The Canadian Journal of Statistics
Vol. 31, No. 4, 2003, Pages 469–482
La revue canadienne de statistique

A geometric approach to transdimensional Markov chain Monte Carlo

Giovanni PETRIS and Luca TARDELLA

Key words and phrases: Bayesian inference; Markov chain Monte Carlo; multimodel inference; regeneration; transdimensional Markov chain Monte Carlo.

MSC 2000: Primary 62F15; secondary 65C05.

Abstract: The authors present theoretical results that show how one can simulate a mixture distribution whose components live in subspaces of different dimension by reformulating the problem in such a way that observations may be drawn from an auxiliary continuous distribution on the largest subspace and then transformed in an appropriate fashion. Motivated by the importance of enlarging the set of available Markov chain Monte Carlo (MCMC) techniques, the authors show how their results can be fruitfully employed in problems such as model selection (or averaging) of nested models, or regeneration of Markov chains for evaluating standard deviations of estimated expectations derived from MCMC simulations.

Une approche géométrique pour les méthodes de Monte-Carlo à chaîne de Markov

Résumé : Les auteurs présentent des résultats théoriques qui montrent comment il est possible de simuler un mélange de lois dont les composantes vivent dans des sous-espaces de dimensions différentes en reformulant le problème de sorte que les observations puissent être tirées d'une loi continue auxiliaire définie sur le plus grand sous-espace et ensuite transformées de façon appropriée. Motivés par l'importance d'élargir l'éventail disponible de méthodes de Monte-Carlo à chaîne de Markov (MCCM), les auteurs montrent comment leurs résultats peuvent être mis à profit, entre autres, dans des situations de choix (ou de compromis) entre divers modèles emboîtés ou de regénération de chaînes de Markov pour l'évaluation de l'écart type d'estimations d'espérances déduites de simulations par MCCM.

1. INTRODUCTION

The use of Markov chain Monte Carlo (MCMC) algorithms has become more and more important in the last decade. In particular, they are used as the main tool to cope with statistical problems where a distribution is known only up to a proportionality constant and some approximation to its features is desired. The wide availability of algorithms that are easy to implement, such as the Gibbs sampler and some of its variants, have allowed for an exponential growth of the routine application of simulation-based Bayesian inference.

The theoretical development of these techniques and the enlargement of their scope of application have been boosted by their extension to distributions supported on subspaces of variable dimension starting from Carlin & Chib (1995) and the reversible jump of Green (1995). Since then, a lot of research effort has been devoted to enlarging the availability of transdimensional simulation techniques as well as enhancing the applicability of those currently available and attempting to overcome some of their intrinsic difficulties. However, it is apparent that a lot of expertise with these techniques is needed for a safe and successful implementation, and often problem-specific difficulties can arise (Godsill & Troughton 1998; Dellaportas, Forster & Ntzoufras 2002; Rotondi 2002). There is a clear need in the statistics community for generic approaches to Bayesian multimodel inference that are easy to implement. This is also testified by current ongoing research mainly aimed at proposing improved variants of the reversible jump (Al-Awhadhi, Hurn & Jennison 2001; Brooks, Giudici & Roberts 2003; Green 2003).

One of the most interesting alternatives to the reversible jump, which is currently the most popular technique for transdimensional Markov chain Monte Carlo, is probably the birth-and-death approach of Stephens (2000). A detailed comparison of the two approaches in the context

This content downloaded by the authorized user from 192.168.72.233 on Wed, 14 Nov 2012 14:05:31 PM
All use subject to JSTOR Terms and Conditions

of mixture model inference is contained in Cappé, Robert & Rydén (2003). The approach of Stephens (2000), however, suffers from the lack of efficient natural ways of specifying the birth proposals in a general context.

In this paper, we develop the theoretical basis for an original, more automatic approach to multimodel Markov chain Monte Carlo for the nested-model case. We focus on the problem of generating a sample from a distribution $\bar{\mu}$—typically the posterior distribution of a parameter of interest—supported by a sequence of nested hyperplanes in $\mathbb{R}^K$. First we show how to construct, in a natural fashion, an absolutely continuous distribution $\bar{\tau}$ on $\mathbb{R}^K$ and a transformation ϕ from $\mathbb{R}^K$ to itself such that $\bar{\tau}\phi^{-1}$ is equal to the target distribution $\bar{\mu}$. In this way, even if one is not able to simulate directly from $\bar{\tau}$, one can easily generate a finite realization $\zeta_1, \ldots, \zeta_N$ of an ergodic Markov chain having limiting distribution $\bar{\tau}$, and then approximate a sample from $\bar{\mu}$ with $\phi(\zeta_1), \ldots, \phi(\zeta_N)$. We also show how to use the same basic device as a building block to construct a Markov kernel with limiting distribution $\bar{\mu}$ directly in the original parameter space, and we show how this can be useful within a Gibbs sampler (Section 3) and for simulating a regenerative chain (Section 4).

The distinguishing feature of our proposal is the automatic transformation of the mixture of the posterior densities appearing in the nested models into a comprehensive global density which has the further advantage of offering a geometric intuition of what is behind the jumps between different models that are to be simulated. Furthermore, with the approach we propose, there is no need to evaluate Jacobians of involved transformations. We believe the availability of effective and more automatic alternatives to the reversible jump might help many researchers to entertain nested models in a fully Bayesian fashion through Markov chain Monte Carlo approximations.

The layout of the paper is as follows. In Section 2, we present the main theoretical results which the proposed method relies upon. Section 3 illustrates, in the controlled context of nested linear models, the performance of our proposal for a simulation-based multimodel Bayesian inference. The choice of this relatively simple model for illustrative purposes is suggested by the availability, when using a natural conjugate prior, of posterior model probabilities. This allows us to compare the true posterior with the simulation results based on our proposed method. For the data analyzed in this example, we show that our method produces reliable estimates of the posterior distribution.

In Section 4, we explore a different kind of application of the theoretical results developed in Section 2. Starting from an idea by Brockwell & Kadane (2002), we show an alternative algorithm for simulating a regenerative Markov chain with a prescribed limiting distribution. In the final section, we examine the features of our strategy in the broader context of currently available techniques. In addition, we mention other applications and some possible extensions of the proposed sampler.

2. MAIN RESULTS

We start by illustrating the core idea in the simple case of an unnormalized probability distribution for θ, having an absolutely continuous component and a component degenerate at one point. In what follows, for the sake of notational simplicity and without loss of generality, we assume that point to be the origin. Let η_K and δ_K denote K-dimensional Lebesgue measure and Dirac measure (concentrated on the origin of $\mathbb{R}^K$), let f_0 be a measurable positive function defined on $\mathbb{R}^K$ and let f_K be a constant. Consider the measure on $\mathbb{R}^K$

$$\mu(d\theta) = f_0(\theta)\eta_K(d\theta) + f_K\delta_K(d\theta), \tag{1}$$

that we assume to be finite, but not necessarily a probability measure. We use the notation $\bar{\mu}$ for the probability proportional to μ, i.e.,

$$\bar{\mu}(\cdot) = \frac{\mu(\cdot)}{\mu(\mathbb{R}^K)}.$$

This content downloaded by the authorized user from 192.168.72.233 on Wed, 14 Nov 2012 14:05:31 PM
All use subject to JSTOR Terms and Conditions

Our first goal is to determine an absolutely continuous measure τ on $\mathbb{R}^K$ and a function $\phi: \mathbb{R}^K \to \mathbb{R}^K$ having the property that $\tau\phi^{-1}(B) = \mu(B)$ for every Borel set B. Note that this implies that the total mass of the measure is preserved, i.e., $\mu(\mathbb{R}^K) = \tau(\mathbb{R}^K)$. Moreover, if a random vector $\tilde{\zeta}$ has distribution proportional to τ, then the distribution of $\tilde{\theta} = \phi(\tilde{\zeta})$ is proportional to the original μ. In order to define the function ϕ, let $B_K(r) = \{\zeta \in \mathbb{R}^K : \|\zeta\| \le r\}$ be the K-dimensional closed ball of radius r, centered at the origin, and consider the radial contraction

$$\psi_K(\zeta; r) = \frac{\zeta}{\|\zeta\|}\left(\|\zeta\|^K - r^K\right)^{1/K}, \quad \zeta \in \mathbb{R}^K,\ \zeta \notin B_K(r).$$

The inverse function, defined for $\theta \neq 0$, is the radial expansion

$$\psi_K^{-1}(\theta; r) = \frac{\theta}{\|\theta\|}\left(\|\theta\|^K + r^K\right)^{1/K}.$$

It is not difficult to prove, using polar coordinates, that for any r, both ψ_K and ψ_K^{-1} preserve the Lebesgue measure. Thus, roughly speaking, one can use ψ_K^{-1} to move the absolutely continuous part of μ away from the origin, temporarily leaving an empty ball $B_K(r)$, and then spread the mass f_K corresponding to the degenerate component of μ uniformly into this ball. See Figure ?? and Figure 2 to visualize the procedure for $K = 1$ and $K = 2$.

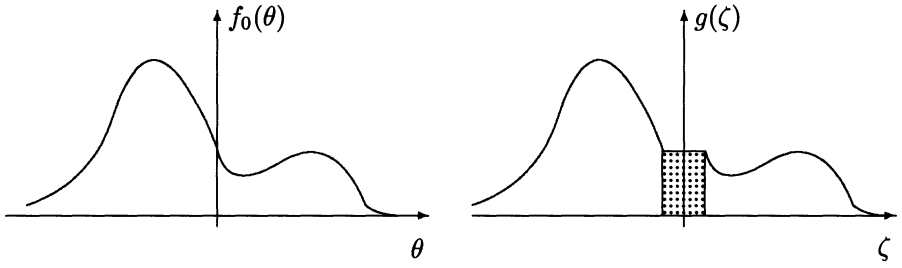

FIGURE 1: Plots of f_0 and g in the one-dimensional case ($K = 1$). The area of the dotted region is equal to the mass corresponding to the degenerate component f_K.

More formally, define

$$g(\zeta) = \begin{cases} cf_K & \text{if } \zeta \in B_K(r), \\ f_0\{\psi_K(\zeta; r)\} & \text{if } \zeta \notin B_K(r), \end{cases}$$

with c and r arbitrary positive constants satisfying $c\eta_K\{B_K(r)\} = 1$. Then, taking $\tau(d\zeta) = g(\zeta)\eta_K(d\zeta)$ and

$$\phi(\zeta) = \begin{cases} 0 & \text{if } \zeta \in B_K(r), \\ \psi_K(\zeta; r) & \text{if } \zeta \notin B_K(r), \end{cases}$$

we have exactly what we were looking for, namely an absolutely continuous measure τ and a function ϕ such that $\tau\phi^{-1} = \mu$. If f_0 is continuous, a default convenient choice for c and r can be derived by requiring that g be continuous as well. The value of g, as ζ approaches the boundary of the ball from outside, tends to $f_0(0)$, so this must be the constant value of g within the ball if one wants this function to be continuous. Hence, $c = f_0(0)/f_K$. Then, the necessary condition that the total mass assigned by the transformed density to the ball $B_K(r)$ equal f_K can be solved for r, giving $r = \left\{f_K\Gamma(K/2+1)/f_0(0)\right\}^{1/K}/\sqrt{\pi}$. Note that, with this choice of c and r, if f_0 is unimodal, then so is g.

This content downloaded by the authorized user from 192.168.72.233 on Wed, 14 Nov 2012 14:05:31 PM
All use subject to JSTOR Terms and Conditions

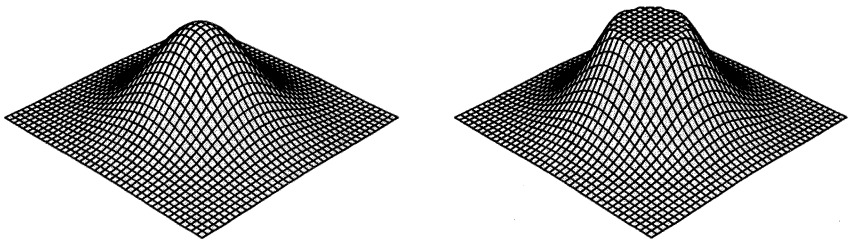

FIGURE 2: Plots of the surfaces of f_0 (left) and g (right) in the two-dimensional case ($K = 2$). Similarly to the one-dimensional case, the volume beneath the constant density in the center of the right surface is equal to the mass corresponding to the degenerate component f_K.

We now consider formally the more general case of a mixture of two nested components, one having full support, and the second supported by the hyperplane H_k, defined by constraining the last k coordinates to be zero. We will write $\theta = (\theta_{k,1}, \theta_{k,2})$ for the generic point of $\mathbb{R}^K$, with $\theta_{k,1}$ and $\theta_{k,2}$ being, respectively, the first $h = K - k$ and last k components. Consider, for a fixed k in $\{1, \ldots, K\}$, the finite measure

$$\mu(d\theta) = f_0(\theta)\eta_K(d\theta) + f_k(\theta_{k,1})\eta_h(d\theta_{k,1})\delta_k(d\theta_{k,2}), \tag{2}$$

where f_0 and f_k are measurable positive functions defined on $\mathbb{R}^K$ and $\mathbb{R}^h$, respectively. To clarify the notation, we note that the integer indexing each density is the dimension of the degenerate part of the corresponding component, i.e., the number of coordinates that are set to zero. The next theorem (see Appendix for a proof) shows how the function ψ_k can be used to transform the measure (2) into an absolutely continuous measure. The reader will realize that the idea sketched at the beginning of this section, visualized in Figure **??** and Figure 2, corresponds to the special case $h = 0$, $k = K$. Figure 3 illustrates what happens in the intermediate case $K = 2$ and $k = h = 1$.

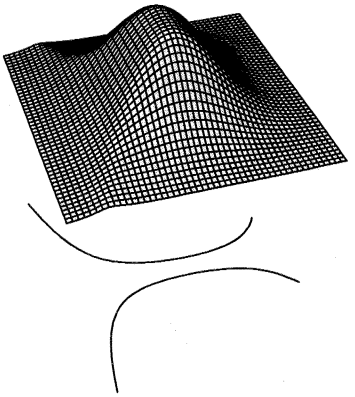

FIGURE 3: Plot of an instance of g as determined by Theorem 1 for $K = 2$ and $k = h = 1$: the component density f_0 is a standard bivariate normal; the component density f_1 is a univariate normal with mean -0.6 and variance 2. A random draw Z from g in the region between the two curves projected on the bottom of the surface corresponds to a random draw $\theta = \phi(Z)$ from the univariate component f_1.

This content downloaded by the authorized user from 192.168.72.233 on Wed, 14 Nov 2012 14:05:31 PM
All use subject to JSTOR Terms and Conditions

THEOREM 1. *Let c and r be two measurable positive functions defined on $\mathbb{R}^h$ such that $c\eta_k\{B_k(r)\} \equiv 1$. Define the transformation of $\mathbb{R}^K$ in itself*

$$\phi(\zeta) = \begin{cases} (\zeta_{k,1}, 0) & \text{if } \zeta_{k,2} \in B_k\{r(\zeta_{k,1})\}, \\ (\zeta_{k,1}, \psi_k\{\zeta_{k,2}; r(\zeta_{k,1})\}) & \text{if } \zeta_{k,2} \notin B_k\{r(\zeta_{k,1})\}. \end{cases}$$

Consider the density g defined on $\mathbb{R}^K$ by

$$g(\zeta) = \begin{cases} f_k(\zeta_{k,1})c(\zeta_{k,1}) & \text{if } \zeta_{k,2} \in B_k\{r(\zeta_{k,1})\}, \\ f_0\{\phi(\zeta)\} & \text{if } \zeta_{k,2} \notin B_k\{r(\zeta_{k,1})\}, \end{cases}$$

and let $\tau(d\zeta) = g(\zeta)\eta_K(d\zeta)$. Then $\tau\phi^{-1} = \mu$. Moreover, if f_0 and f_k are continuous and $c(\zeta_{k,1})f_k(\zeta_{k,1}) = f_0(\zeta_{k,1}, 0)$ for every ζ, then g is continuous.

Remark 1. When one has to deal with a target measure μ that has more than two components, supported by a family of nested hyperplanes, one can make use of Theorem 1 repeatedly. Suppose, for instance, that μ is given by

$$\begin{aligned} &f_0(\theta)\eta_K(d\theta) + f_1(\theta_{1,1})\eta_{K-1}(d\theta_{1,1})\delta_1(d\theta_{1,2}) + \cdots \\ &\quad \cdots + f_{K-1}\eta_1(d\theta_{K-1,1})\delta_{K-1}(d\theta_{K-1,2}) + f_K\delta_K(d\theta). \end{aligned} \tag{3}$$

One can first apply the theorem to the first two components, obtaining an absolutely continuous measure and a transformation of $\mathbb{R}^K$ in itself, say τ_1 and ϕ_1. If g_1 is the density of τ_1, one can then apply the theorem to the measure

$$g_1(\zeta)\eta_K(d\zeta) + f_2(\zeta_{2,1})\eta_{K-2}(d\zeta_{2,1})\delta_2(d\zeta_{2,2}),$$

to construct a measure τ_2 with density g_2 and a transformation ϕ_2. Proceeding in this way, one finally obtains an absolutely continuous measure $\tau = \tau_K$ with density $g = g_K$ and a function $\phi = \phi_1 \circ \cdots \circ \phi_K$ such that $\tau\phi^{-1} = \mu$. The pseudo-code of a simple algorithm that efficiently evaluates g and ϕ can be found in Petris & Tardella (2000).

Remark 2. The new density g of Theorem 1 can be thought of as a slight reshaping of the density f_0, the one with largest support, determined by the embedding of another density supported on a restricted space. This reshaping effect is less pronounced whenever the opening is made around a point with high f_0 density—keeping constant the value of the other density—or when the other density is small relatively to f_0. Geometrically, this is explained by the fact that the radius function $r(\cdot)$ of the opening determined by $\psi_k(\cdot, r(\cdot))$ is related to the value of the two densities according to

$$\frac{f_0(\zeta_{k,1}, 0)}{f_k(\zeta_{k,1})}\eta_k[B_k\{r(\cdot)\}] \equiv 1.$$

This suggests that in order to obtain a more regular shape of the resulting density g, one can reparameterize f_0 and f_k so that their maximum is attained at or near the origin. In the particular case of $k = K$, when f_K is a constant, this is particularly evident (see Figure 4).

Remark 3. In order to understand how practical our approach is, one has to consider that it is not difficult to write a generic function, based on the preceding remarks. Such a function takes as input the list of the densities $f_0, \ldots, f_k, \ldots, f_K$ and, possibly, a list of centering locations $x_1^0, \ldots, x_k^0, \ldots, x_K^0$ such that $f_k(x - x_k^0)$ has a mode around the origin, and returns, for any specified $x \in \mathbb{R}^K$, the density $g(x)$ and the transformed value $\phi(x)$. A function that evaluates ϕ^{-1} can be similarly coded. The generic functions described above have been written by the authors in R (Ihaka & Gentleman 1996) and are available upon request.

This content downloaded by the authorized user from 192.168.72.233 on Wed, 14 Nov 2012 14:05:31 PM
All use subject to JSTOR Terms and Conditions

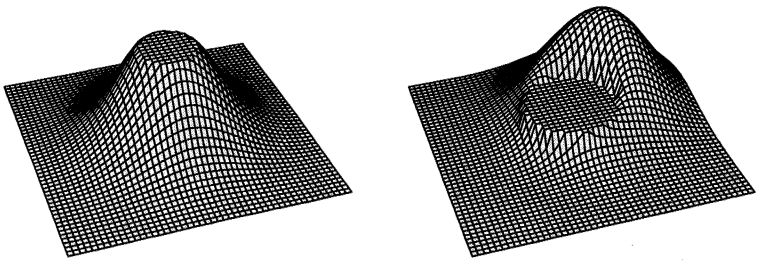

FIGURE 4: Plot of two instances of g as determined by Theorem 1 for $K = 2$ and $k = 2$: on the left, the component density f_0 is a standard bivariate normal and the degenerate component density f_2 is equal to 0.2; on the right, the component density f_0 is a bivariate normal with mean vector $(1, 1)$. The degenerate component density f_2 is equal to 0.2 in both cases.

When one has two models to compare, the smallest one being defined by the constraint $\theta_{k,2} = 0$ for a vector parameter $\theta = (\theta_{k,1}, \theta_{k,2})$, Theorem 1 can be used directly to find the density g. One can then simulate a Markov chain having g as its invariant density, using standard methods for Markov chain Monte Carlo in a Euclidean space. The stream of simulated random variables can then be transformed, applying the transformation ϕ, into a stream of random variables from a process that, although not Markovian, is ergodic and has the prescribed limiting distribution $\bar{\mu}$. When there are additional parameters ω that are common to the two models, so that the entire vector of parameters is (θ, ω), one can think of using again Theorem 1 inside a Gibbs sampler to transform the distribution of θ given ω—which has the form (2)—into a continuous distribution with density $g_\omega(\zeta)$. Updating θ from its full conditional distribution could be done through a Metropolis step on the transformed density $g_\omega(\zeta)$. However, some care is needed in determining the starting value, i.e., the current ζ^{cur}, for the Metropolis update, since ϕ_ω is not one to one. In fact, if the current θ-value of the chain, θ^{cur}, belongs to the smallest model, i.e., $\theta^{\text{cur}}_{k,2} = 0$, then there is no uniquely defined inverse mapping determining $\zeta^{\text{cur}} = \phi_\omega^{-1}(\theta^{\text{cur}})$, as any ζ^{cur} having the last k components inside the ball $B_k(r)$ will be mapped back to θ^{cur}. Theorem 2 guarantees that in this case, in order to preserve the Markov property of the resulting chain, it is enough to map $\theta^{\text{cur}}_{k,2} = 0$ to a randomly (uniformly) selected point $\zeta^{\text{cur}}_{k,2}$ in $B_k(r)$. From a general point of view, the problem can be formulated in terms of constructing a Markov kernel H in the original parameter space, having $\bar{\mu}$ as its invariant distribution, exploiting the density g and transformation ϕ as defined in Theorem 1 and a Markov kernel K in the auxiliary Euclidean space where g is defined, having $\bar{\tau}$ as its invariant distribution.

In order to give the result in a general form, let $(Z, \mathcal{S}_Z, \tau)$ and $(\Theta, \mathcal{S}_\Theta, \mu)$ be probability spaces, and let ϕ be a measurable function from Z onto Θ such that $\tau\phi^{-1} = \mu$. Let K be a transition kernel on $(Z, \mathcal{S}_Z)$ for which τ is invariant:

$$\tau(B) = \int_Z \tau(d\zeta) K(\zeta; B) \qquad \forall B \in \mathcal{S}_Z.$$

Consider $\tau^*(B \mid \phi(\zeta) = \theta)$, a regular version of τ given $\phi^{-1}\mathcal{S}_\Theta$, and define a transition kernel J from Θ to Z by setting

$$J(\theta; B) = \tau^*(B \mid \theta).$$

The following theorem is proved in the Appendix.

THEOREM 2. *Consider a Markov chain $\tilde{\theta}_0, \tilde{\theta}_1, \ldots$ on $(\Theta, \mathcal{S}_\Theta)$, whose transitions are described by the following steps:*

1. *Draw $\tilde{\zeta}_{t,0}$ according to $J(\tilde{\theta}_t; \cdot)$.*

This content downloaded by the authorized user from 192.168.72.233 on Wed, 14 Nov 2012 14:05:31 PM
All use subject to JSTOR Terms and Conditions

2. *Draw* $\tilde{\zeta}_{t,1}$ *according to* $K(\tilde{\zeta}_{t,0}; \cdot)$.

3. *Set* $\tilde{\theta}_{t+1} = \phi(\tilde{\zeta}_{t,1})$.

Let H *denote the corresponding transition kernel, i.e.,*

$$H(\theta; A) = \int_Z J(\theta; d\zeta) K(\zeta; \phi^{-1} A) \qquad \forall \theta \in \Theta,\ A \in \mathcal{S}_\Theta.$$

Then μ *is an invariant measure for* H.

In the setting of Theorem 1, it is easy to show that a regular conditional probability $\tau^*(B \mid \theta)$ is the distribution degenerate at $(\theta_{k,1}, \psi_k^{-1}\{\theta_{k,2}; r(\theta_{k,1})\})$, when $\theta_{k,2} \neq 0$, and the product of the distribution degenerate at $\theta_{k,1}$ for the first components and the uniform distribution over $B_k\{r(\theta_{k,1})\}$ for the last components, when $\theta_{k,2} = 0$. Hence the updating steps 1–3 of Theorem 2 work as follows: whenever the current $\theta^{\text{cur}} = (\theta_{k,1}^{\text{cur}}, \theta_{k,2}^{\text{cur}})$ has $\theta_{k,2}^{\text{cur}} = 0$, set in the first step $\zeta^{\text{cur}} = (\zeta_{k,1}^{\text{cur}}, \zeta_{k,2}^{\text{cur}})$ with $\zeta_{k,2}^{\text{cur}}$ randomly drawn from a uniform distribution over the ball $B_k\{r(\theta_{k,1}^{\text{cur}})\}$ and $\zeta_{k,1}^{\text{cur}} = \theta_{k,1}^{\text{cur}}$; in the second step update ζ^{cur} according to $K(\zeta^{\text{cur}}, \cdot)$ and the realized ζ^{new} is finally mapped in the third step into $\theta^{\text{new}} = \phi(\zeta^{\text{new}})$. In the event that the starting $\theta^{\text{cur}} \neq 0$, then the first step consists of a deterministic map with $\zeta^{\text{cur}} = (\theta_{k,1}, \psi_k^{-1}\{\theta_{k,2}^{\text{cur}}; r(\theta_{k,1})\})$.

3. APPLICATION TO BAYESIAN MODEL SELECTION

In this section, we present a simple numerical illustration of how the theory developed so far can be usefully applied to Bayesian model selection, when the models under consideration are nested.

We consider a linear model setting, with $n \times (p+1)$ design matrix X, where the first column of X is a column of ones. We will denote by M_k the submodel corresponding to a design matrix obtained by eliminating the last k columns of X ($k = 0, \ldots, p$). The reduced design matrix will be denoted by X_k. Note that X_k is an $n \times (p - k + 1)$ matrix—the subscript k refers to the number of regressors dropped from the full model. The full model can be written as $y \sim \mathcal{N}(X\beta, I\sigma^2)$, and submodel M_k is obtained by setting the last k components of β to zero. With this parameterization, one can use the same likelihood function for all the models under consideration. Within a Bayesian framework, one needs to specify a prior probability that charges all the $p + 1$ events $H_k = \{\beta_0 \neq 0, \ldots, \beta_{p-k} \neq 0, \beta_{p-k+1} = \cdots = \beta_p = 0\}$, $k = 0, \ldots, p$. These events, which correspond to the models M_k, constitute a sequence of nested hyperplanes in the parameter space. We define a prior distribution in a hierarchical way as follows: given H_k and σ^2, $\beta_{k,1} = (\beta_0, \ldots, \beta_{p-k})'$ has a $\mathcal{N}(0, \sigma^2 V_k)$ distribution; given H_k, σ^2 has an $\mathcal{IG}(d/2, a/2)$ distribution; $P(H_k) = \alpha_k$. Introducing the indicator functions

$$R_k(\beta) = \begin{cases} 1 & \text{if} \quad \beta \in H_k, \\ 0 & \text{if} \quad \beta \notin H_k, \end{cases}$$

and the σ-finite measure

$$d\nu(\beta, \sigma^2) = \sum_{k=0}^{p} \eta_{p-k+1}(d\beta_{k,1}) \delta_k(d\beta_{k,2}) \eta(d\sigma^2),$$

the prior density is given by

$$f(\beta, \sigma^2) = \frac{d\pi}{d\nu} = \frac{1}{\Gamma(d/2)} \sum_{k=0}^{p} R_k(\beta) \alpha_k (a/2)^{d/2} (2\pi)^{-(p-k+1)/2} |V_k|^{-1/2} (\sigma^2)^{-(d+p-k+3)/2} \exp\{-Q_k/(2\sigma^2)\},$$

This content downloaded by the authorized user from 192.168.72.233 on Wed, 14 Nov 2012 14:05:31 PM
All use subject to JSTOR Terms and Conditions

with $Q_k = \beta'_{k,1}V_k^{-1}\beta_{k,1}+a$. The posterior density, which can be found using the Bayes theorem, is equal to

$$\begin{aligned} f(\beta,\sigma^2 \mid y) &\propto L(\beta,\sigma^2)f(\beta,\sigma^2) \\ &\propto \sum_{k=0}^{p} R_k(\beta)\left\{\frac{|V_k^*|^{1/2}\alpha_k}{|V_k|^{1/2}(a_k^*/2)^{d^*/2}}\right\}(a_k^*/2)^{d^*/2}(2\pi)^{-(p-k+1)/2}|V_k^*|^{-1/2} \\ &\qquad (\sigma^2)^{-(d^*+p-k+3)/2}\exp\{-Q_k^*/(2\sigma^2)\} \end{aligned} \tag{4}$$

with $Q_k^* = (y - X_k\beta_{k,1})'(y - X_k\beta_{k,1}) + \beta'_{k,1}V_k^{-1}\beta_{k,1} + a = (\beta_{k,1} - m_k^*)'(V_k^*)^{-1}(\beta_{k,1} - m_k^*) + a_k^*$, $V_k^* = (X_k'X_k + V_k^{-1})^{-1}$, $m_k^* = V_k^*X_k'y$, $a_k^* = a + y'y - (m_k^*)'(V_k^*)^{-1}m_k^*$, and $d^* = d+n$.

The posterior probability of submodel M_k is proportional to the integral of the kth summand of (4). In this example, this can be determined in closed form because of the conjugate structure of the prior used in each submodel—essentially completing the normalizing constants:

$$P(M_k \mid y) \propto \frac{|V_k^*|^{1/2}\alpha_k}{|V_k|^{1/2}(a_k^*/2)^{d^*/2}}.$$

Hence, in this setting, we have the opportunity to compare exact posterior model probabilities with Markov chain Monte Carlo estimates.

Let us give some details on how to implement a new computational strategy based on the results of the previous section. An approximation of the posterior distribution can be obtained by sampling σ^2 and β in turn, each from its full conditional distribution. One easily obtains that the full conditional distribution of σ^2 given β when $\beta \in H_k$ (i.e., model M_k is visited) is $\mathcal{IG}((d^* + p - k + 1)/2, Q_k^*/2)$. On the other hand, the full conditional distribution of β is proportional to an expression like (3), i.e.,

$$\sum_{k=0}^{p} f_k(\beta_{k,1})\eta_{p-k+1}(d\beta_{k,1})\delta_k(d\beta_{k,2}), \tag{5}$$

where $f_k(\beta_{k,1}) = \alpha_k(2\pi\sigma^2)^{-(p-k+1)/2}|V_k|^{-1/2}\exp\{-Q_k^*/(2\sigma^2)\}$. Implementing a sampler for this distribution based on the results of the previous section is very easy in practice, once one has coded the generic function mentioned in Remark 3. This function takes as input a list of functions that evaluate the densities f_k and a value of β, and returns the value of the transformed density g of Remark 1 at $\zeta = \phi^{-1}(\beta)$, together with ζ itself. Whenever β does not belong to the full model, i.e., when some of the last β_j are zero, ζ is randomly selected among all those values such that $\phi(\zeta) = \beta$. In view of Theorem 2, this guarantees that, if β^{cur} is the value currently visited by the chain, and one sets $\beta^{\text{new}} = \phi(\zeta^{\text{new}})$, where ζ^{new} is obtained from a Markov kernel having invariant density g and starting at ζ^{cur}, the implied sampler for β has the correct invariant distribution, i.e., the full conditional distribution of β—proportional to (5).

In order to keep our approach as automatic as possible, to generate ζ^{new} from g, we use ARMS (Adaptive Rejection Metropolis Sampling, see Gilks, Best & Tan 1995) along a straight line through ζ^{cur}, having a randomly selected direction. (The complete code for the sampler, written in R with parts in C, is available from the authors upon request.) Exact posterior probabilities, together with their Markov chain Monte Carlo estimates obtained using the sampler described above, are reported in Table 1. Markov chain Monte Carlo estimates are based on 200,000 iterations, after a burn-in of 10,000 iterations.

Let us stress that the seductive and most important distinguishing feature of our basic strategy is that we substantially reduce the expertise of the researcher that is sometimes needed to implement currently used techniques. Consider, for instance, how many pages of articles are often devoted to the explanation of convenient jump proposals in specific models handled with

This content downloaded by the authorized user from 192.168.72.233 on Wed, 14 Nov 2012 14:05:31 PM
All use subject to JSTOR Terms and Conditions

the reversible jump approach. With our method, the user is only asked to: (a) give a guess about where high-density points of each submodel are located and, (b) write functions that evaluate the densities f_k. In particular, there is no need of Jacobians nor of the delicate tuning of proposals. Although step (a) cannot be claimed to be fully automatic, it is certainly more standard and simpler than deriving jump proposals through moment matching conditions or other model-specific heuristics.

TABLE 1: Analytically evaluated posterior probabilities of each model compared to Monte Carlo estimates based on our sampler. Models M_8 and M_9 have posterior probability less than 10^{-7} and were never visited by the sampler.

Model	M_0	M_1	M_2	M_3	M_4	M_5	M_6	M_7
Exact probability	0.0233	0.0741	0.1162	0.1185	0.1278	0.3578	0.1757	0.0066
MCMC estimate	0.0216	0.0763	0.1189	0.1232	0.1284	0.3514	0.1725	0.0076

4. APPLICATION TO REGENERATIVE MARKOV CHAINS

If an ergodic Markov chain has a limiting distribution that charges a particular point (an *atom*) of its state space, then it is possible to identify the *regeneration times* of the chain as the times at which the chain visits the atom. Loosely speaking, at each regeneration time the chain "starts again" probabilistically, and the pieces of the chain, usually referred to as *tours*, between two successive regeneration times are stochastically independent. This independence property can be exploited, in the context of Markov chain Monte Carlo, to obtain accurate estimates of Monte Carlo standard errors that otherwise should have taken into account the dependence structure of the simulated values.

Suppose that a Markov chain Monte Carlo scheme has been implemented for approximating a target distribution

$$\pi(A) \propto \int_A f_0(\theta)\, \eta_K(d\theta)$$

through the realization of an ergodic Markov chain with π as stationary distribution. Let us denote with Q_w the corresponding Markov kernel. Here we exploit the result of Brockwell & Kadane (2002, Th. 2.1), where it is shown how to construct a regenerative Markov chain with the same stationary distribution π by simply relying on a Harris recurrent Markov chain with limiting distribution

$$\pi^*(A) = (1-\lambda)\pi(A) + \lambda I_{\{0\}}(A). \tag{6}$$

The regenerative chain is then exploited to estimate the standard errors of the Markov chain Monte Carlo estimates of features of π. We refer to Brockwell & Kadane (2002) for implementation details. Here we show how the results of our Theorems 1 and 2 can be combined to obtain a new final algorithm, different from that of Brockwell and Kadane, for constructing the basic ingredient, i.e., a Markov chain with (6) as limiting distribution.

One immediately realizes that the mixture π^* in (6) is actually of the form (1), i.e., proportional to a finite measure that can be written as

$$\mu(d\theta) = f_0(\theta)\eta_K(d\theta) + f_K\delta_K(d\theta).$$

Of course the constant f_K and the mixing weight λ are functionally related by

$$\lambda = f_K\left\{\int_{\mathbb{R}_K} f_0(\theta)\, \eta_K(d\theta) + f_K\right\}^{-1}$$

This content downloaded by the authorized user from 192.168.72.233 on Wed, 14 Nov 2012 14:05:31 PM
All use subject to JSTOR Terms and Conditions

so that, fixing either one, the other is automatically determined.

Hence Theorem 1 can be used to define the appropriate measure τ and the function ϕ so that $\mu = \tau\phi^{-1}$. Let $K(\cdot,\cdot)$ denote an appropriate Markov kernel so that τ is invariant for $K(\cdot,\cdot)$. Keeping this notation, we can use Theorem 2 to obtain a Markov kernel $H(\cdot,\cdot)$ so that μ is the invariant distribution.

The only thing to discuss at this point are the guidelines for a working procedure. Notice that we have degrees of freedom in choosing λ or, equivalently, f_K. Also, no mention has been made so far about how to construct an appropriate $K(\cdot,\cdot)$ to approximate τ.

In our limited experience, the following suggestion is likely to be effective. From a pilot run of the original working kernel Q_w used for approximating π, one can get an idea where the center x^* of π is located and how large the total mass $\int_{\mathbb{R}_K} f_0(\theta)\,\eta_K(d\theta)$ is, so that one can set an appropriate mass f_K in order to have a corresponding small weight λ, let us say approximately equal to 10^{-3}. Also, without loss of generality, let us assume that f_0 is such that x^* is located at the origin; otherwise, reparameterize accordingly through a translation. In order to easily get a Markov kernel to simulate from a distribution proportional to τ, one can rely on a Metropolis–Hastings scheme using Q_w as a basis for constructing an appropriate proposal. In fact, having fixed f_K so that λ is approximately equal to 10^{-3}, one can expect that the measure τ is not very different from the original f_0. Geometrically (see also Remark 3) only a relatively small opening around the origin—i.e., a point corresponding to a high density f_0—has to be made to accommodate for the mass f_K. Let us denote with $K(\cdot,\cdot)$ the Markov kernel just derived through this Metropolized scheme.

Now we describe in detail how we implemented this idea with the Dugong dataset used in Brockwell & Kadane (2002), originally taken from Ratkowsky (1983). The data consist of measurements of length (Y) and age (X) of 27 dugongs, and the regression model

$$Y_i \sim \mathcal{N}(\mu_i, \sigma^2), \quad \mu_i = \alpha - \beta\gamma^{X_i},$$

is considered, with unknown parameters $\alpha > 0$, $\beta > 0$, $\gamma \in (0,1)$ and $\sigma^2 > 0$. Prior specification for a Bayesian analysis has been chosen as $\alpha \sim \mathcal{N}(0, 10000)$, $\beta \sim \mathcal{N}(0, 10000)$, $\gamma \sim \mathcal{U}(0,1)$ and $\sigma^2 \sim \mathcal{IG}(0.001, 0.001)$. For approximating posterior quantities relative to $\pi(\alpha,\beta,\gamma,\sigma^2|\text{data})$, Brockwell and Kadane implemented a simple Markov chain Monte Carlo strategy consisting of an hybrid Gibbs–Metropolis with sequential draws from (respectively) the full-conditionals of α, β and σ^2, while for the γ component a simple Metropolis step with independent uniform proposal is used instead of its (unavailable) full-conditional. This corresponds in our previous notation to the working kernel Q_w used to simulate a Markov chain with $\pi(\alpha,\beta,\gamma,\sigma^2|\text{data})$ as the limiting distribution. Similarly, to get the Markov kernel $K(\cdot,\cdot)$ mentioned in Theorem 2, we use the same full conditionals of f_0 (here corresponding to the posterior π) for α, β and σ^2 and an independent uniform, for γ, as Metropolis proposals to simulate from the corresponding full-conditionals of the auxiliary density g corresponding to τ and accept the proposed draws with the probability that ensures that τ is stationary for $K(\cdot,\cdot)$.

TABLE 2: Monte Carlo estimate of the parameters and of their standard error evaluated through the regenerative schemes of Petris and Tardella and Brockwell and Kadane.

	Petris and Tardella		Brockwell and Kadane	
	Posterior Mean	Standard Error	Posterior Mean	Standard Error
$\log\alpha$	0.979318	0.000440	0.978697	0.000440
$\log\beta$	−0.022342	0.000010	−0.022591	0.000010
logit γ	1.895405	0.000854	1.888902	0.000859
σ^2	0.008617	0.000004	0.008615	0.000004

This content downloaded by the authorized user from 192.168.72.233 on Wed, 14 Nov 2012 14:05:31 PM
All use subject to JSTOR Terms and Conditions

In order to compare our strategy with that of Brockwell and Kadane (the results are displayed in Table 2), we calibrated the tuning weight λ of their method to get approximately the same expected tour length. For the Petris and Tardella method, with about 10^7 original iterations, run in about 7.5 hours, 3193 tours were obtained with average tour length of about 3128; the estimated coefficient of variation of tour length divided by total length was 3.31×10^{-4}. For the Brockwell and Kadane method, with 10^7 original iterations, run in about 10 hours, 3422 tours were obtained with average tour length of about 2921; the estimated coefficient of variation of the tour length divided by the total length was 3.98×10^{-4}. The parameter estimates and estimated standard errors are very close for the two methods, as was to be expected, and computing time is reduced by about 25% using our approach. Both procedures were written in R. From a general point of view, an appealing feature of our strategy is that it relies on a geometric understanding of the modified chain and avoids the use of a multinormal distribution to guess the shape of f_0 as suggested by Brockwell and Kadane.

5. DISCUSSION

Our basic theoretical results have been applied to develop a sampler which is substantially different from the reversible jump, for we are able to transform and visualize within a Euclidean structure the problem of simulating from mixtures of two distributions with nested supports. The spirit of our approach is certainly more similar to the Carlin & Chib (1995) approach though their way of transforming the varying dimensional problem into a single space of fixed dimension ends up using a much larger space. The saturated space approach of Brooks, Giudici & Roberts (2003) is based on making each model have the same dimension by adding model specific latent variables. However, in that case one still has a set of distinct Euclidean spaces although of common dimension, and the need for effective jump proposals is not avoided. On the other hand, using our method, one has to deal with a unique Euclidean space, within which the various models are basically represented by different regions (see Figure 3) and, from a geometric standpoint, moving from one model to a different one can be done along a smooth curve without discontinuous jumps.

Since the reversible jump is by far the most widely used transdimensional sampler, we will use it as a benchmark to compare our approach to. For a review and comparison of other alternatives to the reversible jump, the interested reader can refer to Dellaportas, Forster & Ntzoufras (2002), Godsill (2001) and Cappé, Robert & Rydén (2003). We believe that a detailed comparison (based on aspects such as accuracy, CPU time, flexibility and scalability, ease of implementation and debugging costs) between two alternative Markov chain Monte Carlo samplers deserves an especially dedicated paper: here we will limit ourselves to some qualitative remarks on the pros and cons of the method proposed in Section 3 versus the reversible jump.

Certainly in the example presented in Section 3, the numerical results were satisfactory. We are convinced that on the same model and data set, a well-tuned reversible jump sampler may produce equally accurate estimates, even though in our limited trials we experienced some difficulties in obtaining a similar precision of the estimates with our "first-attempt" implementation of the reversible jump with few dedicated tuning and monitoring efforts. Also note that monitoring a reversible jump sampler is still a task far from being approached satisfactorily; see Castelloe & Zimmerman (2002). In the example of Section 3, exact posterior probabilities are available and can be used as a benchmark for the good behaviour of the sampler. In typical real applications, where these are not available, the best way to control whether a particular method is behaving properly is to address the problem with a completely different tool and to verify the agreement of the two results. In this light, our method is a contribution towards the important goal of enlarging the pool of methods available for transdimensional Markov chain Monte Carlo. It follows from Remark 3 that our method lends itself more easily than the reversible jump to an automatic implementation. This is traded off both in terms of scalability/flexibility and computational time expenses. As far as speed of computation is concerned, simulating in the largest model subspace is suboptimal, when compared with the reversible jump. Incidentally, we observe that

This content downloaded by the authorized user from 192.168.72.233 on Wed, 14 Nov 2012 14:05:31 PM
All use subject to JSTOR Terms and Conditions

the approach of Carlin & Chib (1995) is even less appealing in this regard because it requires a sampling space of larger dimension and also the problem-specific (nonautomatic) pseudo-priors. It must also be acknowledged that a sampler like the reversible jump is more flexible than the one proposed here in terms of scalability and flexibility, even though its greater flexibility may sometimes turn out to be a drawback because of the practical difficulties that one can encounter in designing efficient proposals and assessing their effectiveness.

Let us conclude with a few words regarding applications and extensions not covered in the paper. Our method has been conceived and used here to deal with distributions with nested support. This is of course a limitation with respect to other methods which can deal also with the non-nested case, but still the benefits of our approach can be enjoyed in relevant nested situations. Although for illustrative purposes we limited ourselves here to an application to linear models, the method illustrated in Section 3 can be applied to many other nested models with only minor modifications. Actually, the case of stationary autoregressive models, with unknown order, has already been considered in Petris & Tardella (2000). The same problem has also been considered, using the reversible jump, in Godsill & Troughton (1998) and Brooks, Giudici & Roberts (2003) in the multivariate case without restriction to the stationary models. This problem has recently been reanalyzed in order to improve the efficiency of the reversible jump scheme in an article especially devoted to it (Brooks & Ehlers 2003).

It is worth noting that many other families of models can be given a nested structure by selecting an appropriate parameterization. This is the case, for example, of mixtures of Normal distributions with an unknown number of components. If π_j is the weight of component j in the mixture $j = 1, \ldots, K$, then identifiability can be enforced through the constraint $\pi_1 \geq \cdots \geq \pi_K$. But then, setting $\pi_K = \pi_{K-1} = \cdots = \pi_{K-k+1} = 0$, $\pi_{K-k} > 0$, one effectively has a mixture with $h = K-k$ components. Taking in turn $k = 0, \ldots, K-1$, one obtains a set of nested models to which the approach of Section 3 can be applied. Successful results about Bayesian estimation of mixtures of normal distributions with an unknown number of components have been obtained by the authors and are the object of a forthcoming paper. Also the treatment of mixtures of binomial distributions, based on the canonical moment representation used in Tardella (2002), is currently under investigation.

Finally, we hint at the possibility that our method may be extended to non-nested situations similar to the spirit of the reversible jump. An anonymous referee remarked that in typical applications of the reversible jump, even when the entire family of models considered is non-nested, jumps occur between pairs of nested models. This suggests that from the current model, one can propose at random a specific smaller or larger model, and then move in the union of the two—the current and the proposed one—using our method. The move has then to be accepted or rejected with a probability that ensures reversibility of the resulting chain. Our method, transforming the union of the two models into a unique Euclidean space, opens the door to the use of generic samplers, e.g., ARMS along a randomly selected direction, that may represent an effective way out of the sometimes difficult problem of choosing the jump proposals in the reversible jump.

APPENDIX

Proof of Theorem 1. We will make use of the fact that the Jacobian of the transformation ϕ is one since ψ_k preserves the Lebesgue measure. Consider two Borel sets A_1 and A_2 in $\mathbb{R}^h$ and $\mathbb{R}^k$ respectively. If $0 \notin A_2$, then

$$\begin{aligned}
\tau\phi^{-1}(A_1 \times A_2) &= \tau(\{\zeta : \phi(\zeta) \in A_1 \times A_2\}) = \int_{\{\zeta:\phi(\zeta)\in A_1\times A_2\}} g(\zeta)\,\eta_K(d\zeta) \\
&= \int_{\{\zeta:\phi(\zeta)\in A_1\times A_2\}} f_0\{\phi(\zeta)\}\,\eta_K(d\zeta) \\
&= \int_{A_1\times A_2} f_0(\theta)\,\eta_K(d\theta) = \mu(A_1 \times A_2).
\end{aligned}$$

This content downloaded by the authorized user from 192.168.72.233 on Wed, 14 Nov 2012 14:05:31 PM
All use subject to JSTOR Terms and Conditions

On the other hand, if $A_2 = \{0\}$, then

$$\begin{aligned}
\tau\phi^{-1}(A_1 \times A_2) &= \tau(\{\zeta : \phi(\zeta) \in A_1 \times A_2\}) \\
&= \int_{\{\zeta:\phi(\zeta)\in A_1\times A_2\}} g(\zeta)\,\eta_K(d\zeta) \\
&= \int_{\{\zeta:\zeta_{k,1}\in A_1,\zeta_{k,2}\in B_k[r(\zeta_{k,1})]\}} f_k(\zeta_{k,1})c(\zeta_{k,1})\,\eta_K(d\zeta) \\
&= \int_{A_1} f_k(\zeta_{k,1})c(\zeta_{k,1})\left(\int_{B_k\{r(\zeta_{k,1})\}} \eta_k(d\zeta_{k,2})\right)\eta_h(d\zeta_{k,1}) \\
&= \int_{A_1} f_k(\zeta_{k,1})c(\zeta_{k,1})\eta_k[B_k\{r(\zeta_{k,1})\}]\,\eta_h(d\theta_{k,1}) \\
&= \int_{A_1} f_k(\theta_{k,1})\,\eta_h(d\theta_{k,1}) = \mu(A_1 \times A_2).
\end{aligned}$$

The two equalities imply that for any Borel set A in $\mathbb{R}^K$, $\tau\phi^{-1}(A) = \mu(A)$. Continuity of g, under the specified additional assumptions, is straightforward.

Proof of Theorem 2. For every $B \in \mathcal{S}_Z$, one has

$$\int_\Theta \mu(d\theta)J(\theta; B) = \int_Z \tau(d\zeta)J\{\phi(\zeta); B\} = \tau(B).$$

Therefore,

$$\begin{aligned}
\int_\Theta \mu(d\theta)H(\theta; A) &= \int_\Theta \mu(d\theta)\int_Z J(\theta, d\zeta)K(\zeta, \phi^{-1}A) = \int_Z \tau(d\zeta)K(\zeta; \phi^{-1}A) \\
&= \tau(\phi^{-1}A) = \mu(A).
\end{aligned}$$

□

ACKNOWLEDGEMENTS

The authors thank the Editor, the Associate Editor and the referees for the helpful comments and suggestions, which improved the presentation of the paper. We are grateful to Laurie Meaux for carefully reviewing the manuscript.

REFERENCES

F. Al-Awhadhi, M. A. Hurn & C. Jennison (2001). *Improving the Acceptance Rate of Reversible Jump MCMC Proposals*. Technical Report, University of Bath, Bath, England.

A. E. Brockwell & J. B. Kadane (2002). *Identification of Regeneration Times in MCMC Simulation, With Application to Adaptive Schemes*. Technical Report, Dept. of Statistics, Carnegie Mellon University, Pittsburgh.

S. P. Brooks & R. S. Ehlers (2003). *Efficient Construction of Reversible Jump MCMC Proposals for Autoregressive Time Series Models*. Technical Report.

S. P. Brooks, P. Giudici & G. O. Roberts (2003). Efficient construction of reversible jump MCMC proposal distributions (with discussion). *Journal of the Royal Statistical Society Series B*, 65, 3–55.

O. Cappé, C. Robert & T. Rydén (2003). Reversible jump, birth-and-death and more general continuous time Markov chain Monte Carlo samplers. *Journal of the Royal Statistical Society Series B*, 65, 679–700.

B. P. Carlin & S. Chib (1995). Bayesian model choice via Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society Series B*, 57, 473–484.

This content downloaded by the authorized user from 192.168.72.233 on Wed, 14 Nov 2012 14:05:31 PM
All use subject to JSTOR Terms and Conditions

J. M. Castelloe & D. M. Zimmerman (2002). *Convergence Assessment for Reversible Jump MCMC Samplers*. Technical Report, Dept. of Statistics and Actuarial Science, University of Iowa, Ames.

P. Dellaportas, J. J. Forster & I. Ntzoufras (2002). On Bayesian model and variable selection using MCMC. *Statistics and Computing*, 12, 27–36.

W. R. Gilks, N. G. Best & K. K. C. Tan (1995). Adaptive rejection Metropolis sampling within Gibbs sampling. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 44, 455–472.

W. R. Gilks, R. M. Neal, N. G. Best & K. K. C. Tan (1997). Corrigendum: Adaptive rejection Metropolis sampling. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 46, 541–542.

S. J. Godsill (2001). On the relationship between Markov chain Monte Carlo methods for model uncertainty. *Journal of Computational and Graphical Statistics*, 10, 230–248.

S. J. Godsill & P. Troughton (1998). A reversible jump sampler for autoregressive time series. In *Proceedings of the 1998 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'98): Washington State Conventon and Trade Center, Seattle, May 12–15, 1998*, IEEE Service Center, Piscataway, NJ, vol. 4: Statistical Signal and Array Processing, Underwater Acoustical Signal Processing, pp. 2257–2260.

P. J. Green (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination, *Biometrika*, 82, 711–732.

P. J. Green (2003). Trans-dimensional MCMC. In *Highly Structured Stochastic Systems* (P. J. Green, N. L. Hjort & S. Richardson, eds.), Oxford Statistical Science Series 27, Oxford University Press.

R. Ihaka & R. Gentleman (1996). R: a language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, 5, 299–314.

G. Petris & L. Tardella (2000). *A New Strategy for Simulating From Mixture Distributions With Applications to Bayesian Model Selection*. Technical Report, Dipartimento di Statistica, Probabilità e Statistiche Applicate, Università di Roma "La Sapienza", Rome, Italy.

D. A. Ratkowsky (1983). *Nonlinear Regression Modeling: A Unified Practical Approach*. Marcel Dekker, New York.

R. Rotondi (2002). On the influence of the proposal distributions on a reversible jump MCMC algorithm applied to the detection of multiple change-points, *Computational Statistics and Data Analysis*, 40, 633–653.

M. A. Stephens (2000). Bayesian analysis of mixture models with an unknown number of components—an alternative to reversible jump methods. *The Annals of Statistics*, 28, 40–74.

L. Tardella (2000). A new Bayesian method for nonparametric capture-recapture models in presence of heterogeneity. *Biometrika*, 89, 807–817.

Received 13 January 2003
Accepted 7 October 2003

Giovanni PETRIS: GPetris@uark.edu
Department of Mathematical Sciences
University of Arkansas, Fayetteville, AR 72701, USA

Luca TARDELLA: Luca.Tardella@uniroma1.it
Dipartimento di Statistica, Probabilità e Statistiche Applicate
Università di Roma "La Sapienza"
Piazzale Aldo Moro, 5, IT-00185 Roma, Italy

This content downloaded by the authorized user from 192.168.72.233 on Wed, 14 Nov 2012 14:05:31 PM
All use subject to JSTOR Terms and Conditions