应用概率统计　第 33 卷
第 1 期　2017 年 2 月

Chinese Journal of Applied Probability and Statistics
Feb., 2017, Vol. 33, No. 1, pp. 21-31
doi: 10.3969/j.issn.1001-4268.2017.01.002

Improper and Proper Posteriors with Improper Hierarchical Priors in a Multivariate Linear Model *

HE Lei[1,2]　　HE DaoJiang[1*]

([1] *Department of Statistics, Anhui Normal University, Wuhu, 241003, China*)

([2] *College of Mathematics and Science, Shanghai Normal University, Shanghai, 200234, China*)

Abstract: In Bayesian analysis, the Markov Chain Monte Carlo (MCMC) algorithm is an efficient and simple method to compute posteriors. However, the chain may appear to converge while the posterior is improper, which will leads to incorrect statistical inferences. In this paper, we focus on the necessary and sufficient conditions for which improper hierarchical priors can yield proper posteriors in a multivariate linear model. In addition, we carry out a simulation study to illustrate the theoretical results, in which the Gibbs sampling and Metropolis-Hasting sampling are employed to generate the posteriors.

Keywords: hierarchical priors; proper posteriors; Gibbs sampling; Metropolis-Hasting sampling

2010 Mathematics Subject Classification: 62F15; 62J05

Citation: He L, He D J. Improper and proper posteriors with improper hierarchical priors in a multivariate linear model [J]. Chinese J. Appl. Probab. Statist., 2017, 33(1): 21–31.

§1. Introduction

In Bayesian analysis, Markov Chain Monte Carlo (MCMC) sampling is widely used to compute the posteriors. However, sometimes it is difficult to diagnose from the results of MCMC algorithm whether the posterior is proper (see [1] and [2]). Therefore, the theoretical investigation is particularly important for us in practical applications.

The resulting posterior should be proper, which is one of the conditions for Markov Chain Monte Carlo algorithm to converge (see [3]). Then an important issue arises, that is, whether the posterior is proper or not when an improper prior is specified. In fact, although MCMC algorithm is an efficient and simple Bayesian computing method, the chain may appear to converge while the posterior is improper. Therefore, we need to analyze convergence conditions theoretically.

*The research was supported by the National Natural Science Foundation of China (Grant No. 11201005) and the Key Project from Anhui Provincial Education Department (Grant No. gxfxZD2016015).

⋆Corresponding author, E-mail: djheahnu@163.com.

Received January 9, 2015. Revised September 15, 2015.

There is some work along this topic, for example, Hobert and Casella [1] pointed out "it may be possible to use null Gibbs chains to make inferences about lower-dimensional functions of the parameters that have proper posteriors"; Gelfand and Sahu [4] discussed a Gibbs sampling issue in an embedded generalized linear models, and obtained some meaningful results, especially for the lower-dimensional posterior. Hadjicostas and Berry [2] also indicated how crucial it is to know whether the posterior is proper.

In this paper we focus on the effect of improper priors on Gibbs sampling in a multivariate linear model. The model is expressed as

$$Y = XB + \varepsilon, \qquad \varepsilon \sim \mathrm{N}_{n\times m}(0, I_n \otimes \Sigma), \tag{1}$$

where Y is an observable $n\times m$ random matrix, X is an $n\times k$ design matrix with $\mathrm{rank}(X) = k \leqslant n$, and B is a $k \times m$ matrix of unknown parameters, ε is an $n \times m$ matrix of random errors which is assumed to follow a matrix normal distribution. $I_n \otimes \Sigma$ denotes the Kronecker product of I_n and Σ, here I_n is the identity matrix of order n, Σ is an $m \times m$ unknown positive definite matrix.

The paper is structured as follows. In Section 2, we present the improper hierarchical priors specification along with discussions about some assumptions. In Section 3, we give the main theoretical results, i.e., the conditions of improper priors yield proper posteriors in the multivariate linear model (1). In Section 4, we address the effects of estimation with respect to the proper and improper posteriors by a simulation. In Section 5, we give an example based on the data of the grain development (see [5]), from which we will show the importance of proper posteriors. Finally, some concluding remarks are given at the end of this paper.

§2. The Specification of Priors

In Bayesian statistical analysis, one of the important issues is to specify priors. Here we first determine the priors for the regression coefficient matrix B and the covariance matrix Σ. The popular noninformative prior, the so-called diffuse prior, which is a constant prior for B, see [6] and [7]. Meanwhile, we assign a hierarchical priors for the covariance matrix Σ that refer to Hobert and Casella [1], Daniels and Kass [8] and Bouriga and Féron [9]. Then the hierarchical priors for B and Σ are specified as follows:

$$\begin{aligned}
\pi(B) &\propto 1;\\
\Sigma \mid \Psi, \beta &\sim \mathrm{IW}_m(\Psi, \beta), \qquad \Psi = \mathrm{diag}(\sigma_1^2, \sigma_2^2, \ldots, \sigma_m^2);\\
\pi(\sigma_i^2) &\propto (\sigma_i^2)^{-a_i} I_{(0,+\infty)}(\sigma_i^2), \qquad i = 1, 2, \ldots, m;
\end{aligned}$$

$$\pi(\beta) \propto \frac{1}{\beta^\delta} I_{(m,+\infty)}(\beta), \tag{2}$$

where a_i's and δ are known, and $\mathrm{IW}_m(\Psi,\beta)$, in the notation of Anderson [10], stands for the inverted Wishart distribution with β degrees of freedom and precision matrix Ψ. The density of $\mathrm{IW}_m(\Psi,\beta)$ is

$$\pi(\Sigma \mid \Psi,\beta) = \frac{1}{2^{m\beta/2}\Gamma_m(\beta/2)} |\Psi|^{\beta/2} |\Sigma|^{-(m+\beta+1)/2} \operatorname{etr}\Big\{-\frac{1}{2}\Sigma^{-1}\Psi\Big\}, \tag{3}$$

where $\operatorname{etr}(A) = \exp(\operatorname{tr}(A))$ for a matrix A, and $\Gamma_m(\cdot)$ is the multivariate gamma function defined as

$$\Gamma_m(\beta) = \pi^{m(m-1)/4} \prod_{i=1}^{m} \Gamma\Big(\beta + \frac{1-i}{2}\Big). \tag{4}$$

For the above hierarchical priors, we need to propound two conditional independence assumptions: (i) $\sigma_1^2, \sigma_2^2, \ldots, \sigma_m^2$ and β are a priori independent; (ii) given Ψ and β, Y is conditionally independent of B and Σ.

§3. Theoretical Results

In this section, we will discuss propriety conditions when improper hierarchical priors can yield the proper joint posteriors. First, we need to derive the joint posterior density of (B,Σ,Ψ,β) given Y by the Bayesian principle.

For model (1), the likelihood of (B,Σ) based on Y is given by

$$L(B,\Sigma \mid Y) = 2\pi^{-mn/2} |\Sigma^{-1}|^{n/2} \operatorname{etr}\Big\{-\frac{1}{2}\big[(B-\widehat{B})^\intercal X^\intercal X (B-\widehat{B}) + S\big]\Big\}, \tag{5}$$

where $\widehat{B}$ and $\widehat{\Sigma}$ are the least squares estimator (LSE) of B and Σ, i.e.,

$$\widehat{B} = (X^\intercal X)^{-1} X^\intercal Y, \qquad \widehat{\Sigma} = \frac{S}{n-k}, \tag{6}$$

where

$$S = (Y - X\widehat{B})^\intercal (Y - X\widehat{B}). \tag{7}$$

It follows (2) and (5) that the joint posterior density of (B,Σ,Ψ,β) is

$$\begin{aligned} \pi(B,\Sigma,\Psi,\beta \mid Y) &\propto L(B,\Sigma \mid Y)\pi(B)\pi(\Sigma \mid \Psi,\beta)\pi(\beta) \prod_{i=1}^{m} \pi(\sigma_i^2) \\ &\propto |\Sigma^{-1}|^{(n+m+\beta+1)/2} \operatorname{etr}\Big\{-\frac{1}{2}\Sigma^{-1}\big[(B-\widehat{B})^\intercal X^\intercal X (B-\widehat{B}) + S + \Psi\big]\Big\} \\ &\quad \cdot \frac{|\Psi|^{\beta/2}}{2^{m\beta/2}\Gamma_m(\beta/2)\beta^\delta} \prod_{i=1}^{m} \pi(\sigma_i^2). \end{aligned} \tag{8}$$

Now we give the sufficient and necessary conditions for the posterior distribution to be proper.

Theorem 1 For the multivariate linear model (1), under the improper hierarchical priors (2), the posterior distribution is proper if and only if

(1) $1-(n-k)/2 < a_i < 1+\beta/2$, $i=1,2,\ldots,m$;

(2) $\delta + \sum\limits_{i=1}^{m} a_i > m+1$.

In order to prove Theorem 1, we first present some lemmas needed.

Lemma 2 Assume that $\mathrm{Re}(\alpha) > (m-1)/2$, let Σ be a positive definite of order m, then

$$\int_{A>0} |A|^{\alpha-(m+1)/2}\,\mathrm{etr}\Big\{-\frac{1}{2}\Sigma^{-1}A\Big\}\mathrm{d}A = \Gamma_m(\alpha)|\Sigma|^{\alpha}2^{m\alpha}.$$

Proof This Lemma is quoted from Liu [11]. □

Lemma 3 Suppose that D and H are two $n\times n$ nonnegative definite matrices, then we have

$$\prod_{i=1}^{n}[\lambda_i(D)+\lambda_i(H)] \leqslant |D+H| \leqslant \prod_{i=1}^{n}[\lambda_i(D)+\lambda_{n-i+1}(H)],$$

where $\lambda_i(A)$ denotes the i-th largest eigenvalue of a matrix A.

Proof See Marshall and Olkin [12]. □

Lemma 4 Let $a,b\in\mathbb{R}$, then

$$\lim_{x\to+\infty}\frac{\Gamma(x+a)}{\Gamma(x+b)x^{a-b}} = 1.$$

Proof The result is due to Bustoz and Ismail [13], see Section 4 therein. □

Now we turn to give the proof of Theorem 1 in details.

Proof of Theorem 1 Integrating (8) over B, we can obtain the joint posterior density of (Σ,Ψ,β), that is

$$\pi(\Sigma,\Psi,\beta\,|\,Y) \propto \frac{|\Sigma^{-1}|^{(n+m+\beta-k+1)/2}\,\mathrm{etr}\{-\Sigma^{-1}(S+\Psi)/2\}|\Psi|^{\beta/2}}{2^{m\beta/2}\Gamma_m(\beta/2)\beta^{\delta}}\prod_{i=1}^{m}\pi(\sigma_i^2). \tag{9}$$

Further, integrating (9) over Σ by Lemma 2, we have

$$\begin{aligned}\pi(\Psi,\beta\,|\,Y) &\propto \int_{\Sigma>0}\pi(\Sigma,\Psi,\beta\,|\,Y)\mathrm{d}\Sigma\\ &\propto \frac{1}{\Gamma_m(\beta/2)\beta^{\delta}}\Gamma_m\Big(\frac{n+\beta-k}{2}\Big)|\Psi|^{\beta/2}|S+\Psi|^{-(n+\beta-k)/2}\prod_{i=1}^{m}\pi(\sigma_i^2).\end{aligned} \tag{10}$$

In order to study the integration of $\pi(\Psi,\beta\,|\,Y)$ on $(0,+\infty)^m$, we first make spectral decomposition of the symmetric matrix S defined in (7). Denote $S = H^{\mathsf{T}}\Lambda H$, where

$\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_m)$, $\lambda_1 \leqslant \lambda_2 \leqslant \cdots \leqslant \lambda_m$ are the sorted eigenvalues of S, and H is an orthogonal matrix. Thus, we have

$$\pi(\Psi, \beta \,|\, Y) \propto \frac{\Gamma_m((n+\beta-k)/2)|\Psi|^{\beta/2}}{\Gamma_m(\beta/2)\beta^{\delta}}|\Psi + H^{\mathsf{T}}\Lambda H|^{-(n+\beta-k)/2}\prod_{i=1}^{m}\pi(\sigma_i^2). \tag{11}$$

By using Lemma 3, we get

$$\prod_{i=1}^{m}(\sigma_i^2 + \lambda_1) \leqslant |\Psi + H^{\mathsf{T}}\Lambda H| \leqslant \prod_{i=1}^{m}(\sigma_i^2 + \lambda_n). \tag{12}$$

It follows from the left side of (12) that

$$\begin{aligned}
\pi(\beta \,|\, Y) &= \int_{(0,+\infty)^m} \pi(\Psi, \beta \,|\, Y)\mathrm{d}\sigma_1^2 \cdots \mathrm{d}\sigma_m^2 \\
&\leqslant \int_{(0,+\infty)^m} \frac{\Gamma_m((n+\beta-k)/2)}{\Gamma_m(\beta/2)\beta^{\delta}} \prod_{i=1}^{m} \frac{(\sigma_i^2)^{\beta/2-a_i}}{(\lambda_1+\sigma_i^2)^{(n+\beta-k)/2}}\mathrm{d}\sigma_1^2 \cdots \mathrm{d}\sigma_m^2 \\
&\propto \frac{\Gamma_m((n+\beta-k)/2)}{\Gamma_m(\beta/2)\beta^{\delta}} \prod_{i=1}^{m} \int_0^{+\infty} \frac{(\sigma_i^2)^{\beta/2-a_i}}{(\lambda_1+\sigma_i^2)^{(n+\beta-k)/2}}\mathrm{d}\sigma_i^2 \\
&:= \pi_{\max}(\beta \,|\, Y).
\end{aligned} \tag{13}$$

In the following, we will first show that the integral

$$\int_0^{+\infty} \frac{(\sigma_i^2)^{\beta/2-a_i}}{(\lambda_1+\sigma_i^2)^{(n+\beta-k)/2}}\mathrm{d}\sigma_i^2$$

in (13) is finite if and only if

$$1 - \frac{n-k}{2} < a_i < 1 + \frac{\beta}{2}. \tag{14}$$

In fact, let $p_i = \beta/2 - a_i$, $q = (n+\beta-k)/2$, using the integral transformation $t = \lambda_1/(\lambda_1 + \sigma_i^2)$, then we have

$$\begin{aligned}
\int_0^{+\infty} \frac{(\sigma_i^2)^{\beta/2-a_i}}{(\lambda_1+\sigma_i^2)^{(n+\beta-k)/2}}\mathrm{d}\sigma_i^2 &= \lambda_1^{p_i-q+1}\int_0^1 t^{q-p_i-2}(1-t)^{p_i}\mathrm{d}t \\
&= \lambda_1^{p_i-q+1}\mathrm{Beta}(q-p_i-1, p_i+1),
\end{aligned}$$

which is finite if and only if

$$q - p_i - 1 > 0, \qquad p_i + 1 > 0.$$

That is to say,

$$\frac{n-k}{2} + a_i - 1 > 0, \qquad \frac{\beta}{2} - a_i + 1 > 0, \tag{15}$$

which is equivalent to

$$1-\frac{n-k}{2}<a_i<1+\frac{\beta}{2}.$$

Now, by the relationship between the gamma function and the beta function, we have

$$\int_0^{+\infty}\frac{(\sigma_i^2)^{\beta/2-a_i}}{(\lambda_1+\sigma_i^2)^{(n+\beta-k)/2}}\mathrm{d}\sigma_i^2=\lambda_1^{-a_i-(n-k)/2+1}\frac{\Gamma(\beta/2-a_i+1)\Gamma((n-k)/2+a_i-1)}{\Gamma((n+\beta-k)/2)}. \tag{16}$$

It follows that

$$\pi_{\max}(\beta\,|\,Y)\propto\frac{\Gamma_m((n+\beta-k)/2)}{\Gamma_m(\beta/2)\beta^\delta}\prod_{i=1}^m\frac{\Gamma(\beta/2-a_i+1)}{\Gamma((n+\beta-k)/2)}. \tag{17}$$

By Lemma 4 and the definition of the multivariate gamma function, it can be shown that, as $\beta\to+\infty$,

$$\prod_{i=1}^m\frac{\Gamma(\beta/2-a_i+1)}{\Gamma((n+\beta-k)/2)}\sim\prod_{i=1}^m\beta^{1-a_i-(n-k)/2},$$
$$\frac{\Gamma_m((n+\beta-k)/2)}{\Gamma_m(\beta/2)}\sim\beta^{m(n-k)/2}.$$

Consequently, as $\beta\to+\infty$,

$$\frac{\Gamma_m((n+\beta-k)/2)}{\Gamma_m(\beta/2)\beta^\delta}\prod_{i=1}^m\frac{\Gamma(\beta/2-a_i+1)}{\Gamma((n+\beta-k)/2)}\sim 1/\beta^{\delta-m+\sum\limits_{i=1}^m a_i}. \tag{18}$$

Note that the right side of (18) is integrable on $(m,+\infty)$ if and only if

$$\delta-m+\sum_{i=1}^m a_i>1. \tag{19}$$

Till now, we have shown that $\pi_{\max}(\beta\,|\,Y)$, an upper bound of $\pi(\beta\,|\,Y)$, is integrable if and only if (14) and (19) hold simultaneously.

On the other hand, from the right side of (12), we have

$$\pi(\beta\,|\,Y)\geqslant\frac{\Gamma_m((n+\beta-k)/2)}{\Gamma_m(\beta/2)\beta^\delta}\prod_{i=1}^m\int_0^{+\infty}\frac{(\sigma_i^2)^{\beta/2-a_i}}{(\lambda_n+\sigma_i^2)^{(n+\beta-k)/2}}\mathrm{d}\sigma_i^2:=\pi_{\min}(\beta\,|\,Y). \tag{20}$$

Similarly, it can be shown that $\pi_{\min}(\beta\,|\,Y)$, an lower bound of $\pi(\beta\,|\,Y)$, is integrable if and only if (14) and (19) are both satisfied. Thus, the proof of Theorem 1 is completed. □

Corollary 5 For the multivariate linear model (1), if all a_i's in the hierarchical priors (2) are taken as 1, then $\delta>1$ is the necessary and sufficient condition for the posterior distribution to be proper.

Proof When $a_i = 1$, $i = 1, 2, \ldots, m$, then the condition (1) in Theorem 1 holds automatically, and the condition (2) turns into $\delta > 1$. $\square$

Remark 6 Proposition 2 in Bouriga and Féron [9] can be regarded as a special case of Corollary 5 where $B = 0$.

To discuss Gibbs sampling in the next section, we here provide the complete conditional distributions of all parameters (B, Σ, Ψ, β), which are listed as follows:

$$[B \,|\, Y, \Sigma, \Psi, \beta] \sim \mathrm{N}(\widehat{B}, (X^\top X)^{-1} \otimes \Sigma), \tag{21}$$

$$[\Sigma \,|\, Y, B, \Psi, \beta] \sim \mathrm{IW}(S + \Psi + (B - \widehat{B})^\top X^\top X(B - \widehat{B}), \beta + n), \tag{22}$$

$$[\sigma_i^2 \,|\, Y, B, \Sigma, \beta] \sim \mathrm{Gamma}\Big(\frac{\beta}{2} - a_i + 1, \frac{\sigma^{ii}}{2}\Big), \qquad i = 1, 2, \ldots, m, \tag{23}$$

$$[\beta \,|\, Y, B, \Sigma, \Psi] \propto \exp\Big\{c\beta - \delta \ln \beta - \ln \Gamma_m\Big(\frac{\beta}{2}\Big)\Big\} I_{]m,+\infty[}(\beta), \tag{24}$$

where $\Sigma^{-1} = (\sigma^{ij})$, and $c = (\ln |\Sigma^{-1}| + \ln |\Psi| - m \ln 2)/2$.

§4. Simulation Results

The proper posterior has been ensured after setting the priors, which is especially important in Bayesian analysis. Because the results of estimation are incorrect when the true posterior is improper. Through the simulation study, we will learn about the behaviour of estimation is different in proper posterior or not.

Now we first describe the calculating process of the Metropolis-Hasting-within-Gibbs sampling method before simulation designed. Let $(B^{(0)}, \Sigma^{(0)}, \Psi^{(0)}, \beta^{(0)})$ be the initial iteration values for (B, Σ, Ψ, β). Suppose that the current values are labeled as $(B^{(t-1)}, \Sigma^{(t-1)}, \Psi^{(t-1)}, \beta^{(t-1)})$, then the updated values $(B^{(t)}, \Sigma^{(t)}, \Psi^{(t)}, \beta^{(t)})$ are generated as follows:

Step 1: draw a sample $B^{(t)}$ from $\mathrm{N}(\widehat{B}, (X^\top X)^{-1} \otimes \Sigma^{(t-1)})$;

Step 2: draw a sample $\Sigma^{(t)}$ from $\mathrm{IW}(S + \Psi^{(t-1)} + (B^{(t)} - \widehat{B})^\top X^\top X(B^{(t)} - \widehat{B}), \beta^{(t-1)} + n)$;

Step 3: draw a sample $\sigma_i^{2(t)}$ from $\mathrm{Gamma}(\beta^{(t-1)}/2 - a_i + 1, \sigma^{ii(t)}/2)$ for $i = 1, 2$, where $\sigma^{ii(t)}$ is the (i, i)-th element in the inverse of $\Sigma^{(t)}$, then $\Psi^{(t)} = \mathrm{diag}(\sigma_1^{2(t)}, \sigma_2^{2(t)})$;

Step 4: draw a sample $\beta^{(t)}$ from the complete conditional distribution of $[\beta \,|\, Y, B^{(t)}, \Sigma^{(t)}, \Psi^{(t)}]$.

For the fourth step, it is difficult to draw from $[\beta \,|\, Y, B^{(t)}, \Sigma^{(t)}, \Psi^{(t)}]$ directly by Gibbs sampling, so the Metropolis-Hasting (M-H) algorithm is employed here to generate $\beta^{(t)}$. A candidate value, labeled v, is drawn from a normal distribution $\mathrm{N}(\beta^{(t-1)}, \tau^2)$ truncated on the interval $(m, +\infty)$, where $m = 2$, and τ^2 is a tuning parameter (see [14]). Denote the right side of (24) by $\varphi(\beta \mid B, \Sigma, \Psi)$, then the candidate value is accepted with probability

$$\min\Big\{1, \frac{\varphi(v \mid B^{(t)}, \Sigma^{(t)}, \Psi^{(t)}) \cdot \Phi((v-2)/\tau)}{\varphi(\beta^{(t-1)} \mid B^{(t)}, \Sigma^{(t)}, \Psi^{(t)}) \cdot \Phi((\beta^{(t-1)}-2)/\tau)}\Big\}.$$

If the candidate value is accepted, then $\beta^{(t)} = v$; otherwise $\beta^{(t)} = \beta^{(t-1)}$. Where $\Phi(\cdot)$ is distribution function of standard normal. The value of τ^2 here is selected as 1 so that the acceptance rate can be around 0.5 for the Markov chain (see [15]).

In the simulation study, we establish the following hierarchical model:

$$\begin{aligned} &\Sigma \mid \Psi, \beta \sim \mathrm{IW}_2(\Psi, 4); \\ &Y_{5\times 2} \sim \mathrm{N}(XB, I_5 \otimes \Sigma), \end{aligned} \tag{25}$$

where $\Psi = \mathrm{diag}(2,2)$ and $X = (x_1, x_2, x_3)$, let $x_1 = (1,1,1,1,1)'$, $x_2 = (1,1,1,0,0)'$, and $x_3 = (1,0,1,0,1)'$. To be convenient for comparison, we set $B = (b_1, b_2)$, among $b_1 = (-5,1,7)'$, $b_2 = (-1,9,5)'$. Moreover, two different cases are considered for the parameters in the priors: (A) $a_1 = 1$, $a_2 = 1$, $\delta = 2$; (B) $a_1 = 1/2$, $a_2 = 1/2$, $\delta = 1$. By Theorem 1, case (A) implies that the posterior is proper, while case (B) yields an improper posterior.

The whole parameters estimation of model (25) are carried out with the M-H algorithms described in above with 5,000 iterations, among which 2,000 are used for the burn-in period. The average values of estimation can be obtained by repeating 1,000 times the previous simulation process. Table 1 summarizes the simulation results for the regression coefficient matrix B and β for the two different cases. For case (A), the results show that the estimated values are approximately identical to the truth ones. Therefore, it is also proved that the proper posterior is of greatest importance. However, these estimations appear to be relatively poor for another case, especially in parameter β.

Table 1 The average values of estimation for $B = (b_{ij})$ and β for both different cases, the values in parentheses refer to the standard errors

	b_{11}	b_{21}	b_{31}	b_{12}	b_{22}	b_{32}	β
Case(A)	-5.0186	0.9929	6.9967	-1.0166	9.10715	4.9323	4.1424
	(1.2498)	(1.3694)	(1.5988)	(1.1023)	(1.3219)	(1.3306)	(2.4234)
Case(B)	-5.2921	0.9299	7.2159	-2.0748	10.2802	6.3472	18.3432
	(6.5989)	(5.5628)	(9.1573)	(36.8914)	(43.6491)	(51.6537)	(20.7892)

§5. Real Data Analysis

The objective of this section is to study the convergence of Gibbs sampling. However, as proposed by [2], the chain may also behave nicely and convergent when the MCMC algorithm is used under a scheme in which the posterior is improper. Consequently, the differences between estimation aren't sharp in such situation, which can be misleading in statistical inference. To reveal it, we shall here use the Anderson's data given by Hossain and Naik [5], which can be assembled in Table 2.

Table 2 Anderson's data

No.	1	2	3	4	5	6	7	8
y_1	40	17	9	15	6	12	5	9
y_2	53	19	10	29	13	27	19	30
x_1	24	11	5	12	7	14	11	18

To investigate the relationship between the weight of grain (y_1), the weight of straw (y_2) and the amount of fertilizer (x_1), we first establish the following multivariate regression model:

$$Y_{8\times 2} = X_{8\times 2}B_{2\times 2} + \varepsilon_{8\times 2}, \qquad \varepsilon \sim \mathrm{N}(0, I_8 \otimes \Sigma),$$

where the elements in the first column of X are all 1, and the priors for B and Σ are specified as that in (2).

In this simulation, the parameters in the priors are specified as (A) $a_1 = 1$, $a_2 = 1$, $\delta = 2$ and (B) $a_1 = 1$, $a_2 = 1$, $\delta = 1$, respectively. Obviously, the posterior is improper for Case (B). For Case (A), the left panel of Figure 1 shows 10,000 observations of β from the Markov chain, and the right one gives the corresponding autocorrelation plot. As we know, the autocorrelation usually decreases with increasing of the iteration step-size. The chain in this case appears to have converged, correctly showing no signs of any problems.

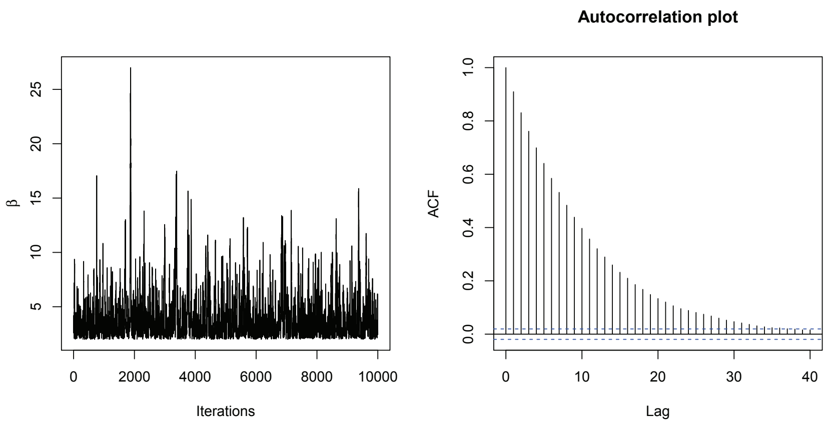


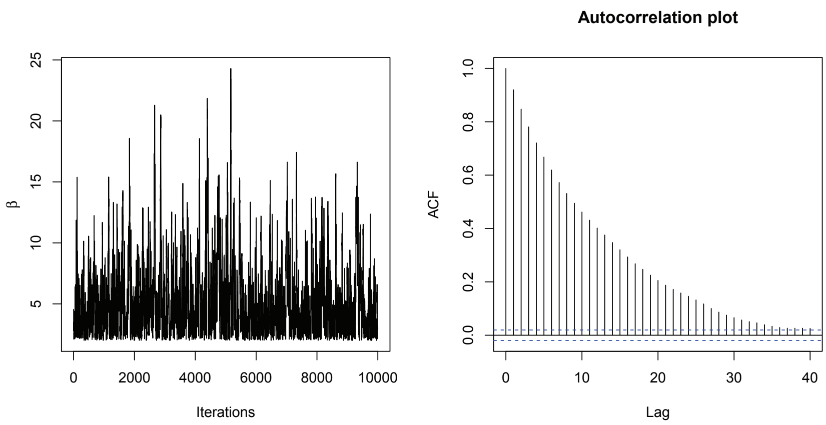
Figure 1 The Markov chain of β from the Gibbs sampling for Case (A) and the corresponding autocorrelation plot

Table 3 reports the posterior means and standard errors in parentheses of regression coefficient matrix B and β, in which 2,000 observations are also used for the burn-in period.

Table 3 The posterior means and standard errors in cases (A) and (B)

	b_{11}	b_{21}	b_{12}	b_{22}	β
Case (A)	-3.6560	1.3952	-2.3161	2.1393	3.8105
	(8.4624)	(0.6055)	(4.2438)	(0.3033)	(5.3348)
Case (B)	-3.8112	1.4060	-2.2740	2.1420	5.3689
	(7.9779)	(0.5699)	(4.1596)	(0.3003)	(5.1481)

For Case (B), Figure 2 shows the 10,000 observations of β from the Gibbs sampling for improper posteriors and the corresponding autocorrelation plot. As seen from the autocorrelation plot, the chain seems reasonable and there is nothing unusual about these observations. However, the estimates are wrong in that the posterior is improper here. The true means and standard deviations do not exist actually. It would be very difficult for us to diagnose from the Markov chain that there is something wrong.



Firgure 2 The Markov chain of β from the Gibbs sampling for Case (B) and the corresponding autocorrelation plot

§6. Concluding Remarks

In this paper, we have derived the necessary and sufficient conditions for the posteriors to be proper in a multivariate linear model with improper hierarchical priors. According to the results of stochastic simulation, we think the estimates are incorrect in that the posterior is improper. Another simulation study based on the Anderson's data is carried out to illustrate the theoretical results. The simulation verifies that it would be hard for us to diagnose from the Markov chain that the posterior is proper or not, which implies that theoretical research of the posteriors is necessary.

References

[1] Hobert J P, Casella G. The effect of improper priors on Gibbs sampling in hierarchical linear mixed models [J]. *J. Amer. Statist. Assoc.*, 1996, **91(436)**: 1461–1473.

[2] Hadjicostas P, Berry S M. Improper and proper posteriors with improper priors in a Poisson-gamma hierarchical model [J]. *Test*, 1999, **8(1)**: 147–166.

[3] Schervish M J, Carlin B P. On the convergence of successive substitution sampling [J]. *J. Comput. Graph. Statist.*, 1992, **1(2)**: 111–127.

[4] Gelfand A E, Sahu S K. Identifiability, improper priors, and Gibbs sampling for generalized linear models [J]. *J. Amer. Statist. Assoc.*, 1999, **94(445)**: 247–253.

[5] Hossain A, Naik D N. Detection of influential observations in multivariate regression [J]. *J. Appl. Statist.*, 1989, **16(1)**: 25–37.

[6] Geisser S. Bayesian estimation in multivariate analysis [J]. *Ann. Math. Statist.*, 1965, **36(1)**: 150–159.

[7] Tiao G C, Zellner A. On the Bayesian estimation of multivariate regression [J]. *J. Roy. Statist. Soc. Ser. B*, 1964, **26(2)**: 277–285.

[8] Daniels M J, Kass R E. Nonconjugate Bayesian estimation of covariance matrices and its use in hierarchical models [J]. *J. Amer. Statist. Assoc.*, 1999, **94(448)**: 1254–1263.

[9] Bouriga M, Féron O. Estimation of covariance matrices based on hierarchical inverse-Wishart priors [J]. *J. Statist. Plann. Inference*, 2013, **143(4)**: 795–808.

[10] Anderson T W. *An Introduction to Multivariate Statistical Analysis* [M]. 3rd ed. Hoboken, New Jersey: Wiley, 2003.

[11] Liu J S. *An Introduction to Wishart Distributions* [M]. Beijing: Science Press, 2005. (in Chinese)

[12] Marshall A W, Olkin I. *Inequalities: Theory of Majorization and Its Applications* [M]. New York: Academic Press, 1979.

[13] Bustoz J, Ismail M E H. On gamma function inequalities [J]. *Math. Comp.*, 1986, **47(176)**: 659–667.

[14] Ntzoufras I. *Bayesian Modeling Using WinBUGS* [M]. Hoboken, New Jersey: Wiley, 2009.

[15] Müller P. A generic approach to posterior integration and Gibbs sampling [R]. West Lafayette: Purdue University, Department of Statistics, 1991.

非正常分层先验下多元线性模型中后验的正常性

贺　磊[1,2]　　何道江[1]

([1]安徽师范大学统计系, 芜湖, 241003; [2]上海师范大学数理学院, 上海, 200234)

摘　要: 在Bayes分析中, MCMC算法是一个简单且行之有效的计算后验的方法. 但是, 有时在非正常后验下得到的Markov链也可能表现出似乎收敛的特征, 这将会导致不正确的统计推断. 为此, 本文给出了在多元线性模型中利用非正常分层先验得到正常后验所需满足的充要条件. 此外, 使用Gibbs方法和 Metropolis-Hasting方法来进行后验抽样, 并通过随机模拟说明了正常后验理论结果的重要性.

关键词: 分层先验; 正常后验; Gibbs抽样; Metropolis-Hasting抽样

中图分类号: O212.8