same as the posterior density obtained by taking $\theta|D_1$ as the prior and D_2 as the data. This result extends to the case $D = (D_1, D_2, \ldots, D_t)$. Suppose that data arrive in different stages. With the first set of data, we can obtain the second order posterior density approximation by (37). This approximation is considered as the prior density for the next data, and the updating procedure for posterior distribution is repeated by using (37).

Lemma 3 *Let ξ be the prior for parameter θ and $D = (D_1, D_2)$ be the observed data. Assuming given θ D_1 and D_2 are independent. Let $L_1(\theta)$ and $L_2(\theta)$ be the likelihood functions based on D_1 and D_2, respectively. Let $\xi(\theta|D)$ be the posterior density of θ given D. Then*

$$\xi(\theta|D) \quad = \quad \frac{\xi(\theta|D_1)L_2(\theta)}{\int \xi(\theta|D_1)L_2(\theta)d\theta} \tag{41}$$

4.2.1 Binomial Model

First we consider a binomial model $Y \sim \mathrm{Bin}(t, \theta)$, where the prior distribution of θ is $\mathrm{Beta}(\alpha, \beta)$. Then, the posterior density for θ is

$$\begin{aligned} p(\theta|y) &\propto \theta^y(1-\theta)^{t-y}\theta^{\alpha-1}(1-\theta)^{\beta-1} \\ &\propto \theta^{y+\alpha-1}(1-\theta)^{t-y+\beta-1}. \end{aligned}$$

So,

$$\theta|y \sim \mathrm{Beta}(\alpha + y, \beta + t - y).$$

Suppose that $t = 58, y = 15, (\alpha, \beta) = (0.5, 4)$. So, the posterior distribution of θ given y is Beta(15.5,47). If the data comes in two stages: $Y_1 \sim \mathrm{Bin}(8, \theta)$ and $Y_2 \sim \mathrm{Bin}(50, \theta)$ with $y_1 = 3$ and $y_2 = 12$, then the posterior distribution of θ given y_1 can be approximated by (37). Using this as the prior and y_2 as second stage data, we can further update the posterior distribution of θ by (37) and compare it with the exact posterior distribution Beta(15.5,47). Figure 6(a) shows the exact and approximate posterior densities of θ given y_1; the two densities are close. Here the exact distribution is Beta(3.5,9) with mean 3.5/(3.5+9)=0.28 and mode (3.5-1)/(3.5+9-2)=0.238, and the MLE based on y_1 is 3/8=0.375. Figure 6(b) shows the densities of exact posterior distribution Beta(15.5,47) and our two-stage approximation. Here the MLE based on y is 15/58=0.259 and the mean of Beta(15.5,47) is 15.5/(15.5+47)=0.248. We also conducted more experiments and found that the approximation (37) may not perform well when the posterior mode in the previous data greatly differs from the MLE of second-stage data.

4.2.2 Two-parameter Logit Model

To update posterior density for two-parameter problems in the two-stage scenario, we need to have an approximate joint posterior density at the first stage, which will be considered as the prior in the second stage. Such an approximation can be obtained by (35) with posterior moments replaced by their approximations; that is,

$$\begin{aligned}\hat{\xi}^t(\theta_1,\theta_2) \;=\; & |\Sigma_t|\phi_2(z^*)[1+q_1(z_1^*)\hat{E}_\xi^t Z_{t_1}+q_1(z_2^*)\hat{E}_\xi^t Z_{t_2}+\frac{1}{2}q_2(z_1^*)(\hat{E}_\xi^t Z_{t1}^2-1) \\ & +q_1(z_1^*)q_1(z_2^*)(\hat{E}_\xi^t Z_{t1}Z_{t2})+\frac{1}{2}q_2(z_2^*)(\hat{E}_\xi^t Z_{t2}^2-1) \\ & +\frac{1}{6}q_3(z_1^*)(\hat{E}_\xi^t Z_{t1}^3-3\hat{E}_\xi^t Z_{t1})+\frac{1}{2}q_2(z_1^*)q_1(z_2^*)(\hat{E}_\xi^t Z_{t1}^2 Z_{t2}-\hat{E}_\xi^t Z_{t2}) \\ & +\frac{1}{2}q_1(z_1^*)q_2(z_2^*)(\hat{E}_\xi^t Z_{t1}Z_{t2}^2-\hat{E}_\xi^t Z_{t1})+\frac{1}{6}q_3(z_2^*)(\hat{E}_\xi^t Z_{t2}^3-3\hat{E}_\xi^t Z_{t2})].\end{aligned} \tag{42}$$

The moment approximations are in terms of the loglikelihood and the prior and their derivatives; see Weng (2010b) and Weng and Hsu (2011).

In the first stage, we consider the same data from Mendenhall et al. (1989) as in Section 4.1.2. The data D_1 contains 24 observations. We take flat priors on both θ_1 and θ_2 and obtain an approximate joint posterior density $\hat{\xi}^t(\theta_1,\theta_2)$ by (42). Next we use this approximation as prior and randomly select 18 observations from D_1 as the data of second stage, denoted as D_2. Then we update the posterior density of θ by (42). Since the exact posterior density does not have a closed form, we compare our two-stage approximation with MCMC results using MCMCpack (2010). The results are in Figure 7. We found that the two densities are rather close, but the approximation method may sometimes give negative values.

4.3 Logit Model

Two-parameter case. Consider the same logit model in Section 4.1.2. We take flat priors on both θ_1 and θ_2 and use the implied density (29) to validate convergence of MCMC samples. To begin, we run the MCMClogit function in an R package MCMCpack (2010) to this data and obtain the posterior of MCMC samples. The comparisons of the density of MCMC samples and the implied density (29) are in Figures 8 and 9, based on posterior samples of sizes 100 (burn-in=50) and 10000 (burn-in=5000), respectively. Here $[\Sigma_t]_{22}=23.25$,

$\hat{\theta}_{t2} = -0.085$, and the posterior means of $q_i(Z_{t2})$, $i = 1, \ldots, 10$ from the MCMC samples in Figure 9 are (-0.224, 0.198, -0.416, 0.329, -0.069, -0.646, 2.844, -3.699, -6.565, 36.080) and $\hat{R}$ is 1.0020 for θ_2. The sample sizes and burn-in are chosen so that one exhibits convergence, but one does not. As expected, Figure 9 with $s = 2$ and $s = 4$ shows nice agreement; however, for Figure 8, the posterior sample and the implied density (29) with $s = 2$, $s = 4$, $s = 20$ and $s = 27$ are disagreements and $\hat{R}$ is 1.2309 for θ_2. Now we take the standard normal priors on both θ_1 and θ_2 and use the implied density (29) to validate convergence of MCMC samples. The comparisons of the density of the posterior samples (size=10000, burn-in=5000) and the implied density (29) with $s = 2$ and $s = 4$ are in Figure 10. Here the posterior means of $q_i(Z_{t2})$, $i = 1, \ldots, 10$ from the MCMC samples are (1.507, 1.535, 0.095, -3.243 , -5.175, 4.035, 28.610, 22.206, -132.075, -336.186) and $\hat{R}$ is 1.0001 for θ_2. The posterior samples has converged, but the implied density (29) failed to diagnose for this example.

Three-parameter case. Next, we consider the data analyzed by Finney (1947). A logistic regression was fit to the data. The model is

$$\log\Big(\frac{p_i}{1-p_i}\Big) = \theta_1 + \theta_2 c_{1i} + \theta_3 c_{2i}, i = 1, \ldots, 39.$$

The residual deviance gives a χ^2-value of 29.772 with 36 degrees of freedom, indicating that the logistic model is quite adequate. The comparisons of the density of MCMC samples (size=10000, burn-in=5000) and the implied density (29) with $s = 2$ and $s = 4$ are in Figure 11. The implied density (29) failed to diagnose for this example. The reason might be that the posterior density has quite heavy tails. Here $[\Sigma_t]_{33} = 1.093$, $\hat{\theta}_{t3} = 2.649$, and the posterior means of $q_i(Z_{t3})$, $i = 1, \ldots, 10$ from the MCMC samples in Figure 11 are (0.557, 0.666, 1.886, 5.581, 18.701, 73.833, 318.033, 1444.880, 6826.201, 32712.78) and $\hat{R}$ is 1.0023 for θ_3.

As a remedy, we truncate some extreme values from the MCMC samples. The comparisons of the density of MCMC samples (size=10000, burn-in=5000) and the implied density (29) with $s = 2$ and $s = 4$ are in Figure 12. Using a larger s gives better results. Here the posterior means of $q_i(Z_{t3})$, $i = 1, \ldots, 10$ from the MCMC samples in Figure 12 are (0.399, 0.086, 0.025, -0.381, -1.335, -0.560, 6.126, 13.857, -11.231, -116.282).

We truncate some extreme values from the MCMC samples to get better results on logit3p model at present section. We have a further discussion on this example and propose a method that transforms Z_t to another pivotal quantity in chapter 5.

Now we take the standard normal priors on θ_1, θ_2, and θ_3, and use the implied density (29) to validate convergence of MCMC samples. Here the posterior means of $q_i(Z_{t3})$, $i = 1, \ldots, 10$ from MCMC samples are (-2.257, 4.217,-5.564, 1.478, 16.393, -43.238, 11.314, 241.204, -627.421, -488.959) and $\hat{R}$ is 1.0013 for θ_3. In this example the posterior samples have heavy tails so we truncate some samples from the MCMC samples. Here the posterior means of $q_i(Z_{t3})$, $i = 1, \ldots, 10$ from the MCMC samples are (-2.279, 4.299, -5.719, 1.509, 16.952, -45.244, 12.290, 253.552, -662.408, -522.589). The comparisons of the density of MCMC samples and the implied density (29) with $s = 2$, $s = 4$ and $s = 14$ are in Figure 13, based on posterior samples of sizes 10000 (burn-in=5000).

Six-parameter case. Next, we consider a six-parameter problem. Kutner et al. (2004) studied the strength of association between several risk factors and the duration of pregnancies based on 102 women. The risk factors were nutritional status (x_1), mother's age (categorized into three groups and represented by two indicator variables x_2 and x_3), history of alcohol use (x_4), and history of tobacco use (x_5). The response of interest, pregnancy duration, was originally categorized into three groups: preterm (less than 36 weeks), intermediate term (from 36 to 37 weeks), and full term (38 weeks or greater). Here we combine the first two categories (coded 1) and let the full term be the second group (coded 0), and then fit a logistic regression model to the data. For details, see Kutner et al. (2004). The model is

$$\log\Big(\frac{p_i}{1-p_i}\Big) = \theta_1 + \theta_2 x_{1i} + \theta_3 x_{2i} + \theta_4 x_{3i} + \theta_5 x_{4i} + \theta_6 x_{5i},\ i = 1, \ldots, 102.$$

We run MCMCpack (2010) to obtain posterior samples of sizes 10000 (burn-in=10000). Figure 14 shows that the implied density (29) with $s = 4$ is quite close to the MCMC result. For example, for θ_6 we have $[\Sigma_t]_{66} = 1.586$, $\hat{\theta}_{t6} = 2.309$, and the posterior means of $q_i(Z_{t6})$ for $i = 1, ..., 4$ from the MCMC samples in Figure 14 are (0.407, 0.237, 0.367, 0.610) and $\hat{R}$ is 1.0114. With these six numbers, we can capture the posterior density of θ_6 pretty well.

4.4 Poisson Regression Model

We consider a data taken from Hinde (1982). The explanatory variable is the length of each roll, and the response variable is the numbers of faults found in 32 rolls of fabric produced in a particular factory. A problem of interest is the number of faults to be proportional to the length of roll. We fit the data using the poisson regression model

$$\log(\mu_i) = \theta_1 + \theta_2 x_i$$

where x_i is the length of ith roll and μ_i is the mean of related response variable. The intercept θ_1 represents the log-mean for zero length roll, while the slope θ_2 represents the change in the log-mean of response variable for every unit increase in the covariate. The loglikelihood and the second partial derivatives are

$$\ell_t(\theta) = \theta_1 \sum_{i=1}^{t} y_i + \theta_2 \sum_{i=1}^{t} x_i y_i - \sum_{i=1}^{t} \exp(\theta_1 + \theta_2 x_i) + C,$$

$$\ell_{11}^{(2)} = -\sum_{i=1}^{t} \exp(\theta_1 + \theta_2 x_i), \ell_{12}^{(2)} = -\sum_{i=1}^{t} x_i \exp(\theta_1 + \theta_2 x_i), \ell_{22}^{(2)} = -\sum_{i=1}^{t} x_i^2 \exp(\theta_1 + \theta_2 x_i).$$

Now we take flat priors on both θ_1 and θ_2 and use the implied density (29) to validate convergence of MCMC samples. We run WinBUGS Lunn et al. (2000) to obtain the posterior MCMC samples of sizes 10000 (burn-in=5000). Figures 15(a2) and (b2) show that the implied density (29) with $s = 2$ and $s = 4$ is quite close to the density of posterior samples. Here we have $[\Sigma_t]_{22} = 3264.80$ and $\hat{\theta}_{t2} = 0.0019$, and the posterior means of $q_i(Z_{t2})$, $i = 1, ..., 4$ from the MCMC samples in Figure 15(b2) are $(0.052, -0.014, 0.047, 0.044)$ and $\hat{R}$ is 1.0004 for θ_2. Now we take the standard normal priors on both θ_1 and θ_2 and use the implied density (29) to validate convergence of MCMC samples. Figures 16(a2) and (b2) show that the implied density (29) with $s = 2$ and $s = 4$ is very close to the MCMC result. Here the posterior means of $q_i(Z_{t2})$, $i = 1, \ldots, 4$ from the MCMC samples in Figure 16(b2) are $(0.235, 0.021, -0.026, -0.080)$ and $\hat{R}$ is 1.0011 for θ_2.

Given sample size $t = 16$, we randomly draw 16 observations from Hinde (1982). As before, we take flat priors on both θ_1 and θ_2. Figures 15(a1) and (b1) show that the implied density (29) with $s = 2$ and $s = 4$ is quite close to the density of posterior samples. Here we have $[\Sigma_t]_{22} = 2487.49$ and $\hat{\theta}_{t2} = 0.00199$, and the posterior means of $q_i(Z_{t2})$, $i = 1, ..., 4$

from the MCMC samples in Figure 15(b1) are $(0.068, -0.030, 0.012, -0.034)$. Now we take the standard normal priors on both θ_1 and θ_2, Figures 16(a1) and (b1) show that the implied density (29) with $s = 2$ and $s = 4$ is very close to the MCMC result. Here the posterior means of $q_i(Z_{t2})$, $i = 1, \ldots, 4$ from the MCMC samples in Figure 16(b1) are $(0.319, 0.026, 0.004, 0.024)$.

4.5 Gamma Model

We consider the example of clotting time of blood given in Hurn et al. (1945); see also McCullagh and Nelder (1989). The data set consists of clotting times in seconds (y) for normal plasma diluted to nine different percentage concentrations with prothrombin-free plasma (ν); clotting was induced by two lots of thromboplastin.

Suppose that Y follows Gamma$(\alpha, 1/\beta)$ with mean $\alpha\beta$. McCullagh and Nelder (1989) take the link function to be inverse; that is,

$$\eta = \mu^{-1} = \frac{1}{\alpha\beta} = \theta_1 + \theta_2 x,$$

where $x = \log(\nu)$. The loglikelihood and second partial derivatives are

$$\ell_t(\theta) = \alpha \sum_{i=1}^{t} \log(\theta_1 + \theta_2 x_i) - \alpha \sum_{i=1}^{t} y_i(\theta_1 + \theta_2 x_i) + C,$$

$$\ell_{11}^{(2)} = -\sum_{i=1}^{t} \frac{\alpha}{(\theta_1 + \theta_2 x_i)^2},\ \ell_{12}^{(2)} = -\sum_{i=1}^{t} \frac{\alpha x_i}{(\theta_1 + \theta_2 x_i)^2},\ \ell_{22}^{(2)} = -\sum_{i=1}^{t} \frac{\alpha x_i^2}{(\theta_1 + \theta_2 x_i)^2}.$$

Since the MCMCpack (2010) does not provide GLM Gamma, we use WinBUGS Lunn et al. (2000). However, GLM Gamma with inverse link produces negative values of η, so we consider the identity link function; that is,

$$\eta = \mu = \alpha\beta = \theta_1 + \theta_2 x,$$

where $x = 1/\log(\nu)$. The loglikelihood and second partial derivatives are

$$\ell_t(\theta) = -\alpha \sum_{i=1}^{t} \log(\theta_1 + \theta_2 x_i) - \alpha \sum_{i=1}^{t} [y_i/(\theta_1 + \theta_2 x_i)] + C,$$

$$\ell_{11}^{(2)} = \alpha \left[(-1)^2 \sum_{i=1}^{t} \frac{1}{(\theta_1 + \theta_2 x_i)^2} + (-1)^3\, 2! \sum_{i=1}^{t} \frac{y_i}{(\theta_1 + \theta_2 x_i)^3} \right],$$

$$\ell_{12}^{(2)} = \alpha \left[(-1)^2 \sum_{i=1}^{t} \frac{x_i}{(\theta_1 + \theta_2 x_i)^2} + (-1)^3 \, 2! \sum_{i=1}^{t} \frac{x_i y_i}{(\theta_1 + \theta_2 x_i)^3} \right],$$

$$\ell_{22}^{(2)} = \alpha \left[(-1)^2 \sum_{i=1}^{t} \frac{x_i^2}{(\theta_1 + \theta_2 x_i)^2} + (-1)^3 \, 2! \sum_{i=1}^{t} \frac{x_i^2 y_i}{(\theta_1 + \theta_2 x_i)^3} \right].$$

We take flat priors on both θ_1 and θ_2 and use the implied density (29) to validate convergence of MCMC samples. With WinBUGS Lunn et al. (2000) we obtain the posterior MCMC samples of sizes 10000 (burn-in=20000). In this example, the number of burn-in needs to be large to converge. Figures 17(a2) and (b2) show that the implied density (29) with $s = 2$ and $s = 4$ is close to the density of MCMC samples. Here $[\Sigma_t]_{22} = 0.01247$ and $\hat{\theta}_{t2} = 154.4327$, and the posterior means of $q_i(Z_{t2})$, $i = 1, ..., 4$ from the MCMC samples in Figure 17(b2) are $(0.450, 0.290, 0.699, 1.128)$ and $\hat{R}$ is 1.0111 for θ_2. Given sample size $t = 9$, we randomly draw 9 observations from Hurn et al. (1945). Figures 17(a1) and (b1) show that the implied density (29) with $s = 2$ and $s = 4$ is close to the density of MCMC samples. Here $[\Sigma_t]_{22} = 0.00713$ and $\hat{\theta}_{t2} = 194.6036$, and the posterior means of $q_i(Z_{t2})$, $i = 1, ..., 4$ from the MCMC samples in Figure 17(b1) are $(0.315, 0.250, 1.143, 2.170)$.

4.6 Mixture Normal

In this section we consider mixture normal models where the posterior density of the parameter may have multimodes. Suppose that a vector of observations $y = (y_1, ..., y_t)$ is drawn from a mixture distribution of two normals:

$$p(y|\theta) = w_1 \phi(y; \theta_1, \sigma_1^2) + w_2 \phi(y; \theta_2, \sigma_2^2), \tag{43}$$

where $\phi(y; \theta, \sigma^2)$ is normal density with mean θ and variance σ^2, $w_1 + w_2 = 1$, w_1 and θ_1 are unknown parameters of interest, θ_2, σ_1^2, and σ_2^2 are assumed known. The conjugate prior distributions are $(w_1, w_2) \sim \text{Beta}(\alpha, \beta)$ and $\theta_i \sim \phi(\mu_i, \tau_i^2)$ for $i = 1, 2$. For detailed discussions of mixture distributions, see Gilks et al. (1995, chapter 24) and Minka (2001). It is convenient to introduce unobserved indicator variables ζ_i with

$$\zeta_i = \begin{cases} 1 & \text{if the } i\text{th observed is drawn from } \phi(\theta_1, \sigma_1^2) \\ 0 & \text{otherwise.} \end{cases}$$

Given w_1, each unobserved variable ζ_i follows Bernoulli distribution with mean w_1.

We take $\theta_2 = 0$ and $\sigma_1^2 = \sigma_2^2 = 1$, and assume that w_1 has prior distribution Beta$(1,1)$. Let θ_1 have a prior density $\phi(0, 100)$. The posterior density of (θ_1, ζ, w_1) given y is

$$\begin{aligned}
p(\theta_1, \zeta, w_1|y) &\propto p(\theta_1, \zeta, w_1, y) \\
&\propto p(\theta_1, w_1)p(y, \zeta|\theta_1, w_1) \\
&\propto p(\theta_1)p(w_1)p(\zeta|w_1)p(y|\theta_1, \zeta) \\
&\propto \exp(-\theta_1^2/200)\prod_i \left[w_1\phi(y_i;\theta_1,1)\right]^{\zeta_i}\left[(1-w_1)\phi(y_i;0,1)\right]^{1-\zeta_i}. \qquad (44)
\end{aligned}$$

Given (θ_1, w_1, y), by (44),

$$p(\zeta|\theta_1, w_1, y) \propto \prod_i \left[w_1\phi(y_i;\theta_1,1)\right]^{\zeta_i}\left[(1-w_1)\phi(y_i;0,1)\right]^{1-\zeta_i}.$$

So

$$p(\zeta_i = 1|\theta_1, w_1, y) \propto w_1\phi(y_i;\theta_1,1), \;\; p(\zeta_i = 0|\theta_1, w_1, y) = (1-w_1)\phi(y_i;0,1);$$

where

$$r_i = w_1\phi(y_i,\theta_1,1)/[w_1\phi(y_i,\theta_1,1) + (1-w_1)\phi(y_i,0,1)]. \qquad (45)$$

So, the full conditional distribution for ζ_i is Bernoulli distribution with mean r_i. To construct the full conditionals for w_1 and θ_1, by (44),

$$p(w_1|y, \theta_1, \zeta) \;\; \propto \;\; w_1^{\sum\zeta_i}(1-w_1)^{t-\sum\zeta_i};$$

and

$$\begin{aligned}
p(\theta_1|y, w_1, \zeta) &\propto \exp(-\theta_1^2/200)\prod_i\left[\phi(y_i;\theta_1,1)^{\zeta_i}\right] \\
&\propto \exp(-\theta_1^2/200)\exp\left(-\frac{\sum\zeta_i(y_i-\theta_1)^2}{2}\right) \\
&\propto \exp\left(-\frac{1}{2}\left[(\frac{1}{100}+\sum\zeta_i)(\theta_1 - \frac{\sum\zeta_i y_i}{\frac{1}{100}+\sum\zeta_i})^2\right]\right).
\end{aligned}$$

So, the full conditional for w_1 is a Beta distribution, and the full conditional for θ_1 is a normal distribution with mean $b/(1/100+h)$ and variance $1/(1/100+h)$, where $h = \sum\zeta_i$ and $b = \sum\zeta_i y_i$. The Gibbs sampler is easy to apply for the mixture normals because the full conditional posterior distributions - $\pi(\zeta|\theta_1, w_1, y)$, $\pi(w_1|\theta_1, \zeta, y)$, $\pi(\theta_1|w_1, \zeta, y)$ -

have standard forms and can be easily sampled from. One cycle of the Gibbs sampler is described below.

Step 1 -Simulate ζ_i :

$$(\zeta_i|\theta_1, w_1, y) \sim \text{Bernoulli}(r_i), \text{ for } i = 1, 2, \ldots, t;$$

where r_i is in (45).

Step 2 -Simulate w_1 :

$$(w_1|\theta_1, \zeta, y) \sim \text{Beta}(h+1, t+1-h),$$

where $h = \sum \zeta_i$.

Step 3 -Simulate θ_1 :

$$(\theta_1|w_1, \zeta, y) \sim \phi\left(\frac{b}{\frac{1}{100}+h}, \frac{1}{\frac{1}{100}+h}\right),$$

where $b = \sum \zeta_i y_i$.

The loglikelihood and the second partial derivatives are

$$\ell_t(\theta) = \sum_{i=1}^{t} \log\left[w_1\phi(y_i; \theta_1, 1) + (1-w_1)\phi(y_i; 0, 1)\right],$$

where $\theta = (w_1, \theta_1)$.

$$\ell_{11}^{(2)} = -\sum_i \frac{[\phi(y_i;\theta_1,1) - \phi(y_i;0,1)]^2}{f^2}, \ell_{12}^{(2)} = -\sum_i \frac{(y_i-\theta_1)\phi(y_i;\theta_1,1)\phi(y_i;0,1)}{f^2},$$

$$\ell_{22}^{(2)} = -\sum_i \left\{\frac{w_1(1-w_1)[(y_i-\theta_1)^2-1]\phi(y_i;\theta_1,1)\phi(y_i;0,1) - w_1^2\phi(y_i;\theta_1,1)^2}{f^2}\right\}.$$

where $f = w_1\phi(y_i;\theta_1,1) + (1-w_1)\phi(y_i;0,1)$. In this simulation we set $w_1 = 0.5$ and $\theta_1 = 2$, and generate a data of sample size 20. Then, we take the prior of θ_1 as N(0,100) and use the implied density (29) to validate convergence of MCMC samples. We run R function to obtain posterior samples of θ_1 by Gibbs sampler with sample sizes 5000 (burn-in=5000). The initial value for Gibbs sampler is $(\theta_1^{(0)}, w_1^{(0)}) = (-2, 0.5)$. Figure 18 shows that the implied density (29) with $s = 2$ and $s = 4$ is very close to the MCMC result. For this data, the posterior distribution has a single mode and the posterior means of $q_i(Z_{t2})$ with $Z_{t2} = [\Sigma_t]_{22}(\theta_1 - \hat{\theta}_{t1})$, $i = 1, \ldots, 4$ from the Gibbs samples are $(-0.00076, 0.0786, 0.0184, 0.1693)$ and $\hat{R}$ is 1.0024 for θ_1. Next we let $\sigma_2^2 = 10$, meanwhile generating another data of sample size 20. We obtain the posterior samples of θ_1 by Gibbs sampler with sample sizes 5000 (burn-in=5000). The initial value for Gibbs sampler is also $(\theta_1^{(0)}, w_1^{(0)}) = (-2, 0.5)$. The posterior samples have two modes. Figure 19(a) shows that the density of MCMC samples is close to the exact density by numerical integration and 19(b) shows the implied density (29) failed to diagnose for this example. The posterior means of $q_i(Z_{t2})$, $i = 1, \ldots, 4$ from the Gibbs samples are $(-2.643, 86.271, -925.97, 5850.66)$ and $\hat{R}$ is 1.0016 for θ_1.

We have a further discussion on this example and propose a method that transforms Z_t to another pivotal quantity in chapter 5.

5 Concluding Remarks

We have compared second order Bayesian asymptotics and found some agreements and some disagreements. We found that our $O(t^{-1/2})$ term is arithmetically equivalent to Johnson's, but the $O(t^{-1})$ term is not. Since the derivation is tedious and difficult to detect errors, simulation studies are conducted to further compare these expansions. The simulations confirmed that the two expressions for $O(t^{-1/2})$ term yield close results, and revealed that our $O(t^{-1})$ term gives better performance than Johnson's. Note that the emphasis here is on comparison of the two expansions, rather than the regularity conditions for the expansions. We also conduct simulations to compare the order $O(t^{-3/2})$ expansions with Tierney and Kadane (1986) result.

Since the asymptotic posterior distribution depends on observed data, we try different samples to see its effect on the asymptotic approximations. Here we have a further discussion from both analytical and numerical (experimental) result.

Analytical results. In the appendix of Tierney and Kadane (1986), Laplace's method provides an approximation for integrals of the form $\int e^{tL(\theta)} d\theta$ when t is large. If L has a unique maximum at $\hat{\theta}$ and $\sigma^2 = -1/\nabla^2 L(\hat{\theta})$, then

$$\int e^{tL(\theta)} d\theta = \sqrt{2\pi}\sigma t^{-1/2} e^{tL(\hat{\theta})} (1 + \frac{a}{t} + \frac{b}{t^2} + O(t^{-3})), \tag{46}$$

where, setting $L_k = \nabla^k L(\hat{\theta})$, the constants a and b are given by

$$a = \frac{1}{8}\sigma^4 L_4 + \frac{5}{24}\sigma^6 L_3^2$$

and

$$b = \frac{1}{48}\sigma^6 L_6 + \frac{35}{384}\sigma^8 L_4^2 + \frac{7}{48}\sigma^8 L_3 L_5 + \frac{35}{64}\sigma^{10} L_3^2 L_4 + \frac{385}{1152}\sigma^{12} L_3^4.$$

Result (46) remains valid if L is replaced by a sufficiently well-behaved sequence $L^{(t)}$ of functions. In this case the coefficients a and b may depend on t, but this dependence will be suppressed. If a and b do indeed depend on t, we will assume regularity conditions for the sequence $L^{(t)}$ that insure that a and b are bounded in t.

In Weng (2010a) (see Theorem 2, page 752), to ensure the marginal posterior density (Eq(23)), the following condition is required:

(A1) For each $r > 0$, $E_\xi^t(\|Z_t\|^r) = O(1)$.

Here $O(1)$ means convergence of a sequence of real numbers as $t \to \infty$. However, it is difficult to judge $O(1)$ from the posterior moments of Z_t.

Numerical (experimental) result. Consider the same logit model in Section 4.1.2. Given sample sizes $t = 12$ and $t = 15$, we randomly draw three samples from 24 observations, respectively, and the comparisons of the density (Eq(23)) with exact density (numerical integration) are in Figures 20 and 21 with $s = 2$. Figure 20 contains the results of three different samples for $t = 12$. The posterior means of the three samples Z_{t2}^r, $r = 1, \ldots, 10$ from numerical integration are in Table 1.1, where the seventh to tenth posterior moments of Z_{t2} for the three samples are large and it does not seem to satisfy the condition (A1). The approximations are not well in Figures 20(a), (b) and (c).

Figure 21 contains the results of three different samples for $t = 15$. The posterior means of the three samples Z_{t2}^r, $r = 1, \ldots, 10$ from numerical integration are in Table 1.2, where the seventh to tenth posterior moments of Z_{t2} for the three samples are also large and it does not seem to satisfy the condition (A1). In this example we found that the posterior moments of the first sample are relatively smaller than the other two samples. So, the approximations are not well in Figures 21(b) and (c); however, the approximation is well in Figure 21(a).

Now we consider the same observed data from the previous paragraph. The comparisons of the density (Eq(23)) and the MCMC density with both $s = 2$ and $s = 7$ are in Figures 22 and 23, where the sample sizes are $t = 12$ and $t = 15$, respectively. The posterior means of the three samples Z_{t2}^r, $r = 1, \ldots, 6$ from MCMC samples are in Table 2.1, where the fifth to sixth posterior moments of Z_{t2} in the first sample are larger than the other two samples. The first sample does not seem to satisfy the condition (A1) and the approximation is not well in Figure 22-1. However, the posterior moments of the second and third samples are relatively smaller than the first sample and the approximations are well in Figures 22-2 and 22-3.

The posterior means of the three samples Z_{t2}^r, $r = 1, \ldots, 6$ from MCMC samples are in Table 2.2, where the fifth to sixth posterior moments of Z_{t2} in the third sample are larger than the other two samples. The third sample does not seem to satisfy the condition (A1) and the approximation is not well in Figure 23-3. However, the posterior moments of the first and second samples are relatively smaller than the third sample and the approximations are well in Figures 23-1 and 23-2.

	posterior moments (exact density)									
data	Z_{t2}	Z_{t2}^2	Z_{t2}^3	Z_{t2}^4	Z_{t2}^5	Z_{t2}^6	Z_{t2}^7	Z_{t2}^8	Z_{t2}^9	Z_{t2}^{10}
1	-0.55	1.76	-3.53	11.94	-37.29	139.1	-521.6	2100.7	-8596.0	36413.0
2	-0.44	1.58	-2.68	9.40	-26.51	99.36	-351.4	1406.9	-5562.5	23389.6
3	-0.27	1.43	-1.72	7.50	-17.20	72.31	-226.5	966.7	-3549.9	15526.2

Table 1.1 sample size $t = 12$.

	posterior moments (exact density)									
data	Z_{t2}	Z_{t2}^2	Z_{t2}^3	Z_{t2}^4	Z_{t2}^5	Z_{t2}^6	Z_{t2}^7	Z_{t2}^8	Z_{t2}^9	Z_{t2}^{10}
1	-0.19	1.28	-1.10	5.74	-10.14	47.47	-124.4	565.8	-1845.8	8401.6
2	-0.54	1.70	-3.44	11.34	-35.62	130.9	-491.2	1961.0	-8013.8	33771.0
3	-0.75	2.14	-5.39	18.01	-62.90	239.6	-949.5	3908.4	-16496.7	71105.2

Table 1.2 sample size $t = 15$.

	posterior moments (MCMC density)					
data	Z_{t2}	Z_{t2}^2	Z_{t2}^3	Z_{t2}^4	Z_{t2}^5	Z_{t2}^6
1	-0.51	1.75	-3.56	13.26	-47.28	208.5
2	-0.39	1.54	-2.29	8.37	-20.85	79.53
3	-0.24	1.36	-1.48	6.46	-13.14	54.34

Table 2.1 sample size $t = 12$.

	posterior moments (MCMC density)					
data	Z_{t2}	Z_{t2}^2	Z_{t2}^3	Z_{t2}^4	Z_{t2}^5	Z_{t2}^6
1	-0.14	1.20	-0.79	4.67	-6.23	32.96
2	-0.52	1.64	-3.22	10.59	-33.25	123.8
3	-0.78	2.45	-7.41	30.61	-142.1	743.2

Table 2.2 sample size $t = 15$.

In Section 4.2.1, we have described the binomial model $Y \sim \text{Bin}(t, \theta)$ with prior Beta(α, β) of θ for multi-stage data and compared the posterior density (Eq(37)) with the exact distribution. Now we try different priors to see its effect on the approximations. We compared the exact posterior distribution of θ given the first-stage data y_1 with posterior density (Eq(37)) for four different priors; the results are in Figures 24(a), (b), (c) and (d), respectively. Next, we try four different second-stage data for each of the settings in Figures 24(a), (b), (c) and (d). The results are in Figures 25 to 28. We have some observations from Figure 25 to Figure 28. First, the approximated posterior density (Eq(37)) is near the exact distribution when the MLE of second-stage data does not differ much from the posterior mode in the previous data, see Figures 25(a)(b), 26(a)(b), 27(a)(b) and 28(a)(b). Secondly, the approximation (Eq(37)) may not perform well and may have some negative values when the posterior mode in the previous data greatly differs from the MLE of second-stage data; see Figures 25(c), 26(c), 27(c) and 28(c). Thirdly, the problem of negative values by approximated posterior density (Eq(37)) may be improved when the sample size increases; see Figures 25(d), 26(d), 27(d) and 28(d). Our experiments on multi-stage data showed that the analytic approximations are reasonably well when the MLE of second-stage data does not differ much from the posterior mode in the previous data. However, when the MLE of second-stage data greatly differs from the posterior mode in the previous data, the approximations may be not well; the larger the sample size is, the better the result will be.

We also proposed a graphical method for validating convergence of MCMC. Our method is based on a Bayesian Edgeworth expansion for the posterior distribution. The method has been tested on some GLM models and mixture normal models, but the results are mixed. One problem is that in some occasions the posterior densities seem to have heavier tails. In fact, for a Bayesian Edgeworth expansion to be valid, the posterior moments need to be finite. So, intuitively a density with heavy tails may cause some problems. However, in a simulation sample, all moments can be calculated and are finite; therefore, it is difficult to judge whether the real moments are finite or not.

When the posterior moments of Z_t are large, the asymptotic result is often not well. We transform Z_t to $W_t = w * Z_t$, where $0 < w < 1$. Since W_t has smaller moments, we expect the asymptotic result for W_t to be better. By (22), Z_{tp} involves only θ_p so that $P^t_\xi(\theta_p \le a_p) = P^t_\xi(Z_{tp} \le z^*_p)$ for $z^*_p = [\Sigma_t]_{pp}(a_p - \hat{\theta}_{tp})$. Let $\varphi_p = w\theta_p$. We obtain $P^t_\xi(\varphi_p \le wa_p) = P^t_\xi(W_{tp} \le w^*_p)$ and $P^t_\xi(W_{tp} \le w^*_p) = P^t_\xi(Z_{tp} \le w^*_p/w)$, where $w^*_p = wz^*_p = [\Sigma_t]_{pp}(wa_p - w\hat{\theta}_{tp})$. Then, by (21) and (23), we derive the marginal posterior density for φ_p:

$$\xi^t_\varphi(b_p) = \frac{1}{w}[\Sigma_t]_{pp}\left\{\phi(\frac{w^*_p}{w}) + \sum_{\substack{k\in\{1,\dots,3s\}\\ k\neq 3s-1}} \frac{1}{k!}q_k(\frac{w^*_p}{w})\phi(\frac{w^*_p}{w})E^t_\xi(q_k(\frac{W_{tp}}{w})) + O(t^{-\frac{3s+1}{2}+s})\right\}, \quad (47)$$

where $b_p = wa_p$. Two examples below are illustrated.

Logit model (three-parameter case.) Consider the same logit model in Section 4.3. Here we take the posterior MCMC samples of parameters with sample sizes 10000 (burn-in=5000). Now we take $w = 0.5$ and to compare the posterior density (Eq(47)) with MCMC densities in Figure 29, where the posterior means of $q_i(W_{t3})$, $i = 1, \dots, 6$ from the MCMC samples are (0.2786, -0.5833, -0.3912, 1.2865, 1.1674, -4.8815). The results are well.

Mixture model. Consider the same mixture model in Section 4.6. Here we take the posterior samples of θ_1 with sample sizes 5000 (burn-in=5000) and $Z_{t2} = [\Sigma_t]_{22}(\theta_1 - \hat{\theta}_{t1})$. Now we take $w = 0.5$, $w = 0.1$ and to compare the posterior density (Eq(47)) with MCMC densities in Figures 30(a) and (b). In Figure 30(a), the posterior means of $q_i(W_{t2})$, $i = 1, \ldots, 6$ from the MCMC samples are (-1.3218, 20.8179, -112.77, 3561.29, -31987.67, 1253357). The result is not well. In Figure 30(b), the posterior means of $q_i(W_{t2})$, $i = 1, \ldots, 6$ from the MCMC samples are (-0.2643, -0.1272, -0.1407, 3.666, -5.229, 19.425) and we obtain better results for larger s.

When the numbers of both burn-in and MCMC sample are small, the asymptotic result for Z_t is not well. We transform Z_t to W_t to see its effect on the asymptotic result. Some examples below are illustrated.

Logit model (two-parameter case.) Consider the same logit model in Section 4.3. Here we take the posterior MCMC samples of parameters with sample sizes 100 (burn-in=50). Next we take $w = 0.1$ and to compare the posterior densities (Eq(47)) with MCMC densities in Figures 31(a) with $s = 10$ and 31(b) with $s = 20$. The posterior means of $q_i(W_{t2})$, $i = 1, \ldots, 4$ from the MCMC samples are $(-0.0445, -0.9823, 0.1291, 2.895)$. The approximations are not well.

Poisson model. Consider the same Poisson model in Section 4.4. Here we take the posterior MCMC samples of parameters with sample sizes 100 (burn-in=50). The posterior means of $q_i(Z_{t2})$, $i = 1, \ldots, 4$ from the MCMC samples are $(-0.1730, 0.2623, -1.004, 1.0059)$ and $\hat{R}$ is 1.5873 for θ_2. The comparisons of the density (Eq(47)) with the density (MCMC) are in Figures 32(a) with $s = 10$ and 32(b) with $s = 20$. The approximations are not well. Now we take $w = 0.1$ and to compare the posterior densities (Eq(47)) with MCMC densities in Figure 33. The posterior means of $q_i(W_{t2})$, $i = 1, \ldots, 4$ from the MCMC samples are $(-0.0173, -0.9873, 0.0503, 2.9248)$. The approximations are not well.

Mixture model. Consider the same mixture model in Section 4.6. Here we take the posterior samples of θ_1 with sample sizes 50 (burn-in=50). The posterior means of $q_i(Z_{t2})$, $i = 1, \ldots, 4$ from the MCMC samples are $(0.84976, 1.8444, 8.3110, 40.6077)$ and $\hat{R}$ is 1.1375 for θ_1. The posterior samples have two modes. Figure 34(a) shows that the density (MCMC) is not close to the exact density (numerical integration). The comparisons of the density (Eq(47)) with the density (MCMC) are in Figures 34(b) with $s = 7$ and 34(c) with $s = 27$. The approximations are not well. Now we take $w = 0.1$ and to compare the posterior densities (Eq(47)) with MCMC densities in Figure 35. The posterior means of $q_i(W_{t2})$,

$i = 1, \ldots, 4$ from the MCMC samples are $(0.0849, -0.9715, -0.2440, 2.834)$. The approximations are not well.

References

Albert, J. H. and Chib, S. (1993), "Bayesian analysis of binary and polychotomous response data", *Journal of the American Statistical Association*, 88, 669–679.

Brooks, S. P. and Gelman, A. (1998), "General methods for monitoring convergence of iterative simulations", *Journal of Computational and Graphical Statistics*, 7, 434–455.

Dellaportas, P. and Smith, A. F. M. (1993), "Bayesian inference for generalized linear and proportional hazards models via gibbs sampling", *Appl. Statist.*, 42, 443–460.

Erdelyi, A. (1956), *Asympotic Expansions*, New York: Dover Publications.

Finney, D. J. (1947), "The estimation from individual records of the relationship between dose and quantal response", *Biometrika*, 34, 320–334.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (1995), *Bayesian data analysis*, New York : Chapman and Hall.4.

Gelman, A. and Rubin, D. B. (1992), "Inference form iterative simulation using multiple sequences", *Statistical Science*, 7, 457–511.

Geman, S. and Geman, D. (1984), "Stochastic relaxation, gibbs distributions and the bayesian restoration of images", *IEEE Trans. Pattn. Anal. Mach. Intel*, 6, 721–741.

Gilks, W. R., Richardson, S., and Spiegelhalter, D. J. (1995), *Markov Chain Monte Carlo in Practice*, London: Chapman and Hall.

Gilks, W. R. and Wild, P. (1992), "Adaptive rejection sampling for gibbs sampling", *Appl. Statist.*, 41, 337–348.

Glickman, Mark E. (1993), "Paired comparison models with time-varying parameters", Ph.D. thesis, Department of Statistics, Harvard University.

Hinde, J. (1982), *Compound Poisson regression models*, New York: Springer.

Hurn, M. W., Barker, N. W., and Magath, T. D. (1945), "The determination of prothrombin time following the administration of dicumarol with specific reference to thromboplastin", *Med.*, 30, 432–447.

Johnson, R. (1967), "An asymptotic expansion for posterior distributions", *Ann. Math. Statist.*, 38, 1899–1906.

——— (1970), "Asymptotic expansions associated with posterior distributions", *Ann. Math. Statist.*, 41, 851–864.

Kutner, M. H., Nachtsheim, C. J., and Neter, J. (2004), *Applied linear regression models*, Boston: McGraw.

Lunn, D.J., Thomas, A., Best, N., and Spiegelhalter, D. (2000), "Winbugs – a bayesian modelling framework: concepts structure and extensibility", *Statistics and Computing*, 10, 325–337.

Martin, Andrew D., Quinn, Kevin M., and Park, Jong Hee (2010), "MCMCpack: Markov chain Monte Carlo (MCMC) Package", R package version 2.10.0, URL `http://CRAN.R-project.org/package=MCMCpack`.

McCullagh, P. and Nelder, J. A. (1989), *Generalized linear models*, New York : Chapman and Hall.

Mendenhall, W. M., Parsons, J. T., Stringer, S. P., Cassissi, N. J., and Million, R. R. (1989), "T2 oral tongue carcinoma treated with radiotherapy: analysis of local control and complications", *Radiotherapy and Oncology*, 16, 275–282.

Minka, T. P. (2001), "A family of algorithms for approximate bayesian inference", *Department of Electrical Engineering and Computer Science*, 1–75.

Myers, R. H. (1990), *Classical and Modern Regression with Applications*, Boston: PWS-Kent.

R Development Core Team (2010), "R: A language and environment for statistical computing", R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0, URL `http://www.R-project.org`.

Tanner, M. A. (1993), *Tools for Statistical Inference*, New York: Springer.

Tierney, L. and Kadane, J. B. (1986), "Accurate approximations for posterior moments and marginal densities", *Journal of the American Statistical Association*, 81, 82–86.

Weng, R. C. (2003), "On Stein's identity for posterior normality", *Statistica Sinica*, 13, 495–506.

——— (2010a), "A Bayesian Edgeworth expansion by Stein's Identity", *Bayesian Analysis*, 5, 741–764.

——— (2010b), "A note on posterior distributions", in *JSM Proceedings, Section on Bayesian Statistical Science*, 2599–2608, Alexandria, VA: American Statistical Association.

Weng, R. C. and Hsu, C.-H. (2011), "A study of expansions of posterior distributions", *Communications in Statistics - Theory and Methods.*

Weng, R. C. and Tsai, W.-C. (2008), "Asymptotic posterior normality for multiparameter problems", *Journal of Statistical Planning and Inference*, 138, 4068–4080.

Woodroofe, M. (1989), "Very weak expansions for sequentially designed experiments: linear models", *Ann. Statist.*, 17, 1087–1102.

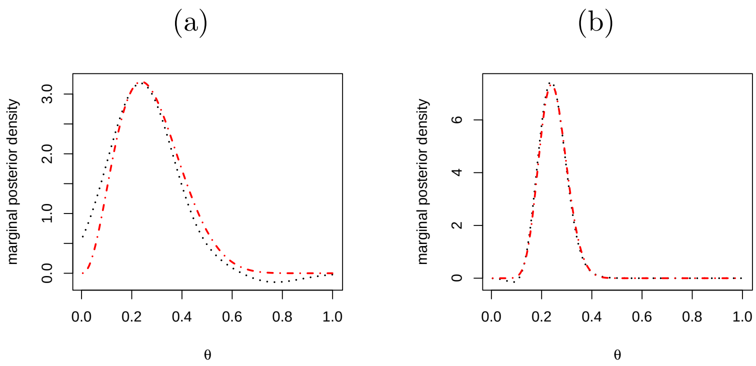


Figure 6: Marginal posterior pdf of θ. Beta-Binomial model.

Sequential update. Dotted: Equation (37); Dot-dashed: Exact distribution.

(a)$(t_1, y_1) = (8, 3), p(\theta|y_1)$.

(b)$(t_1, y_1) = (8, 3), (t_2, y_2) = (50, 12)$ two stage approximation of $p(\theta|y_1, y_2)$.

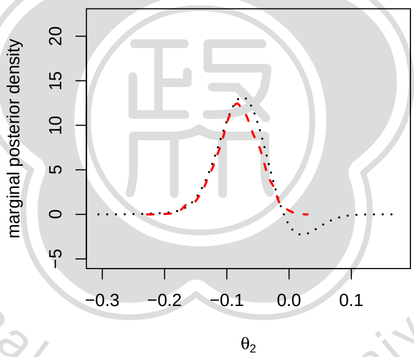


Figure 7: Marginal posterior pdf of θ_2. Sequential update for logit model. Dotted: Equation (37); Dashed: MCMC(burn-in=5000, mcmc=10000) as Exact.

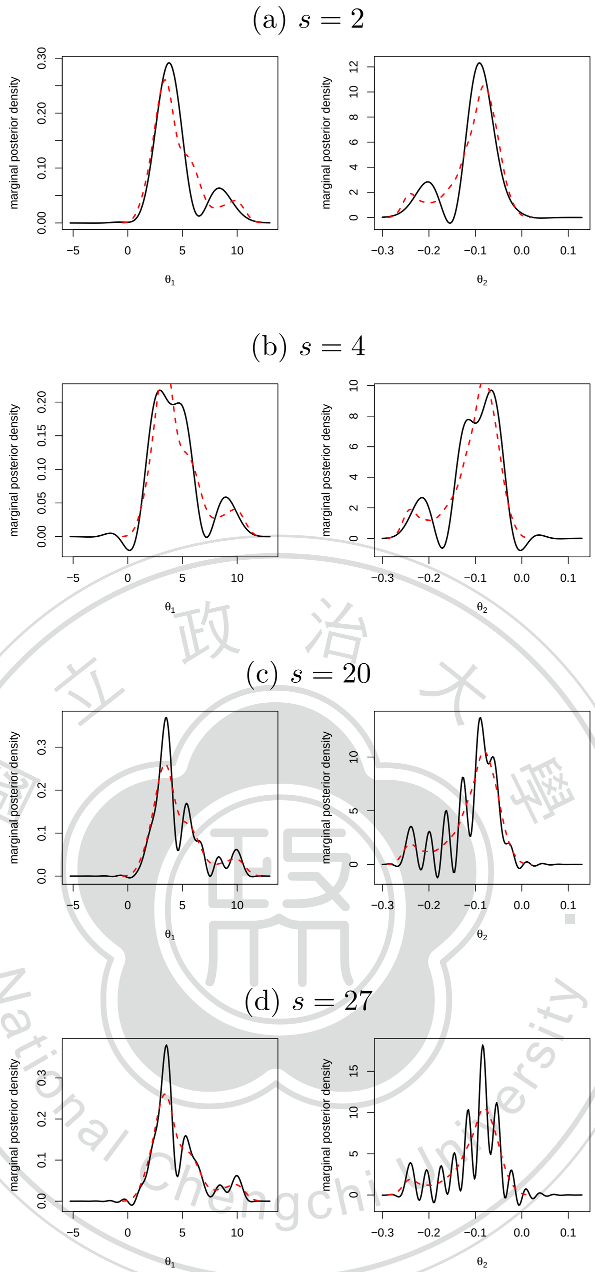


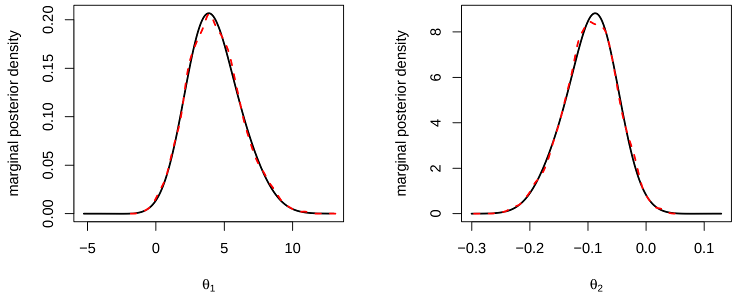
Figure 8: Marginal posterior pdf of θ_1 and θ_2. Logit model with flat prior.

Dashed: MCMC(burn-in=50, mcmc=100);

Solid: Equation (29) with $s = 2$, $s = 4$, $s = 20$ and $s = 27$.

$\hat{R}$ is 1.2309 for θ_2.

(a) $s = 2$



(b) $s = 4$

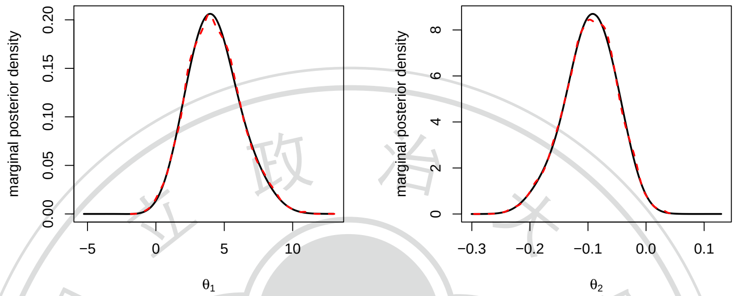


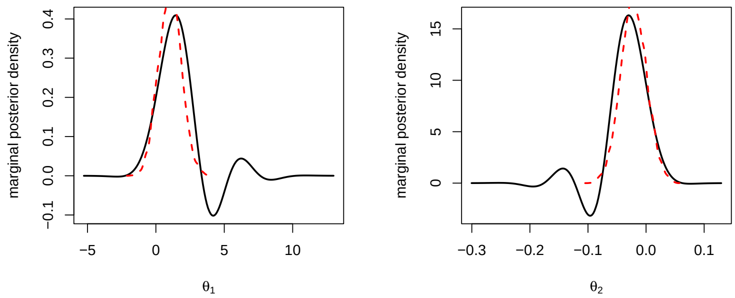
Figure 9: Marginal posterior pdf of θ_1 and θ_2. Logit model with flat prior.

Dashed: MCMC(burn-in=5000, mcmc=10000);

Solid: Equation (29) with $s = 2$ and $s = 4$.

$\hat{R}$ is 1.0020 for θ_2.

(a) $s = 2$



(b) $s = 4$

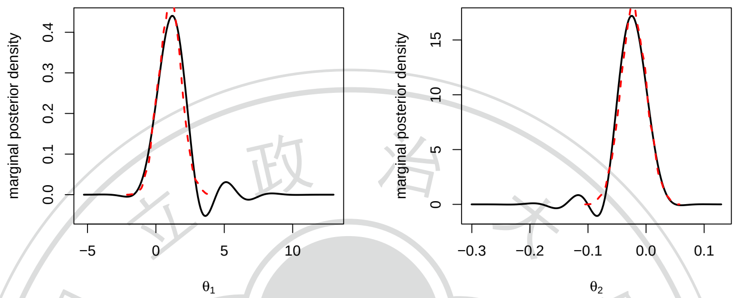


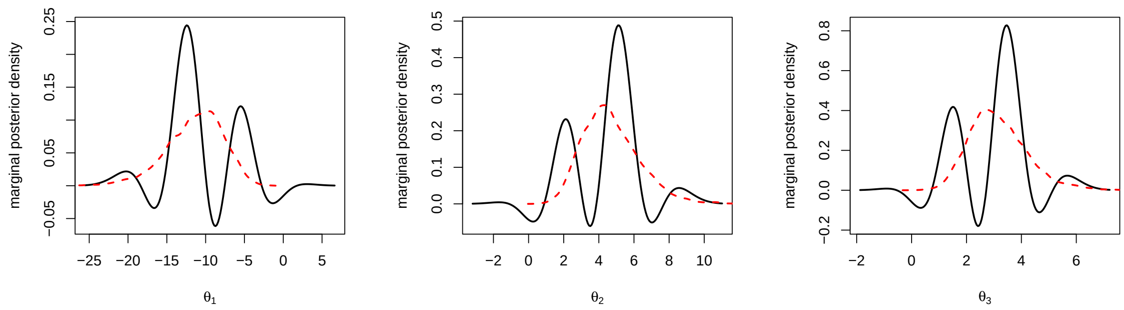
Figure 10: Marginal posterior pdf of θ_1 and θ_2. Logit model with N(0,1) prior.

Dashed: MCMC(burn-in=5000, mcmc=10000);

Solid: Equation (29) with $s = 2$ and $s = 4$.

$\hat{R}$ is 1.0001 for θ_2.

(a)$s = 2$



(b)$s = 4$

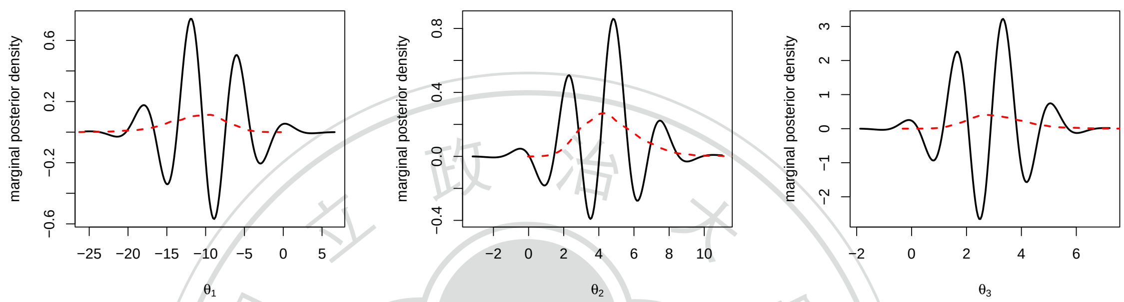


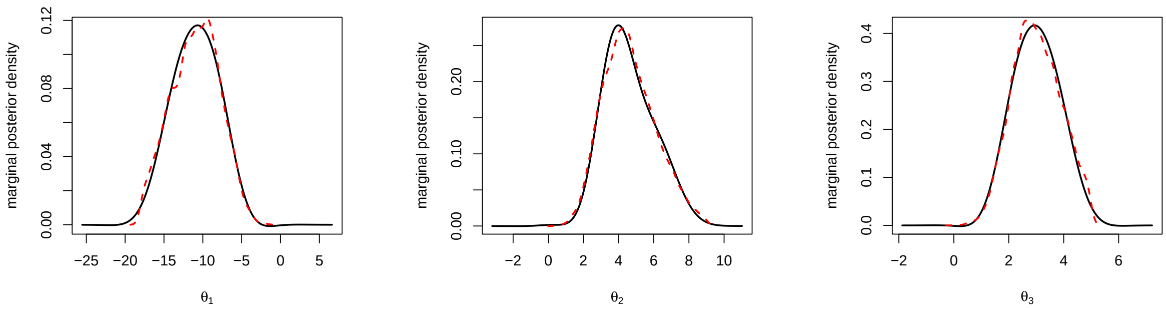
Figure 11: Marginal posterior pdf of θ_1, θ_2 and θ_3. Logit model with flat prior.

Dashed: MCMC(burn-in=5000, mcmc=10000);

Solid: Equation (29) with $s = 2$ and $s = 4$.

$\hat{R}$ is 1.0022 for θ_3.

(a)$s = 2$



(b)$s = 4$

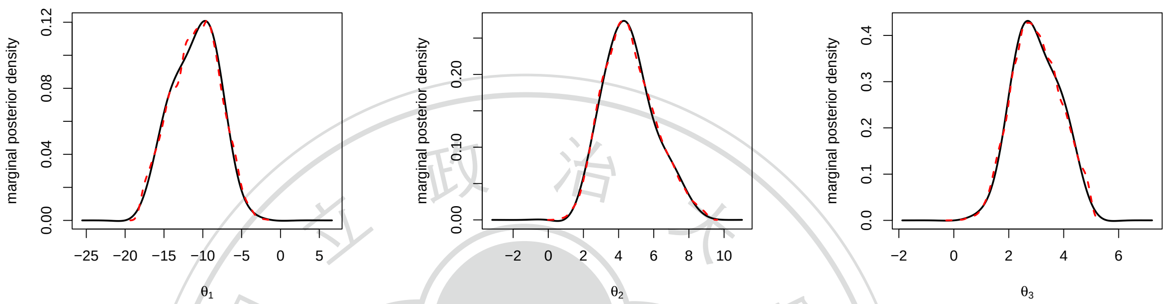


Figure 12: Marginal posterior pdf of θ_1, θ_2 and θ_3 (truncated). Logit model with flat prior.

Dashed: MCMC(burn-in=5000, mcmc=10000);

Solid: Equation (29) with $s = 2$ and $s = 4$.

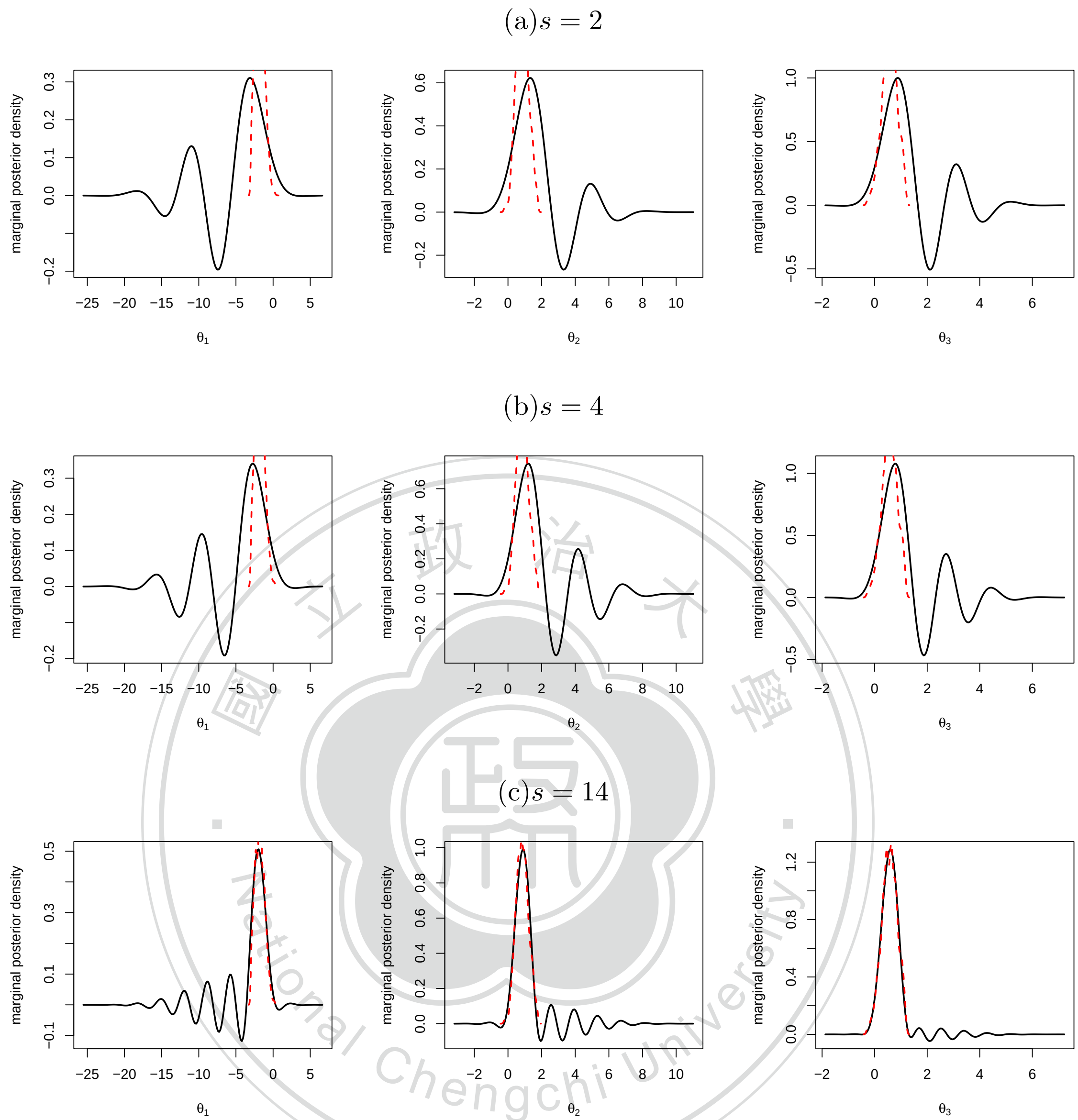


Figure 13: Marginal posterior pdf of θ_1, θ_2 and θ_3 (truncated). Logit model with N(0,1) prior.

Dashed: MCMC(burn-in=5000, mcmc=10000);

Solid: Equation (29) with $s = 2$, $s = 4$ and $s = 14$.

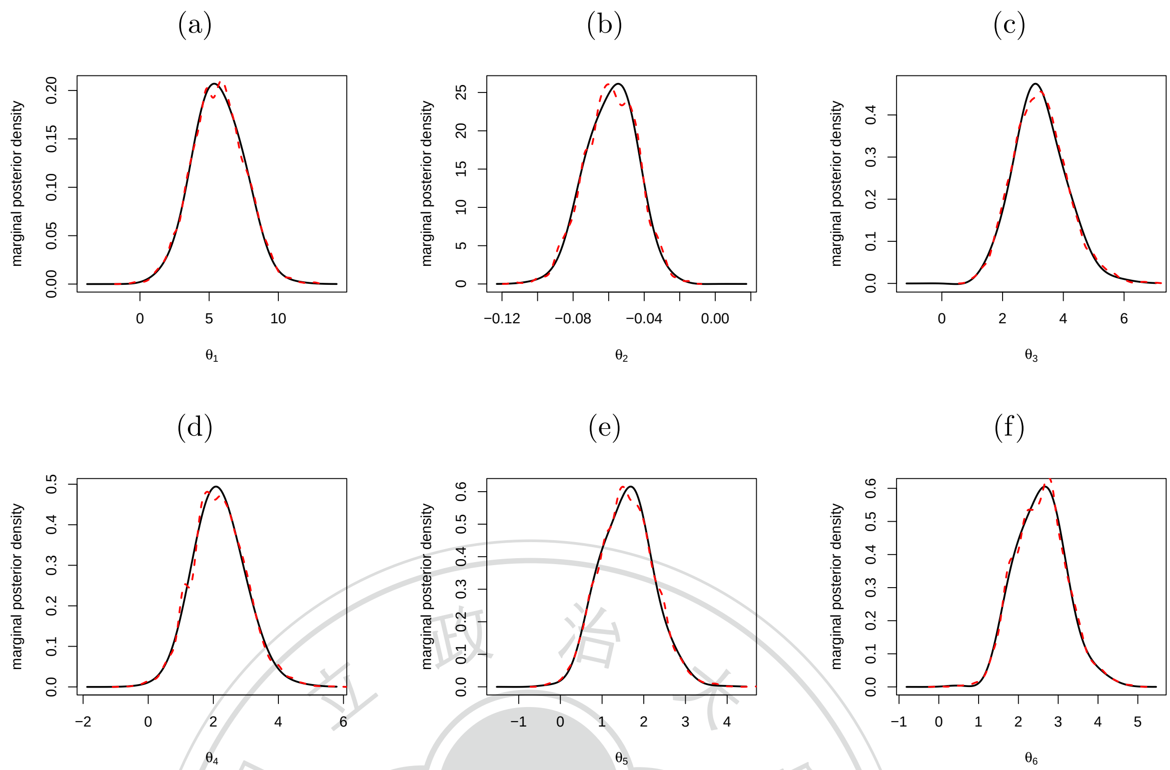


Figure 14: Marginal posterior pdf of $\theta_1 \ldots \theta_6$. Logit model with flat prior.

Dashed: MCMC(burn-in=10000, mcmc=10000);

Solid: Equation (29) with $s = 4$.

$\hat{R}$ is 1.0114 for θ_6.

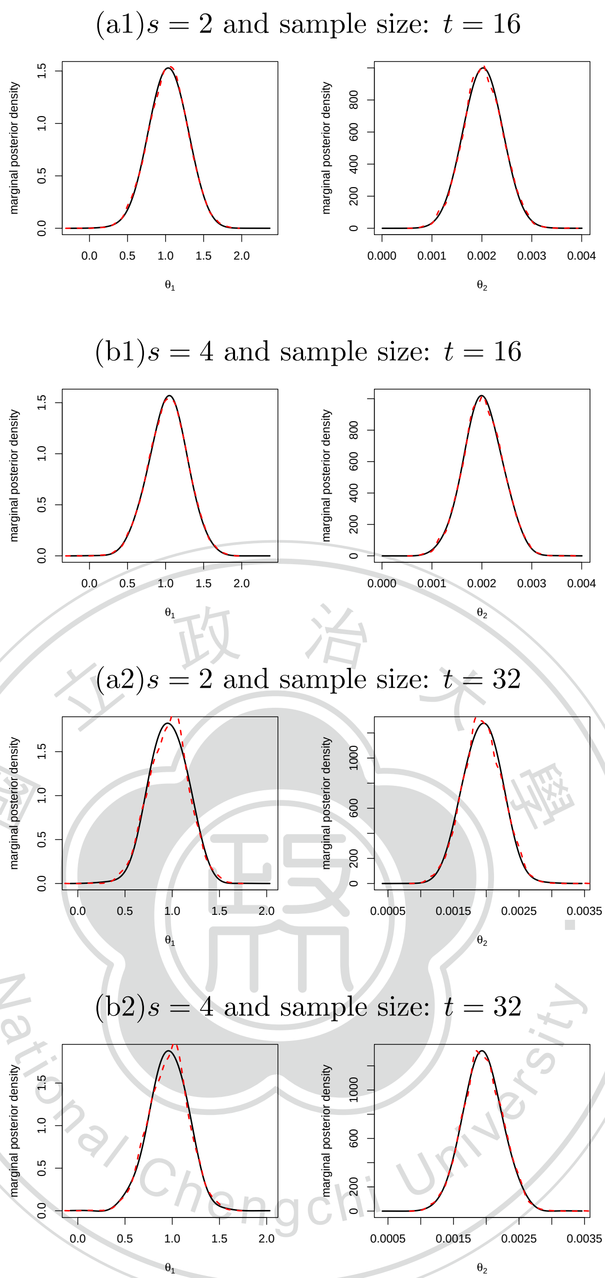


Figure 15: Marginal posterior pdf of θ_1 and θ_2. Poisson model with flat prior.

Dashed: MCMC(burn-in=5000, mcmc=10000).

Solid: Equation (29) with $s = 2$ and $s = 4$.

$\hat{R}$ is 1.0004 for θ_2 with t=32.

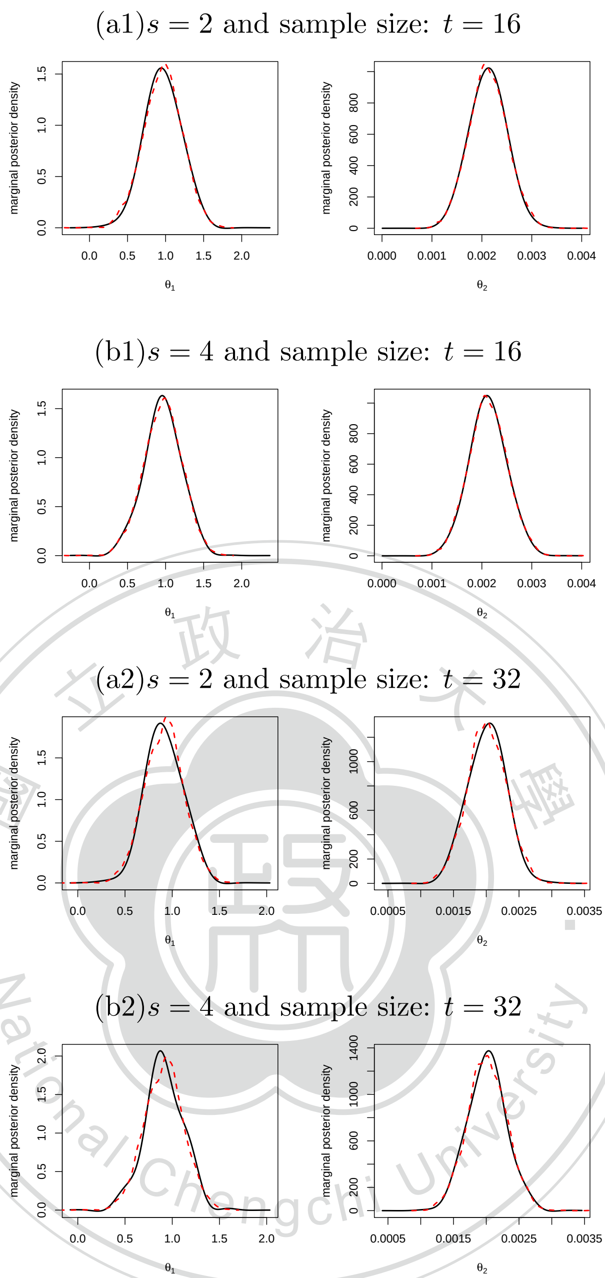


Figure 16: Marginal posterior pdf of θ_1 and θ_2. Poisson model with N(0,1) prior.

Dashed: MCMC(burn-in=5000, mcmc=10000);

Solid: Equation (29) with $s = 2$ and $s = 4$.

$\hat{R}$ is 1.0011 for θ_2 with t=32.

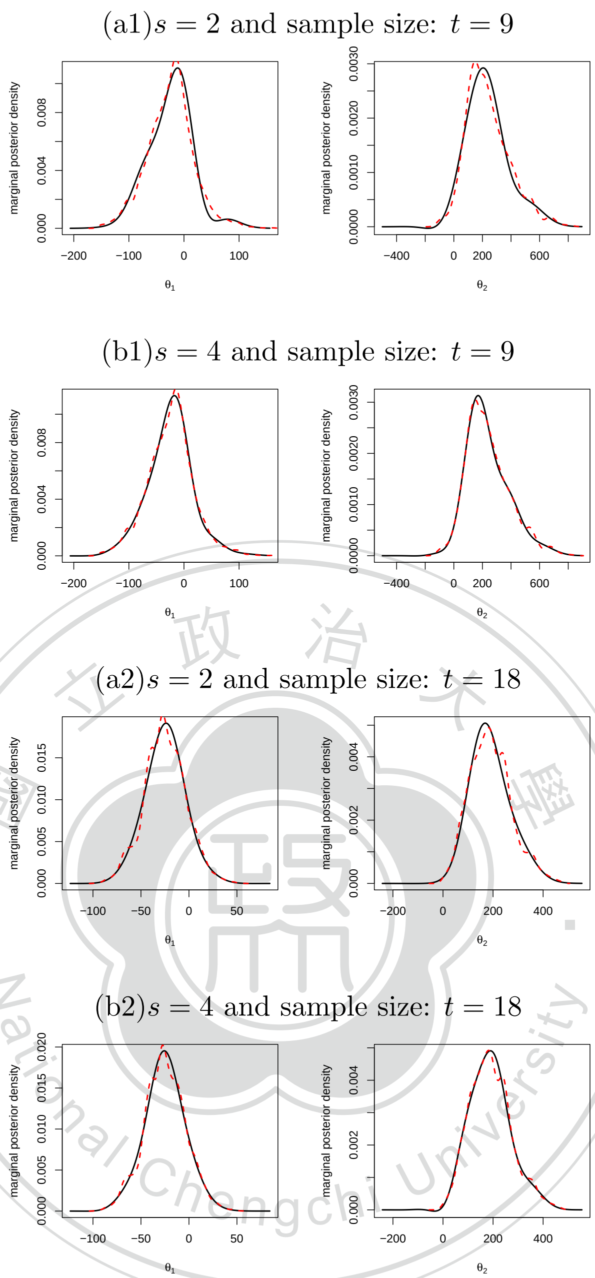


Figure 17: Marginal posterior pdf of θ_1 and θ_2. Gamma model(identity link) with flat prior.

Dashed: MCMC(burn-in=20000, mcmc=10000);

Solid: Equation (29) with $s = 2$ and $s = 4$.

$\hat{R}$ is 1.0111 for θ_2 with t=18.

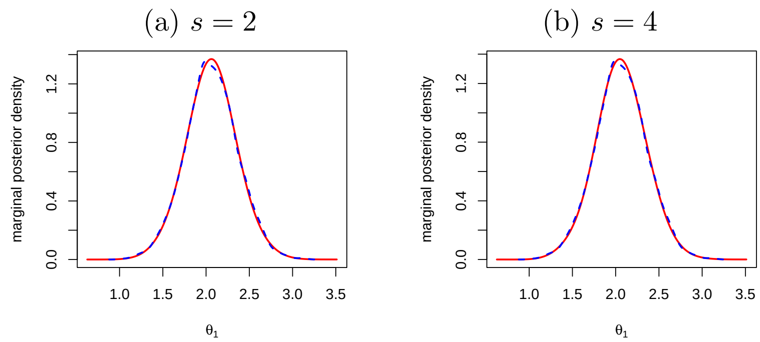


Figure 18: Marginal posterior pdf of θ_1 with $\hat{R} = 1.0024$. Mixture normal model.
Dashed: MCMC(burn-in=5000, mcmc=5000);
Solid: Equation (29) with $s = 2$ and $s = 4$.

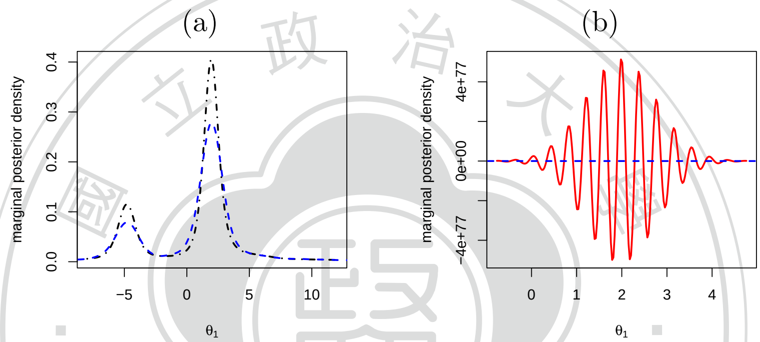


Figure 19: Marginal posterior pdf of θ_1 with $\hat{R} = 1.0016$. Mixture normal model.
Dashed: MCMC(burn-in=5000, mcmc=5000); Solid: Equation (29).
(a)Dot-dashed: Exact distribution by numerical integration; (b)$s = 27$.

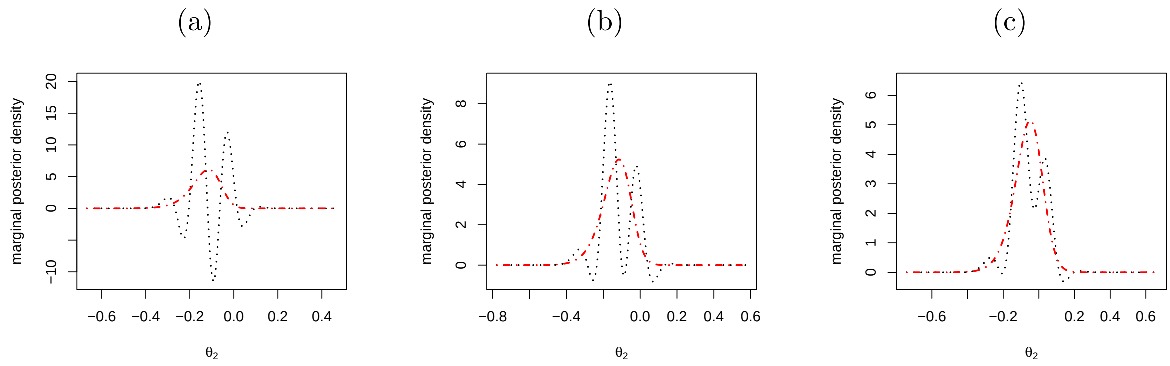


Figure 20: Marginal posterior pdf of θ_2 with $t = 12$. Logit2p-flat model. Dotted: Equation (23); Dot-dashed: Exact distribution by numerical integration. (a), (b) and (c) with $s = 2$.

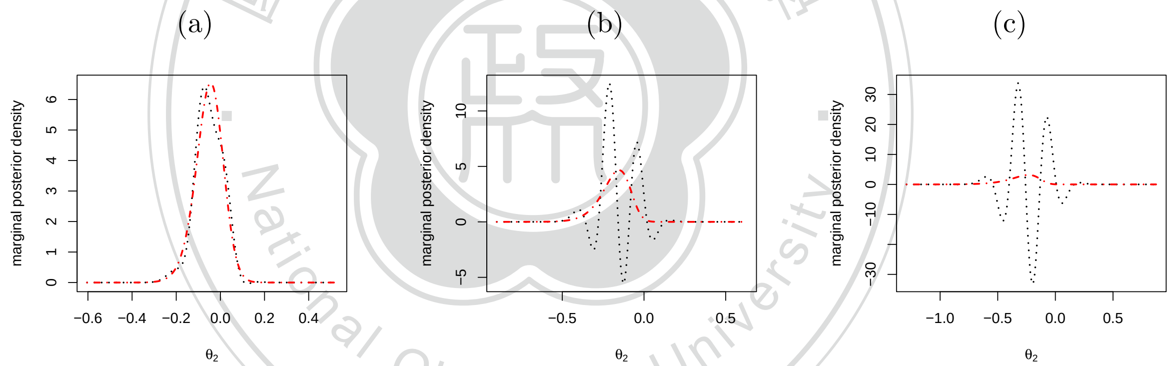


Figure 21: Marginal posterior pdf of θ_2 with $t = 15$. Logit2p-flat model. Dotted: Equation (23); Dot-dashed: Exact distribution by numerical integration. (a), (b) and (c) with $s = 2$.

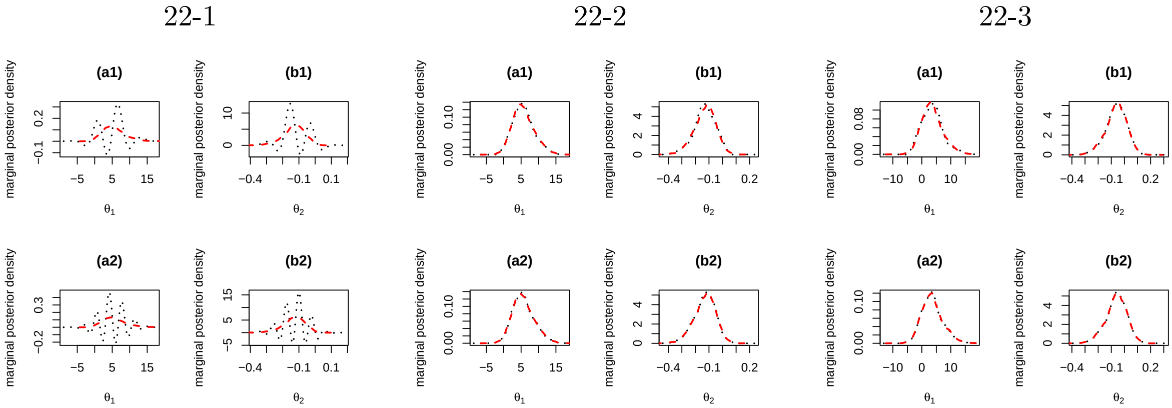


Figure 22: Marginal posterior pdf of θ_1 and θ_2 with $t = 12$. Logit2p-flat model. Dashed: MCMC(burn-in=5000, mcmc=10000); Dotted: Equation (23). (a1) and (b1) with $s = 2$; (a2) and (b2) with $s = 7$.

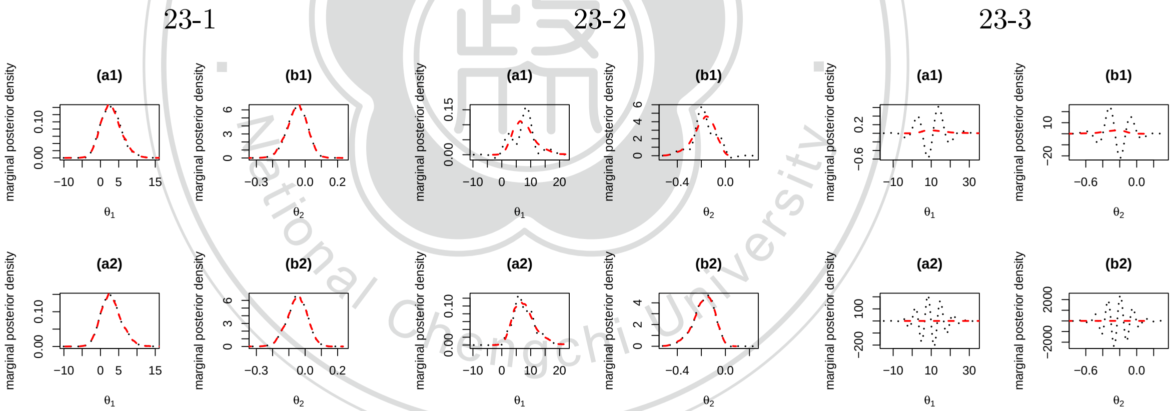


Figure 23: Marginal posterior pdf of θ_1 and θ_2 with $t = 15$. Logit2p-flat model. Dashed: MCMC(burn-in=5000, mcmc=10000); Dotted: Equation (23). (a1) and (b1) with $s = 2$; (a2) and (b2) with $s = 7$.

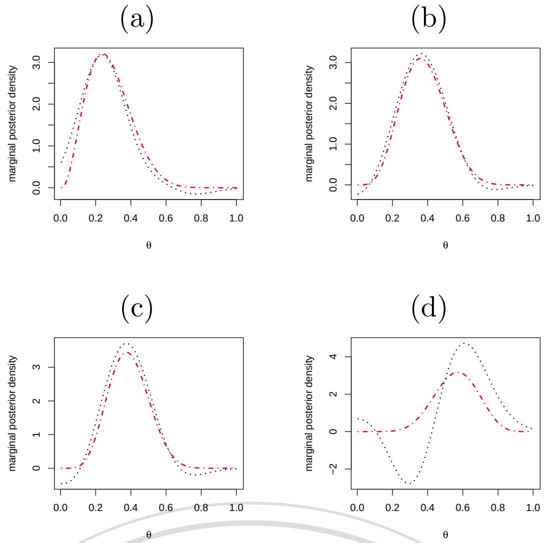


Figure 24: Marginal posterior pdf of θ. Beta-Binomial model $(t_1, y_1) = (8, 3)$.
Dotted: Equation (37); Dot-dashed: Exact distribution.
(a) prior Beta$(0.5, 4)$. (b) prior Beta$(2.5, 4)$. (c) prior Beta$(4, 6)$. (d) prior Beta$(6, 2)$.

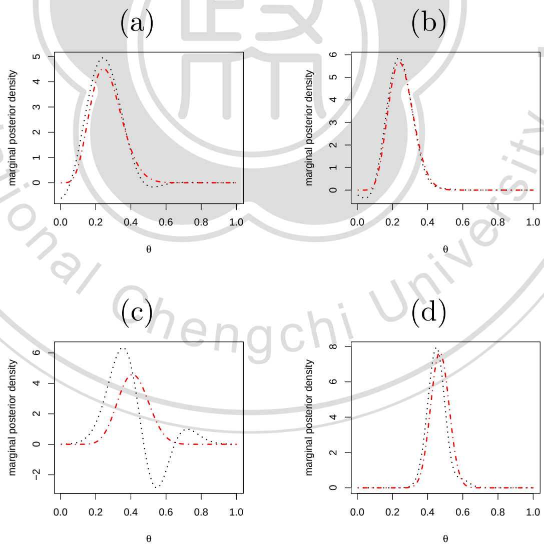


Figure 25: Marginal posterior pdf of θ. Beta-Binomial prior Beta$(0.5, 4)$. $(t_1, y_1) = (8, 3)$.
Dotted: Equation (37); Dot-dashed: Exact distribution.
(a)$(t_2, y_2) = (12, 3)$. (b)$(t_2, y_2) = (25, 6)$. (c)$(t_2, y_2) = (20, 10)$. (d)$(t_2, y_2) = (80, 40)$.

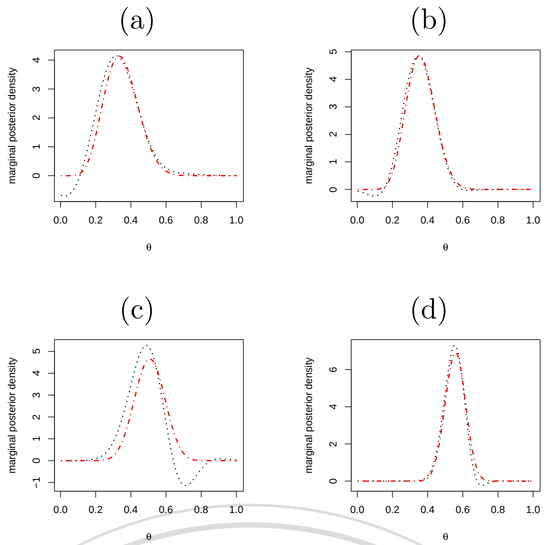


Figure 26: Marginal posterior pdf of θ. Beta-Binomial prior Beta(2.5, 4). $(t_1, y_1) = (8, 3)$

Dotted: Equation (37); Dot-dashed: Exact distribution.

(a)$(t_2, y_2) = (10, 3)$. (b)$(t_2, y_2) = (20, 7)$. (c)$(t_2, y_2) = (20, 12)$. (d)$(t_2, y_2) = (60, 36)$

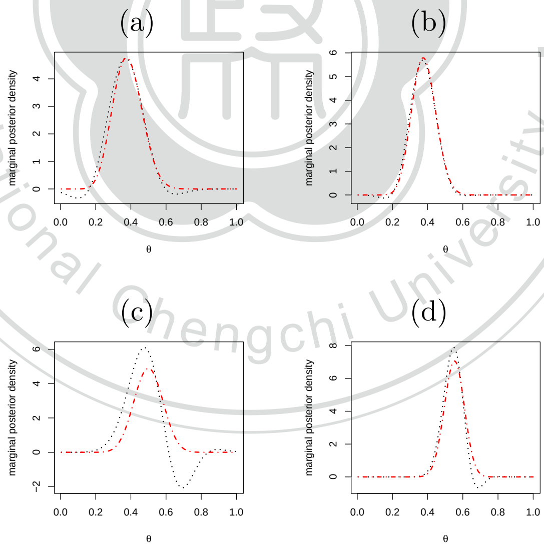


Figure 27: Marginal posterior pdf of θ. Beta-Binomial prior Beta(4, 6). $(t_1, y_1) = (8, 3)$

Dotted: Equation (37); Dot-dashed: Exact distribution.

(a)$(t_2, y_2) = (16, 6)$. (b)$(t_2, y_2) = (32, 12)$. (c)$(t_2, y_2) = (20, 12)$. (d)$(t_2, y_2) = (60, 36)$

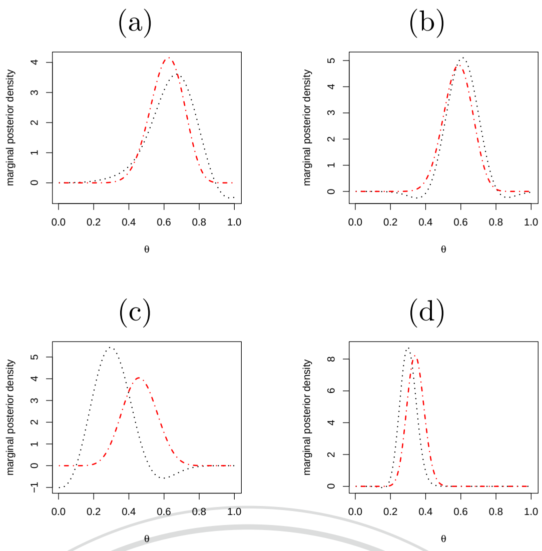
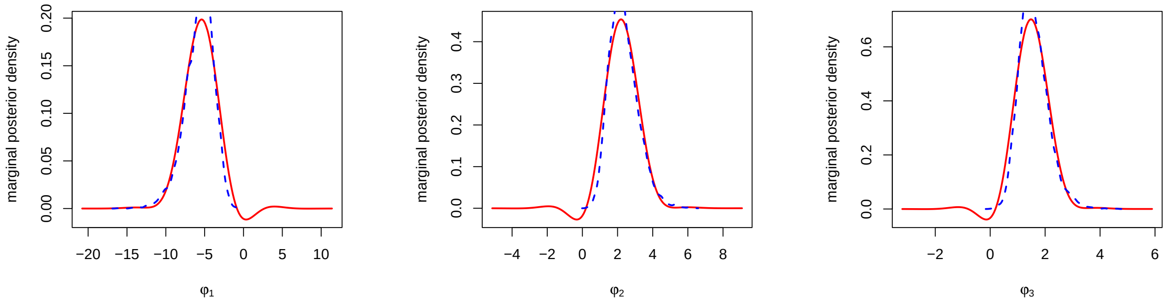


Figure 28: Marginal posterior pdf of θ. Beta-Binomial prior Beta$(6,2)$. $(t_1, y_1) = (8, 3)$

Dotted: Equation (37); Dot-dashed: Exact distribution.

(a)$(t_2, y_2) = (10, 7)$. (b)$(t_2, y_2) = (20, 12)$. (c)$(t_2, y_2) = (10, 2)$. (d)$(t_2, y_2) = (80, 16)$

(a)$s = 2$



(b)$s = 4$

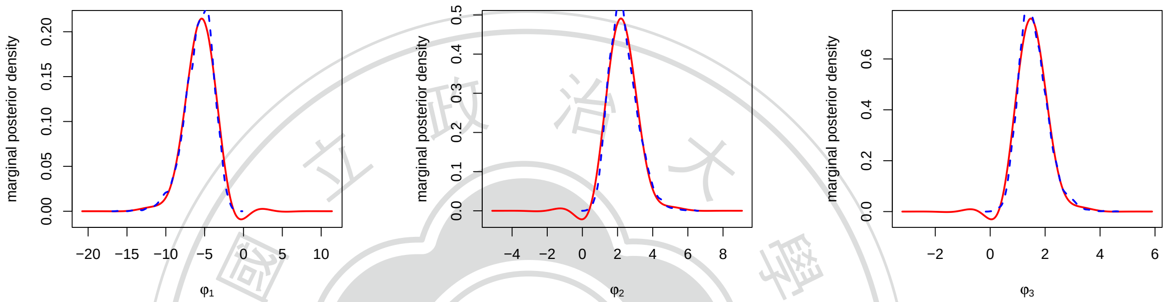


Figure 29: Marginal posterior pdf of φ_1, φ_2 and φ_3 ($W_t = 0.5 * Z_t$). Logit model with flat prior.

Dashed: MCMC(burn-in=5000, mcmc=10000);

Solid: Equation (47) with $s = 2$ and $s = 4$.

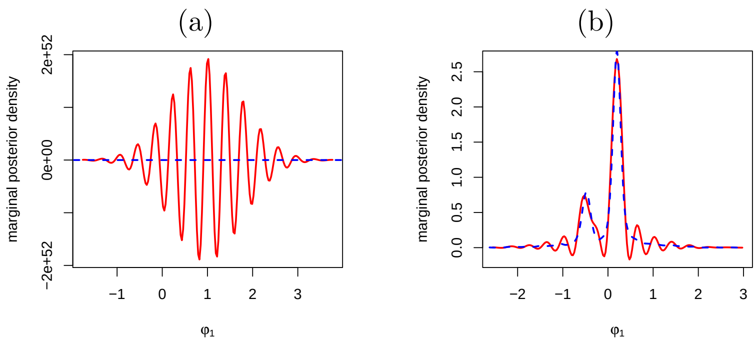


Figure 30: Marginal posterior pdf of φ_1. Mixture normal model.

Dashed: MCMC(burn-in=5000, mcmc=5000); Solid: Equation (47).

(a)$W_t = 0.5 * Z_t$ with $s = 27$; (b)$W_t = 0.1 * Z_t$ with $s = 27$.

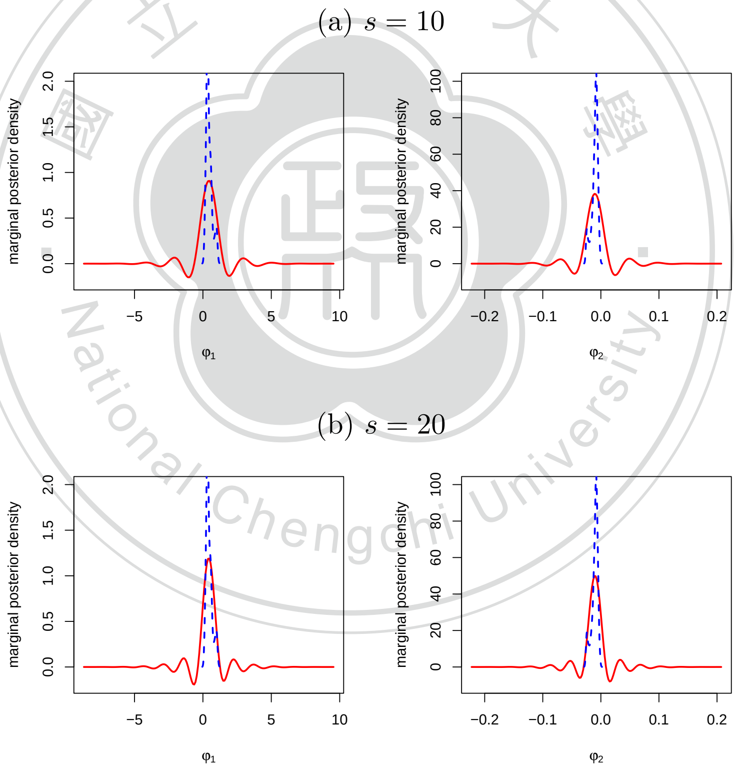
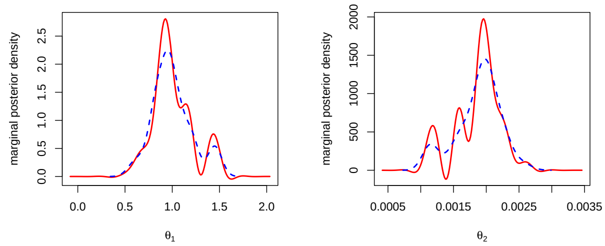


Figure 31: Marginal posterior pdf of φ_1 and φ_2 ($W_t = 0.1 * Z_t$). Logit model with flat prior.

Dashed: MCMC(burn-in=50, mcmc=100); Solid: Equation (47). (a) $s = 10$; (b) $s = 20$.

(a) $s = 10$



(b) $s = 20$

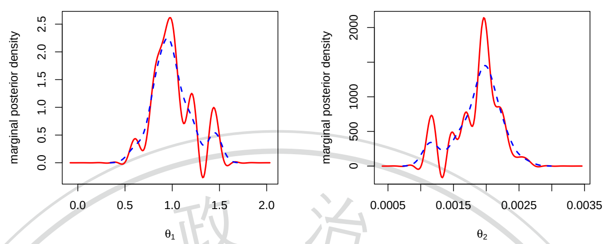
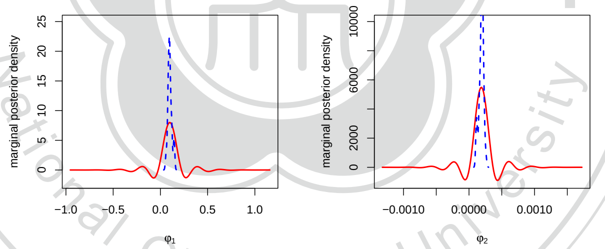


Figure 32: Marginal posterior pdf of θ_1 and θ_2. Poisson-flat model. Dashed: MCMC(burn-in=50, mcmc=100); Solid: Equation (23). (a) $s = 10$; (b) $s = 20$. $\hat{R}$ is 1.5873 for θ_2.

(a) $s = 10$



(b) $s = 20$

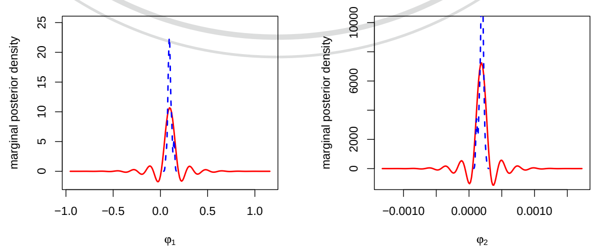


Figure 33: Marginal posterior pdf of φ_1 and φ_2 ($W_t = 0.1 * Z_t$). Poisson-flat model. Dashed: MCMC(burn-in=50, mcmc=100); Solid: Equation (47). (a) $s = 10$; (b) $s = 20$.

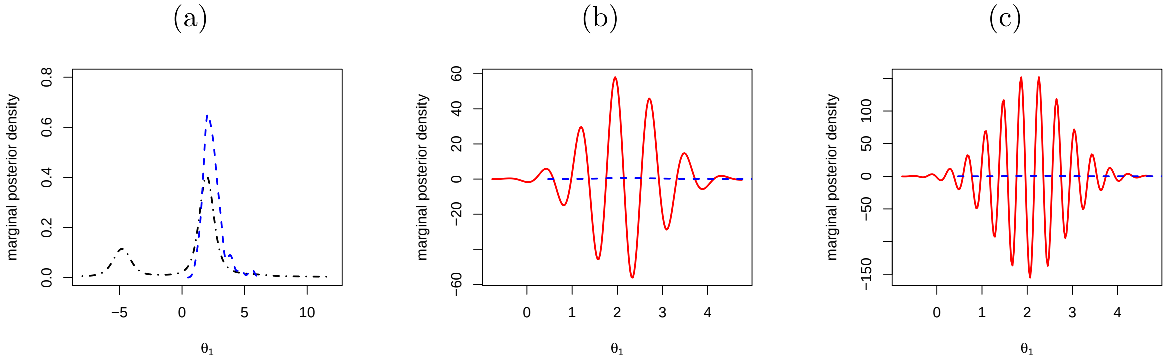


Figure 34: Marginal posterior pdf of θ_1 with $\hat{R} = 1.1375$. Mixture normal model. Dashed: MCMC(burn-in=50, mcmc=50); Solid: Equation (23). (a)Dot-dashed: Exact distribution by numerical integration; (b)$s = 7$; (c)$s = 27$.

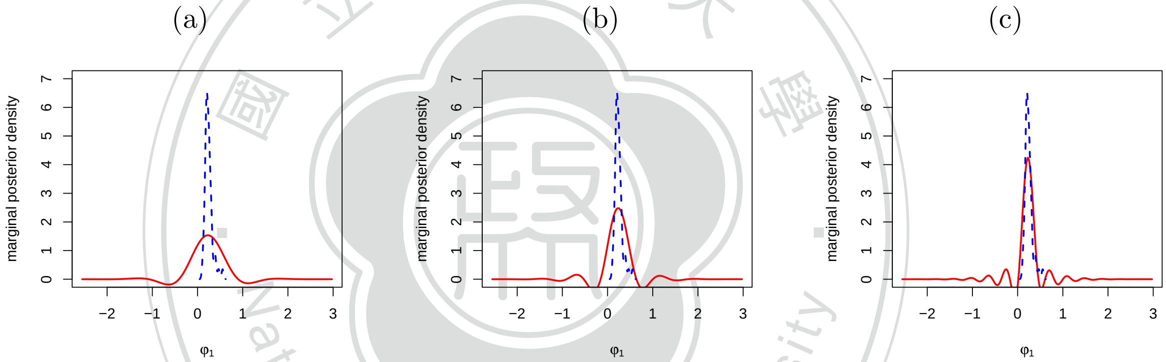


Figure 35: Marginal posterior pdf of φ_1 ($W_t = 0.1 * Z_t$). Mixture normal model. Dashed: MCMC(burnin=50, mcmc=50); Solid: Equation (47). (a) $s = 2$; (b) $s = 7$; (c) $s = 27$.