Statistica Sinica **22** (2012), 1199-1215
doi:http://dx.doi.org/10.5705/ss.2010.263

A SWITCHING MARKOV CHAIN MONTE CARLO METHOD FOR STATISTICAL IDENTIFIABILITY OF NONLINEAR PHARMACOKINETICS MODELS

Seongho Kim and Lang Li

University of Louisville and Indiana University School of Medicine

Abstract: We study the convergence rate of MCMC on the statistically unidentifiable nonlinear model involving the Michaelis-Menten kinetic equation. We have shown that, under certain conditions, the convergence diagnosis of Raftery and Lewis (1992) is consistent with the convergence rate argued by Brooks and Roberts (1999). Therefore, different MCMC schemes developed in linear models are further extended and compared in nonlinear models. We demonstrate that the single component MCMC (SCM) scheme is faster than the group component MCMC (GCM) scheme on unidentifiable models, while GCM is faster than SCM when the model is statistically identifiable. A novel MCMC method is then developed using both SCM and GCM schemes, termed the Switching MCMC (SWM) method. The proposed SWM possesses an advantage in that it is able to estimate parameters regardless of the statistically identifiable situations. In addition, simulations and data analysis suggest a better performance of the proposed SWM algorithm than SCM and GCM.

Key words and phrases: Convergence rate, Michaelis-Menten (MM) kinetics, Monte Carlo Markov chain (MCMC), statistical identifiability, pharmacokinetics (PK), switching algorithm.

1. Introduction

In nonlinear models, mathematical identifiability refers to the identifiability of the model parameters from noise-free data. This is also called structural or deterministic identifiability (Cobelli and DiStefano III (1980)). Statistical identifiability is the identifiability of parameters provided from the noise data (Miao et al. (2009)). This is also termed numerical identifiability (Godfrey and Fitch (1984)).

Historically, neither mathematical nor statistical identifiability is a theoretical concern in Bayesian approach. Lindley (1971) claimed that non-identifiability causes no real difficulties for Bayesian analysis since it is always possible to resolve the issue via proper priors. Many investigators observed that non-identifiability did not preclude Bayesian learning (Poirier (1998); Gelfand and Sahu (1999); Eberly and Carlin (2000)). Neath and Samaniego (1997) characterized a class

of prior distributions that lead to posterior estimates of unidentifiable parameters superior to their prior estimates. Xie and Carlin (2006) carefully quantified Bayesian learning for unidentifiable models and parameters.

Statistically unidentifiable models may lead to poor convergence of MCMC algorithms (Eberly and Carlin (2000)). In linear models with the normal assumption, the theoretical convergence speed of MCMC (i.e. Gibbs sampling) was extensively investigated by Roberts and Sahu (1997). They showed that the convergence speeds of different Gibbs sampling schemes heavily depend on the correlation structures among parameters. Unidentifiable parameters in the linear model often have high correlations, resulting in slow convergence. To overcome the slow convergence of unidentifiable parameters in linear models, the model reparametrization can be an effective approach. For example, Gelfand, Sahu, and Carlin (1995) demonstrated that a hierarchically centered parameterization can lead to faster convergence and mixing.

Similar to linear models, unidentifiable parameters in nonlinear models can slow MCMC algorithms under the Bayesian framework. However, little research has been conducted to investigate the convergence speed of MCMC algorithms for nonlinear models when they have unidentifiable parameters. Unlike linear models, model reparametrization is seriously limited by the functional form of a nonlinear model.

Because of the non-closed form of conditional distributions in nonlinear models, various Metropolis-Hastings algorithms become routine strategies (Gilks, Richardson, and Spiegelhalter (1996)). Consequently, theoretical guidance for how to evaluate the speed of convergence of Gibbs sampling in linear models (Roberts, Gelman, and Gilks (1997)) is not applicable to nonlinear models. Practically, MCMC diagnosis tests are utilized to assess their convergence (Cowles and Carlin (1996)), and different MCMC convergence diagnoses have their own asymptotic theories. Therefore, it is desirable to investigate whether MCMC schemes developed in linear models can be extended to nonlinear models, and whether they keep the same performance. However, without comparable evaluation criteria of MCMC convergence between two classes of models, it is impossible to compare their performances.

In this paper, we speak to three tasks.

- Select a fast MCMC scheme for statistically unidentifiable nonlinear models.
- Investigate whether the order of convergence speed of two MCMC chains determined by Raftery and Lewis (1992) is comparable to the order defined by Robert and Sahu (1997).
- Develop a novel MCMC method that can be used for parameter estimation regardless of the statistical identifiable situation.

We hope our research can provide some insight on the optimal MCMC scheme selection for nonlinear models and assert either the similarity or difference among different MCMC algorithms.

The rest of the paper is organized as follows. Section 2 describes a midazolam (MDZ) pharmacokinetics (PK) model whose metabolism follows a Michaelis-Menten equation. In particular, its unidentifiable PK parameters are introduced. This description of the PK model helps us to define the statistical model and to later evaluate MCMC performances. It is noteworthy that unidentifiable PK parameters usually arrise at the subject specific level. Hence, in this paper, we focus on a nonlinear model for subject's PK data. In Section 3, a single component MCMC and a group component MCMC are defined, and a novel random switching MCMC is proposed for a nonlinear model. In Section 4, the speed of convergence judged by the diagnosis test of Raftery and Lewis (1992) is shown to be consistent with the order defined by Robert and Sahu (1997). In Section 5, the performances of these MCMC algorithms are compared based on MDZ clinical trial data, as well as on simulated data. Conclusions and discussion are in Section 6. Note that all statistical analyses and simulations are performed using the statistical package R (R Development Core Team), and the R code for the proposed algorithms can be found in the supplement at `http://www.stat.sinica.edu.tw/statistica`.

2. Michaelis-Menten Equation and Midazolam Pharmacokinetics Model

2.1. Michaelis-menten equation and statistical identifiability

The drug metabolism rate follows the Michaelis-Menten (MM) kinetics equation (Atkinson et al. (2001)):

$$V(t) = \frac{dC(t)}{dt} = \frac{Vmax \cdot C(t)}{Km + C(t)}, \tag{2.1}$$

where $V(t)$ is the velocity of the reaction, $Vmax$ is the maximum velocity, Km is the MM constant, and $C(t)$ is the drug concentration. Monod (1949) first applied the MM equation to microbiology for the growth rate of microorganisms. The MM equation (2.1) generally describes the relationship between the rate of substrate conversion by an enzyme to the concentration of the substrate. In this relationship, $V(t)$ is the rate of conversion, $Vmax$ is the maximum rate of conversion, and $C(t)$ is the substrate concentration. The MM constant Km is equivalent to the substrate concentration at which the rate of conversion is half of $Vmax$. Km approximates the affinity of enzyme for the substrate. A small Km indicates high affinity, and a substrate with a smaller Km will approach $Vmax$ more quickly. Very high $C(t)$ values are required to approach $Vmax$, which is

reached only when $C(t)$ is high enough to saturate the enzyme. For additional details, we refer the reader to the paper by Hein and Niemann (1962).

In PK studies, statistical identifiability often occurs with the MM equation. Suppose the observed data $y(t)$ follows a normal distribution with the MM equation at a time point t given the parameter $\theta = (Vmax, Km)$:

$$y(t) \sim \mathbf{N}\{logf(\theta, t), \sigma^2\}, \tag{2.2}$$

where $f(\theta, t) = V(t)$. However, when Km is much higher than the concentration $C(t)$,

$$f(\theta, t) = \frac{Vmax \cdot C(t)}{Km + C(t)} \approx \frac{Vmax}{Km} \cdot C(t) \text{ if } C(t) \ll Km. \tag{2.3}$$

On the other hand, when Km is much smaller than the concentration $C(t)$,

$$f(\theta, t) = \frac{Vmax \cdot C(t)}{Km + C(t)} \approx Vmax \text{ if } C(t) \gg Km. \tag{2.4}$$

That is, if the concentration $C(t)$ is either much less than or much greater than the true value of the parameter Km, we cannot estimate both Km and $Vmax$ separately, there are problems with identifiability.

Definition. For a family of distributions $\{p(x|\theta)|\theta \in \Theta\}$. The parameter θ is called *statistically identifiable* if distinct values of θ correspond to distinct probability density or mass functions.

If the model is not statistically identifiable, there is difficulty in doing inference. Since different parameters wcan yield the same likelihood function value (Casella and Berger (1990, Chap. 11)).

2.2. Midazolam pharmacokinetics and its clinical study

Midazolam (MDZ) is a benzodiazepine used to cause relaxation or sleep before surgery and to block the memory of the procedure. It can be administrated in both oral and intravenous formulations (Gorski et al. (1998)).

A MDZ pharmacokinetics (PK) study was conducted in the General Clinical Research Center (GCRC) at Indiana University. Twenty-two subjects were recruited into this study. Blood samples for MDZ assays were collected in non-heparinized evacuated blood collection tubes at 0.5, 0.75, 1, 1.5, 2, 4, 6, and 9 hours after intravenously dosing MDZ (2.98 mg 4.8 mg). We look only one of 24 subjects' clinical trial data.

Compartmental PK analysis uses kinetic models to describe and predict the concentration-time curve. PK compartmental models are often similar to kinetic

models used in such other scientific disciplines as chemical kinetics and thermodynamics. The simplest PK compartmental model is the one-compartmental PK model with oral dose administration and first-order elimination (Chang (2010)). Here the MDZ PK is assumed to follow a two-compartmental model with the MM equation. The PK is described by the system of the ordinary differential equations (ODEs)

$$\begin{aligned}
&\frac{dA_1(t)}{dt} = -CL \cdot \frac{A_1(t)}{V_1} + CL_{12} \cdot \left(\frac{A_2(t)}{V_2} - \frac{A_1(t)}{V_1}\right), \qquad (2.5)\\
&\frac{dA_2(t)}{dt} = -CL_{12} \cdot \left(\frac{A_2(t)}{V_2} - \frac{A_1(t)}{V_1}\right),\\
&CL = \frac{Qh \cdot CLint}{Qh + CLint};\ CLint = \frac{Vmax}{Km + A_1(t)/V_1};\ (A_1, A_2)|_{t=0} = (Dose, 0),
\end{aligned}$$

where (A_1, A_2) are amounts of drug in systemic and peripheral compartments, respectively, (V_1, V_2) are volumes of distribution in systemic and peripheral compartments, respectively, CL_{12} is the inter-compartment rate constant, CL is the systemic clearance, $CLint$ is the intrinsic hepatic clearance, $Vmax$ is the maximum of velocity, Km is MM constant, and Qh is the hepatic blood flow known as 80 l/h.

Since (2.5) is nonlinear, there is no closed-form solution. Therefore we use the R package *odesolve* to numerically solve for PK models. Because only the systemic concentrations are observable in the study, the predicted systemic concentration at time t is

$$\log f(\theta, t) = \frac{\log A_1(t)}{V_1}, \qquad (2.6)$$

where $\theta = (\theta_1, \theta_2, \sigma^2)'$, $\theta_1 = (\log V_1, \log V_2, \log Vmax, \log CL_{12})'$, and $\theta_2 = \log Km$.

3. Nonlinear Models and Monte Carlo Markov Chain Sampling Schemes

3.1. Nonlinear models for MDZ population pharmacokinetics

We illustrate the nonlinear model for MDZ with the MM equation under the framework of Bayesian analysis. The model has the form

$$[\log y_i|\theta, \sigma^2] \sim \mathbf{N}\{\log f(\theta, t_i), \sigma^2\}, i = 1, \ldots, N, \qquad (3.1)$$

where N is the number of time points, $\theta = (\theta_1, \theta_2, \sigma^2)'$, $\theta_1 = (\log V_1,\ \log V_2,\ \log Vmax,\ \log CL_12)'$, $\theta_2 = \log Km$, and $\log f(\theta, t_i) = (\log A_1(t_i))/V_1$ as described in (2.6). It is noteworthy that, since the observed concentration y_i is always greater than zero, the concentration is log-transformed in order for the

PK model to take advantage of a normal distribution. The prior distributions are

$$[\theta] \sim \mathbf{MVN}\{d, D\}, \left[\frac{1}{\sigma^2}\right] \sim Ga\{\frac{v}{2}, \frac{u \cdot v}{2}\}, \tag{3.2}$$

where $d = (d_1, d_2, \ldots, d_5)'$, $D = \begin{pmatrix} D_{11} \cdots D_{15} \\ \vdots \quad \ddots \quad \vdots \\ D_{51} \cdots D_{55} \end{pmatrix}$, **MVN** stands for a multivariate normal distribution, and Ga refers to a gamma distribution. Then the posterior distribution for (θ, σ^2) is proportional to

$$\begin{aligned} &\prod_{i=1}^{N} [\log y_i | \theta, \sigma^2] \times Prior(\theta, \sigma^2), \\ &Prior(\theta, \sigma^2) = \mathbf{MVN}\{d, D\} \times Ga\{\frac{v}{2}, \frac{u \cdot v}{2}\}. \end{aligned} \tag{3.3}$$

The following are conditional distributions and their kernels to draw the parameters (θ, σ^2):

$$[\theta | \log y, \sigma^2] \propto \exp\Big\{-\frac{1}{2\sigma^2}\sum_{i=1}^{N}\big(\log y_i - \log f(\theta, t_i)\big)^2 - \frac{1}{2}(\theta - d)' D^{-1} (\theta - d)\Big\}, \tag{3.4}$$

$$[\frac{1}{\sigma^2} | \log y, \theta] \sim Ga\Big\{\frac{1}{2}(v + N), \frac{1}{2}\big(\sum_{i=1}^{N}(\log y_i - \log f(\theta, t_i))^2 + v \cdot u\big)\Big\}, \tag{3.5}$$

where $\log y = (\log y_1, \log y_2, \ldots, \log y_N)$. Because of the nonlinear MM equation, the conditional distribution of θ is nonstandard and known up to a normalizing constant. For this reason, we use the random walk Metropolis (RWM) method to draw samples from the full conditional distribution of θ. More detailed random walk description in a PK model can be found in Kim, Hall, and Li (2009).

3.2. Different Monte Carlo Markov chain sampling scheme

Liu, Wong, and Kong (1994) introduced two Gibbs samplers and examined them based on operator theory. To illustrate, we consider a three-dimensional distribution $\pi(\theta_1, \theta_2, \theta_3)$.

Scheme 1: Single component Gibbs sampler

$$(1)\ \theta_1 \sim \pi(\theta_1 | \theta_2, \theta_3);\ (2)\ \theta_2 \sim \pi(\theta_2 | \theta_1, \theta_3);\ (3)\ \theta_3 \sim \pi(\theta_3 | \theta_1, \theta_2). \tag{3.6}$$

Scheme 2: Group component Gibbs sampler

$$(1)\ (\theta_1, \theta_2) \sim \pi(\theta_1, \theta_2 | \theta_3);\ (2)\ \theta_3 \sim \pi(\theta_3 | \theta_1, \theta_2). \tag{3.7}$$

Liu and his colleagues (Liu (1996); Liu, Wong, and Kong (1994)) studied the two schemes using spectral radii (the rate of convergence). They concluded that the second was faster in some cases, but otherwise it was not clear which scheme had the advantage. Further discussion was done by Roberts and Sahu (1997) with a target normal distribution for $\pi(\theta_1, \theta_2, \theta_3)$. In particular, they found that Scheme 2 has the faster rate of convergence if all partial correlations of the Gaussian target density are non-negative. At the same time, they found that Scheme 1 can also be the faster under a condition on the negative partial correlations. We investigate the speed of the convergence of Schemes 1 and 2 using the general MCMC algorithm under statistical identifiability given prior information. From now on, we designate the Single Component MCMC as SCM and the Group Component MCMC as GCM.

It has been shown that unidentifiable parameters are generally highly correlated, and MCMC methods converge slowly (Rannala (2002)). It is also rarely possible to determine statistical identifiability without fitting a model (Hengl et al. (2007)). It is desirable to have a method that can estimate the parameters regardless of the statistically identifiable situation. We propose a novel switching algorithm to fulfill this need.

In nonlinear models, we expect that SCM is a more efficient sampling scheme than GCM in unidentifiable models, while GCM may be more efficient than SCM in identifiable models. Our idea is to take an average of the different behaviors of GCM and SCM. To do this, we propose drawing samples randomly with switching between GCM and SCM. The SWM algorithm is an MCMC algorithm that does this, as follows.

Step 1: Set a switching probability s_p.

Step 2: Generate a random number $u \sim U(0, 1)$

$$\begin{cases} \text{Generate a sample using GCM if } u \geq s_p, \\ \text{Generate a sample using SCM if } u < s_p. \end{cases}$$

Step 3: Repeat *Step* 2 until the number of iterations is reached.

The SWM has as tuning parameter the switching probability s_p. GCM, SCM, and SWM are implemented with RWM. In GCM, we directly use (3.4); however, in SCM, it is divided into two components as

$$\begin{cases} [\theta|\log y, \sigma^2] \propto \exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^{N}\big(\log y_i - \log f(\theta, t_i)\big)^2 - \frac{1}{2}(\theta - d)'D^{-1}(\theta - d)\}, \\ [\theta|\log y, \sigma^2] \propto \exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^{N}\big(\log y_i - \log f(\theta, t_i)\big)^2 - \frac{1}{2}(\theta - d)'D^{-1}(\theta - d)\}. \end{cases} \tag{3.8}$$

SCM iterates (3.5) along with (3.8) to obtain the required estimates by RWM, while GCM iterates (3.4) and (3.5). In addition, SWM iterates between GCM and SCM.

4. Evaluation Criteria for MCMC Convergence

4.1. Introduction to convergence evaluation for MCMC in linear models

We revisit the convergence rates of SCM and GCM in the linear models of Section 3.2. We focus on a three-dimensional Gaussian target distribution with the inverse covariance-variance matrix

$$Q = \begin{bmatrix} 1 & q_{12} & q_{13} \\ q_{12} & 1 & q_{23} \\ q_{13} & q_{23} & 1 \end{bmatrix} (= \Sigma^{-1}). \tag{4.1}$$

The convergence rate of the GCM that blocks the first two components together is $1 - |Q|/(1 - q_{12}^2)$, and the SCM's convergence rate is $\max\{(b \pm \sqrt{b^2 - 4c})/2\}$, Roberts and Sahu (1997), where $b = q_{12}^2 + q_{13}^2 + q_{23}^2 - q_{12}q_{13}q_{23}$ and $c = q_{12}q_{13}q_{23}$. Using these results and the scaled inverse matrix (4.1), the convergence rates are examined in Figure 1. Without loss of generality, convergence rates were compared according to q_{12} and q_{13} ($-1 \leq q_{12}, q_{13} \leq 1$) given $q_{23} = \{-0.99, -0.7, -0.3, -0.01, 0, 0.01, 0.3, 0.7, 0.99\}$. The unidentifiable parameters are characterized by q_{12}, which is either close to 1 or -1. Based on Figure 1, the relative convergence rates between GCM and SCM depend heavily on q_{23}: when q_{23} is close to 0, GCM is faster than SCM almost all the time; when q_{23} is away from 0, a faster convergence rate is that of GCM or SCM. It is difficult to specify conditions for it.

Figure 1 suggests that the selection of a fast MCMC scheme is problem-dependent, even in the linear models. In nonlinear models, additional challenges appear. Since there is no closed form of marginal or conditional distributions of model parameters in nonlinear models, their convergence speeds cannot be assessed as (4.1) in the linear models. Some practical evaluation criteria of convergence, such as the method proposed by Raftery and Lewis (1992), have been routinely used. It is important then to compare the consistency between practical and theoretical convergence rates of MCMC algorithms for nonlinear models.

4.2. The consistency between the Raftery/Lewis and Roberts/Sahu convergence criteria

We use the notation of Raftery and Lewis (1992). Their method is based on the estimation of a particular quantile of some scalar function $\theta(X)$. The method

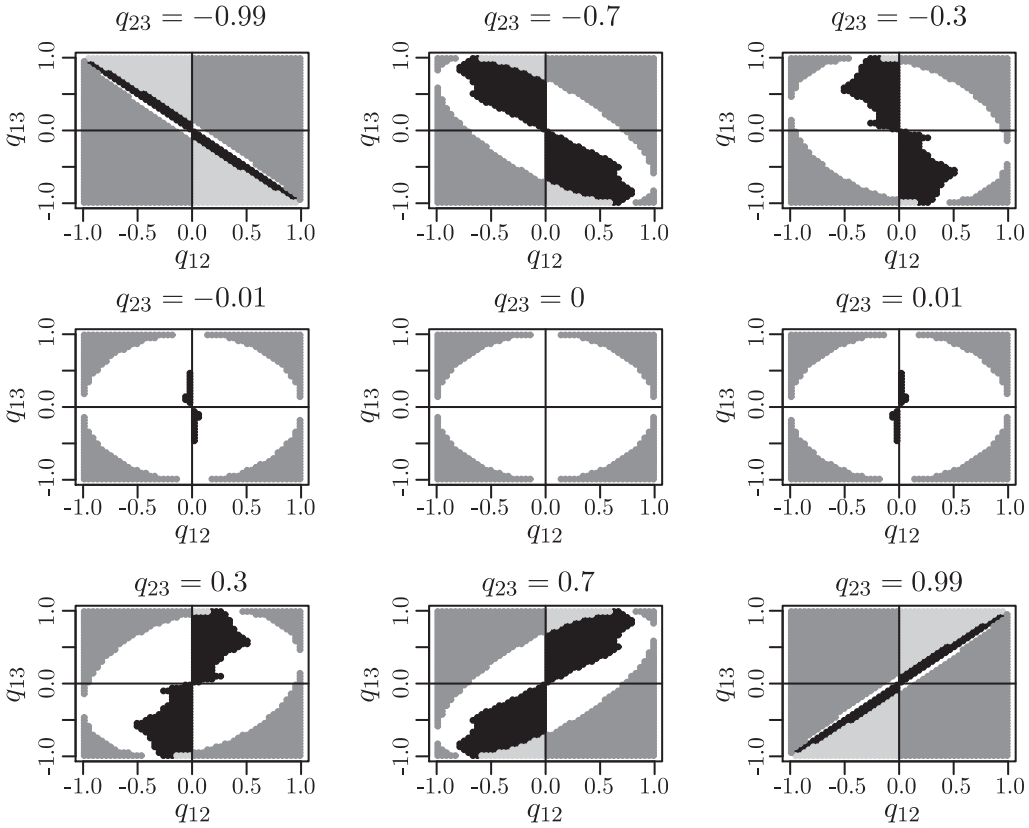

Figure 1. Comparison of the convergence rates of Schemes 1 and 2 of the Gibbs sampler based on the scaled inverse covariance-variance matrix. The inverse matrix is positive definite in the black and white regions and not positive definite in the darkgray and gray regions. Scheme 1 is faster than Scheme 2 in the black and darkgray regions, while Scheme 2 is faster than Scheme 1 in the white and gray regions.

first calculates $\theta(X^t)$ at each iteration t, then takes $Z_t = \delta(\theta(X^t) \leq u)$, where $\delta(\cdot)$ is the indicator function. The sequence $\{Z_t\}$ is not a Markov chain, but the new sequence $Z_t^{(k)} = Z_{(1+(t-1)k)}$ is approximately a Markov chain. Practically, k is a small number (El Adlouni, Favre, and Bobee (2006); Cowles and Carlin (1996)). The Raftery and Lewis (1992) approach is implemented with sequence $Z_t (= Z_t^{(1)})$, as suggested in El Adlouni, Favre, and Bobee (2006). Let

$$P = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix} \tag{4.2}$$

be the transition matrix for the sequence Z_t, where $\alpha = Prob(Z_t = 1 | Z_t = 0)$

and $\beta = Prob(Z_t = 0|Z_t = 1)$. The required number of iterations to converge is

$$n = n(\alpha, \beta, \epsilon, p, r) = \frac{\log((\alpha+\beta)\epsilon/\max(\alpha,\beta))}{\log|1-\alpha-\beta|} + \frac{(2-\alpha-\beta)\alpha\beta(\Phi^{-1}((p+1)/2))^2}{r(\alpha+\beta)^3},$$

where Φ^{-1} is the inverse standard normal cumulative distribution function. Note that it is clear that the convergence rate of the sequence Z_t is $\rho = 1-\alpha-\beta$, in view of P.

Lemma 1. *If $\beta > \alpha(4-\alpha)/(\alpha+2)$ or $\alpha > \beta(4-\beta)/(\beta+2)$, then $n(\alpha, \beta, p, r)$ is a monotonically decreasing function with respect to $\alpha+\beta$, where $0 < \alpha+\beta < 1$, $\alpha,\ \beta > 0$.*

Proof. Without loss of generality, take α, ϵ, p, r fixed, and $t = \alpha+\beta(< 1)$. Then $n(t)$ is defined as

$$n(t) = n_1(t) + n_2(t) = \frac{\log(t\cdot\epsilon/\max(\alpha, t-\alpha))}{\log|1-t|} + \frac{(2-t)\alpha(t-\alpha)(\Phi^{-1}\big((p+1)/2)\big)^2}{r\cdot t^3}. \tag{4.3}$$

It is not difficult to show that $n_1(t)$ is monotonically decreasing with respect to $t(< 1)$. On the other hand, with additional calculation, the maximum of $n_2(t)$ is $6\alpha/(\alpha+2)$. Therefore, if $t > 6\alpha/(\alpha+2)$, $n_2(t)$ is monotone decreasing.

A corollary states the consistency of the theoretical and practical convergence of an MCMC algorithm.

Corollary 1. *Suppose P_1 and P_2 are transition matrices with transition probabilities (α_1, β_1) and (α_2, β_2), respectively, and likewise the transition probabilities satisfy the conditions of the Lemma. Then $\rho_1 \geq \rho_2$ if and only if $n_1 \geq n_2$, where ρ_i,n_i are the convergence rate and the required number of iterations for P_i,$i = 1, 2$.*

Although the method of Raftery and Lewis (1992) can underestimate the convergence rate, as argued by Brooks and Roberts (1999), the order of the convergence rates between the Raftery and Lewis method and the true theoretical convergence rate are consistent under mild conditions. Indeed, all the simulations in Brooks and Roberts (1999) and El Adlouni, Favre, and Bobee (2006) support our argument (for our case, see the transition probabilities probabilities α and β that are necessary for estimating the required number of iterations (I) in Table 1.). Accordingly, we use the method of Raftery and Lewis (1992) for the purpose of comparing the speed of the convergence among different MCMC schemes.

Table 1. Experiment with MDZ data: The estimates of means and standard deviations (mean ± SD) of the parameters $\log V_1$, $\log V_2$, $\log Vmax$, $\log Km$, $\log CL_{12}$, σ^2, the ratio of maximum of concentration to estimated Km, recommended numbers of iterations (I), acceptance rate (AR) for SCM, GCM, and SWM, and α, β, ρ $(= \alpha + \beta)$ according to the RWM constant $c = 1$ and optimal ARs ($c = 0.15$, 0.004, 0.05 for SCM, GCM, SWM, respectively).

	Optimal ARs ($c = 0.15$; 0.004; 0.05)		
	SCM	GCM	SWM
$\log V_1$	4.32±0.07	4.63±0.20	4.33±0.08
$\log V_2$	4.61±0.39	2.90±0.32	4.59±0.46
$\log Vmax$	6.88±0.43	6.79±0.32	6.72±0.31
$\log Km$	1.75±1.23	0.86±0.71	1.13±1.04
$\log CL_{12}$	3.13±0.23	3.53±0.20	3.21±0.22
σ^2	0.004±0.006	0.091±0.101	0.006±0.010
$\frac{\max_t C(t)}{Km}$	10.0	24.4	18.6
I	34.1	186.8	59.8
AR	0.239	0.227	0.223
α	0.025	0.024	0.024
β	0.962	0.937	0.945
$\rho(= \alpha + \beta)$	0.987	0.961	0.969

5. MDZ Data Analysis and MM Simulation Studies

5.1. MDZ data analysis

The MDZ PK model described in Sections 2 and 3 was employed for data analysis. Only one individual was selected, which PK model might be statistically unidentifiable. The previously published MDZ PK meta-analysis (Li et al. (2007)) was utilized for prior distributions. The following PK parameter values were implemented in the prior distribution (8): $d = [\log 67.05, \log 45.08, \log 4284, \log 2.11, \log 42.9]'$; $v = 4$; $u = 11/2,000$;

$$D = \begin{bmatrix} 0.5 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0.5 \end{bmatrix}.$$

Two runs of 100,000 iterations were conducted, with burn-in $= 1,000$ and thin $=$ 5. The RWM was set up such that $c = 0.15$, 0.004, and 0.05 for SCM, GCM, and SWM, respectively, for which similar optimal acceptance rates (AR) are achieved (Roberts, Gelman, and Gilks (1997)) as shown in Table 1.

In parameter estimation (Table 1) and model fitting (Figure 2(a)), the Km estimates are smaller than the observed concentrations, $\max_t C(t)/Km$ ranges 10.0 to 24.4. This suggests that Km may not be identifiable from $Vmax$. Table 1 presents the recommended number of iterations (I) and convergence rate (ρ)

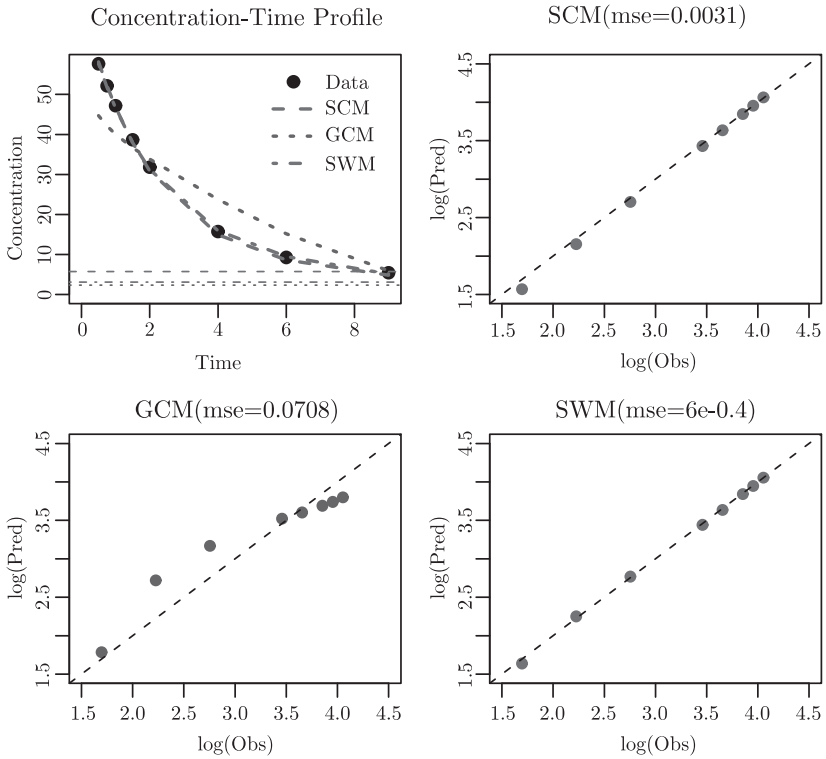

Figure 2(a). Experiment with MDZ data: Concentration-Time Profile and Mean Squared Error (a) and MCMC Trace plots (b) when RWM constant c has optimal ARs (c = 0.15, 0.004, 0.05 for SCM, GCM, SWM, respectively). The red is of SCM, green of GCM, and blue of SWM. The estimates of Km for each method are depicted as the dotted horizontal lines in the concentration-time profile (top-left plot) in (a). Their estimates of Km are depicted in Table 1.

for each method. SCM has five times smaller I than GCM (34.1 vs 186.8), while SCM has larger ρ than GCM (0.987 vs 0.961). These results are consistent with our theoretical results. As expected, SWM's I (59.8) and ρ (0.969) fall between those of SCM and GCM. It should be noted that the recommended number of iterations (I) is the normalized measure which is the ratio of the total number of iterations to converge and the lower bound of the number of iterations to converge.

Figure 2(b) shows the trace plots for all parameters using SCM, GCM, and SWM on the PK MDZ data. The parameter traces of GCM (green plot) have relatively poor mixing and convergence. The estimates of V_2 and Km of GCM are smaller than those of SCM and SWM, as shown in Table 1. The mean squared errors (MSEs) of SCM, GCM, and SWM are 0.0031, 0.0708, and 0.0006,

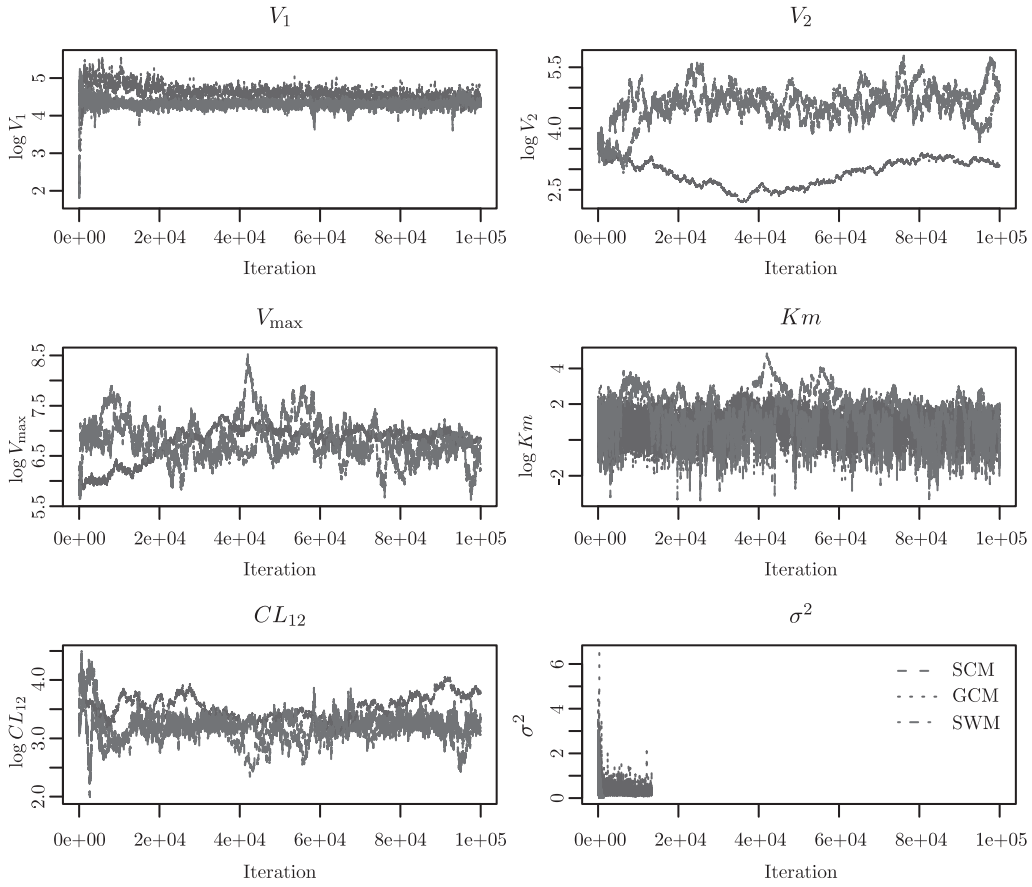

Figure 2(b). Experiment with MDZ data: Concentration-Time Profile and Mean Squared Error (a) and MCMC Trace plots (b) when RWM constant c has optimal ARs (c = 0.15, 0.004, 0.05 for SCM, GCM, SWM, respectively). The red is of SCM, green of GCM, and blue of SWM. The estimates of Km for each method are depicted as the dotted horizontal lines in the concentration-time profile (top-left plot) in (a). Their estimates of Km are depicted in Table 1.

respectively. In addition, SWM fits better than the others, likely due to its better mixing procedure that uses both GCM and SCM.

5.2. MM simulation model and its description

To validate the performances of GCM, SCM, and SWM, simulation studies were designed and compared. Two datasets (identifiable and unidentifiable) were generated from the MM equation with N (the number of time points) = 20 and fixing σ^2 as 0.2^2 for simplicity. The concentration C was generated between 0.01

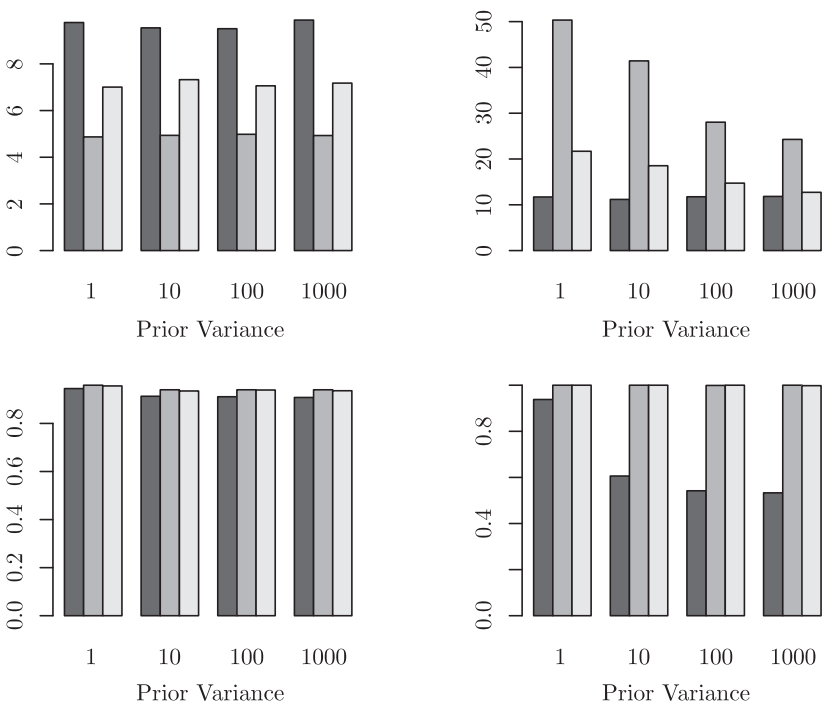

Figure 3. Simulation: Recommended numbers of iterations (I; first row) and 95% coverage probability (CP; second row) for the MM constant Km of SCM (dark bar), GCM (bright bar), and SWM (brightest bar), according to the identifiable model (left column) and the non-identifiable model (right column).

and 1 with equal size, and the true value of $Vmax$ was 1 in both cases. In the simulation, $Km = 1$ for the identifiable model, $Km = 100$ for the unidentifiable model. The variance component of the prior information was 1, 10, 10^2, and 10^3, to reflect how informative is the prior. The initial values of parameters were taken as the true values. To implement this simulation, we used the expressions (3.1)−(3.5), (3.8) with

$$\theta = (\theta_1, \theta_2)' = (\log Vmax, \log Km)'; f(\theta, C) = \frac{Vmax \cdot C}{Km + C}. \tag{5.1}$$

In each analysis, a run of 100,000 iterations was conducted, with burn-in = 3,000 and thin = 5. The performance of the recommended number (I) and 95% Coverage Probability (CP) of the MM constant Km were compared among SCM, GCM, and SWM.

Figure 3 presents the average of the total number of iterations (I) recommended for Km by the method of Raftery and Lewis (1992). In case of the

statistically identifiable model, GCM was almost two times faster than SCM (SCM/GCM = 2.00, 1.93, 1.91, 2.00 for prior variances of $1, 10^1, 10^2, 10^3$), while GCM was roughly three times slower than SCM (GCM/SCM = 4.30, 3.71, 2.39, 2.05) when the model was statistically unidentifiable. Not surprisingly, the proposed SWM had a robust performance in both situations, with recommended numbers of iterations (I) between those of SCM and GCM.

Comparing 95% CP (the second row in Figure 3), only SCM had a strong dependency on the prior variance when the model was not identifiable: 95% CP = (93.8%, 60.6%, 54.2%, 53.5%) with prior variance = ($1,10^1,10^2,10^3$), respectively. It is worth noting that GCM had better 95% CPs than SCM even when the model was unidentifiable (100%, 100%, 99.9%, 100%). Although SWM is adapted from GCM and SCM, SWM performed well (95.6%, 93.5%, 93.9%, 93.6%; 100%, 100%, 100%, 99.8%).

6. Conclusions

We have discussed the impact of unidentifiable nonlinear models and parameter estimation on MCMC algorithms.

We have shown that, under certain conditions, the convergence diagnosis of Raftery and Lewis (1992) is consistent with that argued by Brooks and Roberts (1999). The former diagnosis can generally be used for MCMC algorithms of both of linear and nonlinear models, while the latter is derived only for linear models. Our work bridges these two. Two MCMC schemes, GCM and SCM, and the proposed method, SWM, were compared in identifiable or unidentifiable nonlinear models with real data analysis and simulation studies.

Using one subject's data from the MDZ pharmacokinetic study and its two compartmental model, we have found evidence that Km is unidentifiable. Data analysis further confirmed that SCM converges faster than GCM when the model is unidentifiable, and SWM's performance falls in between that of SCM and GCM. Both SCM and SWM provided better fits than GCM in the data analysis.

In the simulation studies, when the model is identifiable, GCM is faster than SCM and SWM. They all have comparable coverage probabilities, however. When the model is unidentifiable, SCM is faster than GCM, but SCM has worse coverage probability than GCM and SWM. On the other hand, SWM shows better performance than GCM in both convergence rate and coverage probability.

In conclusion, the SCM is a fast MCMC scheme for statistically unidentifiable nonlinear models. In order to estimate parameters by SCM, one should, however, recognize the identifiability of the model, and that might be impossible to know before parameter estimation. On the other hand, the proposed SWM shows a robust performance for both identifiable and non-identifiable situations in data analysis and in simulation. Especially, it is faster and better mixing than

GCM, and its coverage probability is better than SCM when the model is not identifiable. Furthermore, it possesses the great advantage that can be applied to estimate parameters regardless of the statistically identifiability situation.

Acknowledgement

Dr. Seongho Kim is supported by DOE grant, DE-EM0000197, and an Intramural Research Incentive Grant from the University of Louisville (S.K.). Dr. Lang Li is supported by NIH RO1 GMS 74217 (L.L.).

References

Atkinson, A. J., Daniels, C. E., Dedrick, R. L., Grudzinskas, C. V. and Markey, S. P. (2001). *Principles of Clinical Pharmacology*. Academic Press, San Diego.

Brooks, S. and Roberts, G. (1999). On quantile estimation and MCMC convergence. *Biometrika* **86**, 710-717.

Casella, G. and Berger, R. L. (1990). *Statistical inference*. Duxbury Press, Belmont, CA.

Chang, M. (2010). *Monte Carlo Simulation for the Pharmaceutical Industry: Concepts, Algorithms, and Case Studies*. CRC Press; Boca Raton, FL.

Cobelli, C. and DiStefano III, J. (1980). Parameter and structural identifiability concepts and ambiguities: a critical review and analysis. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **239**, 7-24.

Cowles, M. K. and Carlin, B. P. (1996). Markov chain Monte Carlo convergence diagnostics: A comparative review. *J. Amer. Statist. Assoc.* **91**, 883-904.

Eberly, L. E. and Carlin, B. P. (2000). Identifiability and convergence issues for Markov chain Monte Carlo fitting of spatial models. *Statist. Med.* **19**, 2279-2294

El Adlouni, S., Favre, A. and Bobee, B. (2006). Comparison of methodologies to assess the convergence of Markov chain Monte Carlo methods. *Computat. Statist. Data Anal.* **50**, 2685-2701.

Gelfand, A. E. and Sahu, S. K. (1999). Identifiability, improper priors, and Gibbs sampling for generalized linear models. *J. Amer. Statist. Assoc.* **94**, 247-225.

Gelfand, A. E., Sahu, S. K. and Carlin, B. P. (1995). Efficient parametrizations for normal linear mixed models. *Biometrika* **82**, 479-488.

Gilks, W. R., Richardson, S. and Spiegelhalter, D. G. (1996). Markov Chain Monte Carlo in practice. Chapman and Hall, London.

Godfrey, K. R. and Fitch, W. R. (1984). The deterministic identifiability of nonlinear pharmacokinetic models. *J. Pharmacokin. Biopharm.* **12**, 177-191.

Gorski, J. C., Jones, D. R., Haehner-Daniels, B. D., Hamman, M. A., O'Mara Jr., E. M. and Hall, S. D. (1998). The contribution of intestinal and hepatic CYP3A to the interaction between midazolam and clarithromycin. *Clin Pharmacol Ther* **64**, 133-143.

Hein, G. E. and Niemann, C. (1962). Steric course and specificity of a Chymotrypsin catalyzed reactions. I. *J. Amer. Chem. Soc.* **84**, 4487-4494.

Hengl, S., Kreutz, C., Timmer, J. and Maiwald, T. (2007). Data-based identifiability analysis of non-linear dynamical models. *Bioinformatics* **23**, 2612-2618.

Kim, S., Hall, S. D. and Li, L. (2009). A novel Gibbs maximum a posteriori (GMAP) approach on Bayesian nonlinear mixed-effects population pharmacokinetics (PK) models. *J. Biopharma. Stat.* **19**, 700-720

Li, L., Yu, M., Chin, R., Lucksiri, A., Flockhart, D. A. and Hall, S. D. (2007). Drug-drug interaction prediction: a Bayesian meta-analysis approach. *Stat. Med.* **26**, 3700-3721

Lindley, D. V. (1971). *Bayesian Statistics: A review*. SIAM, Philadelphia, PA.

Liu, J. S. (1996). Metropolized independent sampling with comparisons to rejection sampling and importance sampling. *Statist. Comput.* **6**, 113-119.

Liu, J. S., Wong, W. H. and Kong, A. (1994). Covariance structure of the Gibbs sampler with applications to the comparisons of estimators and augmentation schemes. *Biometrika* **81**, 27-40

Miao, H., Dykes, C., Demeter, L. M. and Wu, H. (2009). Differential equation modeling of HIV viral fitness experiments: model identification, model selection, and multimodel inference. *Biometrics* **65**, 292-300.

Monod, J. (1949). The growth of bacterial cultures. *Ann. Rev. Microb.* **3**, 371-394

Neath, A. and Samaniego, F. J. (1997). On the efficacy of Bayesian inference for nonidentifiable models. *Amer. Statist.* **51**, 225-232.

Poirier, D. J. (1998). Revising beliefs in non-identified models. *Econometric Theory* **14**, 483-509

Rannala, B. (2002). Identifiability of parameters in MCMC Bayesian inference of phylogeny. *Syst. Biol.* **51**, 754-760.

Raftery, A. E. and Lewis, S. M. (1992). One long run with diagnostics: implementation strategies for Markov chain Monte Carlo. *Statist. Sci.* **7**, 493-497.

Roberts, G. O., Gelman, A. and Gilks, W. R. (1997). Weak convergence and optimal scaling of random walk Metropolis algorithms. *Ann. Appl. Prob.* **7**, 110-20.

Roberts, G. O. and Sahu, S. K. (1997). Updating schemes, correlation structure, blocking and parameterization for the Gibbs sampler. *J. Roy. Statist. Soc. Ser. B* **59**, 291-317.

Xie, Y. and Carlin, B. P. (2006). Measures of Bayesian learning and identifiability in hierarchical models. *J. Statist. Plann. Inference* **136**, 3458-3477.

Department of Bioinformatics and Biostatistics, University of Louisville, Louisville, KY, 40202, USA.

E-mail: s0kim023@louisville.edu

Department of Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, IN, 46032, USA.

E-mail: lali@iupui.edu

(Received November 2010; accepted July 2011)