Statistica Sinica **18**(2008), 1641-1651

AN EXTENSION OF PESKUN AND TIERNEY ORDERINGS TO CONTINUOUS TIME MARKOV CHAINS

Fabrizio Leisen and Antonietta Mira

University of Insubria

Abstract: Peskun ordering is a partial ordering defined on the space of transition matrices of discrete time Markov chains. If the Markov chains are reversible with respect to a common stationary distribution π, Peskun ordering implies an ordering on the asymptotic variances of the resulting Markov chain Monte Carlo estimators of integrals with respect to π. Peskun ordering is also relevant in the framework of time-invariant estimating equations in that it provides a necessary condition for ordering the asymptotic variances of the resulting estimators. Tierney ordering extends Peskun ordering from finite to general state spaces. In this paper Peskun and Tierney orderings are extended from discrete time to continuous time Markov chains.

Key words and phrases: Asymptotic variance, covariance ordering, efficiency ordering, MCMC, time-invariance estimating equations.

1. Introduction

The class of Markov chains (MCs) that are stationary with respect to a specified distribution, π, plays an important role in two separate but connected fields, namely Markov chain Monte Carlo methods (MCMC), Hastings (1970), and time-invariance estimating equations (TIEE), Baddeley (2000).

In MCMC we are interested in estimating the expected value, $E_\pi f$, of a function f with respect to a distribution π. If such integral cannot be computed analytically, either because the state space is too large or because π and/or f are too complicated, we construct a Markov chain that has π as its unique stationary and limiting distribution. The MC is run for n time-steps to produce a simulated path: $x_1, x_2, \cdots x_n$, possibly after a burn-in period that allows the MC to forget its initial distribution and to reach the stationary regime. We then estimate $\mu = E_\pi f$ by $\hat{\mu}_n = (1/n)\sum_{i=1}^{n} f(x_i)$. Under regularity conditions, Tierney (1994), the Strong Law of Large Numbers and the Central Limit Theorem ensure that $\hat{\mu}$ is asymptotically unbiased and normally distributed.

Time-invariant estimating equations is a general framework to construct estimators for generic models. Suppose we have a model indexed by a parameter,

π_θ, and we are interested in estimating θ. We construct a MC that has the model of interest as its stationary distribution. An unbiased estimating equation is obtained by equating to zero the generator of the Markov chain applied to some function, S, defined on the sample space and evaluated at the data, x. A natural way to evaluate the performance of time-invariant estimators is the Godambe-Heyde asymptotic variance, Godambe and Kale (1991).

Both in the MCMC and the TIEE framework there are some degrees of freedom on how to choose the Markov chain since, given the distribution or the model of interest, there are many MCs that are stationary with respect to it. In the MCMC context this raises the following question: given two Markov chains, Q_1 and Q_2, both ergodic with respect to π, which one produces estimators of $E_\pi f$ with smaller asymptotic variance? The corresponding question in the TIEE framework is the following: given two Markov chains stationary with respect to π_θ, which one produces time-invariant estimators of θ with smaller Godambe-Heyde asymptotic variance?

Peskun (1973) first addressed the question in the MCMC context by proposing a partial ordering on the space of discrete time Markov chains defined on finite state spaces. The ordering was later extended by Tierney (1998) to general state space MCs, but the discrete time assumption was retained. In their papers, Peskun and Tierney demonstrate that their respective orderings imply an ordering on the resulting MCMC estimators in terms of asymptotic variances; i.e., in terms of their efficiency. A related partial ordering, the covariance ordering, was later introduced by Mira and Geyer (1999). While Peskun ordering gives a sufficient condition for efficiency ordering, covariance ordering is equivalent to efficiency ordering.

The related question in the TIEE framework was first addressed by Mira and Baddeley (2001). The authors show that Peskun ordering gives a necessary but not sufficient condition for Godambe-Heyde ordering.

Both in the MCMC and in the TIEE framework one often has to deal with continuous time Markov chains. In particular, in the MCMC framework, there has been success in constructing an efficient proposal distribution for the Metropolis-Hastings algorithm by using the Euler discretizations of the transition probabilities of a Langevin diffusion process that has π as its stationary distribution. The seminal paper along these lines appeared in the physics literature (Doll, Rossky and Friedman (1978)), and the idea was only later brought into the mainstream statistical literature in a discussion by Besag (1994). Theoretical convergence properties, in terms of speed of convergence to stationarity, of these type of MCMC algorithms have been extensively studied, see for example Roberts and Tweedie (1996a,b) and Stramer and Tweedie (1999a,b). On the other hand, to our knowledge, there is still no theoretical discussion of the properties of these diffusion Metropolis-Hastings-type algorithms in terms of the asymptotic variance of the resulting estimators; i.e., in terms of efficiency orderings. As for the TIEE framework, there are state-space models that naturally appear as stationary distributions of continuous time processes. Just to give one example, a Gibbs point processes can be seen as the stationary distribution of a spatial birth and death process. For more examples refer to Baddeley (2000), Mira and Baddeley (2001), Mira and Baddeley (2007). In order to study the performance of the resulting time-invariant estimators in this context, the extension of Peskun ordering proposed in the present paper could be highly beneficial.

The aim of this paper is to extend Peskun ordering to the case of continuous time Markov chains. Despite the fact that Peskun ordering can be relevant in two different general frameworks, as we noted above, we mainly focus on the original MCMC context where Peskun ordering was introduced. In Section 2 we review the Peskun (2.1) and Tierney (2.2) orderings defined in discrete time for finite and general state space Markov chains, respectively. We then show how continuous time Markov chains could be exploited for MCMC purposes (Section 3). The core of the paper is in Section 4 and 5, where new orderings are defined for continuous time Markov chains in finite and general state spaces, respectively. An example (Section 6) and some remarks on possible future research directions conclude the paper.

2. Ordering of Markov Chains in the MCMC Setting

In this section we review the definitions of orderings defined on the space of Markov chains that are stationary with respect to a common distribution of interest. These orderings are relevant for MCMC purposes in that they link the performance of the resulting MCMC estimators to some characteristic of the transition kernel used to update the underlying MC.

In particular we first give a definition of the efficiency ordering (Mira (2001)) that holds both for finite and general state spaces. We then review the Peskun (1973) and Tierney (1998) orderings and the results that connect these orderings to the efficiency of the resulting MCMC estimators.

2.1. Efficiency ordering

We begin by giving some definitions and setting up the notation. Let $Q = \{q_{ij}\}_{ij \in E}$ be a time-invariant transition matrix; i.e.,

$$q_{ij} = P(X_{t+1} = j | X_t = i), \qquad \forall t,$$

where the MC takes values on a finite state space E, and X_t denote the position of the MC at time t. We identify Markov chains with the corresponding transition matrices.

Let $\mathcal{S}$ be the class of Markov chains stationary with respect to some given distribution of interest, say π (i.e., $Q \in \mathcal{S}$ if $\pi Q = \pi$), $\mathcal{R}$ be the subset of the reversible ones, and $L^2(\pi)$ be space of all functions that have a finite variance with respect to π. Let $v(f, Q)$ be the limit, as n tends to infinity, of n times the variance of the MCMC estimator, $\hat{\mu}_n$, computed on a π-stationary chain updated using the transition matrix Q:

$$v(f, Q) = \lim_{n\to\infty} n\mathrm{Var}_\pi(\hat{\mu}_n) = \mathrm{Var}_\pi(f(X))\Big[1 + 2\sum_{j=1}^{\infty} \rho_j\Big],$$

where $\rho_j = \mathrm{Cor}_\pi[f(X_0), f(X_j)]$ is the lag-j autocorrelation along the simulated π-stationary chain.

Definition 1. Let $Q_1, Q_2 \in \mathcal{S}$. Q_1 is uniformly more efficient than Q_2, $Q_1 \succeq_E Q_2$, if $v(f, Q_1) \leq v(f, Q_2)$ for all $f \in L^2(\pi)$.

2.2. Peskun ordering for discrete time finite state space Markov chains

Definition 2. Given two Markov chains $Q_1, Q_2 \in \mathcal{S}$, $Q_1 = \{q_{(1)ij}\}_{i,j\in E}$ and $Q_2 = \{q_{(2)ij}\}_{i,j\in E}$, we say that Q_1 is better than Q_2 in the Peskun sense, and write $Q_1 \succeq_P Q_2$, if

$$q_{(1)ij} \geq q_{(2)ij}, \qquad \forall i \neq j.$$

Peskun ordering is also known as off the diagonal ordering because, in order that $Q_1 \succeq_P Q_2$, each of the off-diagonal elements of Q_1 has to be greater than or equal to the corresponding off-diagonal elements in Q_2. This means that Q_1 has a higher probability of moving around in the state space than Q_2, and therefore the corresponding Markov chain explores the space in a more efficient way (better mixing). Thus, we expect that the resulting MCMC estimates will be more precise than the ones obtained by averaging along a Markov chain generated via Q_2. This intuition is stated more rigorously in the Peskun (1973) theorem.

Theorem 1. *Given two Markov chains $Q_1, Q_2 \in \mathcal{R}$, $Q_1 \succeq_P Q_2 \Rightarrow Q_1 \succeq_E Q_2$.*

The first use of this ordering appears in Peskun (1973), where the author shows that the Metropolis-Hastings algorithm, Hastings (1970), the main algorithm used in MCMC, dominates a class of competitors reversible with respect to some π, all with the same propose/accept updating structure, and with symmetric acceptance probability (see also Baddeley (2000)).

2.3. Tierney ordering for discrete time general state space Markov chains

We identify Markov chains with the corresponding transition kernels: $Q(x, A) = \Pr(X_n \in A | X_{n-1} = x)$ for every set A.

Definition 3. If Q_1 and Q_2 are transition kernels on a measurable space with stationary distribution π, then Q_1 dominates Q_2 in the Tierney ordering, $Q_1 \succeq_T Q_2$, if for $\pi-$almost all x in the state space we have $Q_1(x, B\backslash\{x\}) \geq Q_2(x, B\backslash\{x\})$ for all measurable B.

The theorem of Tierney (1998) extends Theorem 2.1.1 by Peskun (1973).

Theorem 2. *Given two Markov chains* $Q_1, Q_2 \in \mathcal{R}$, $Q_1 \succeq_T Q_2 \Rightarrow Q_1 \succeq_E Q_2$.

3. Continuous Time Markov Chains for MCMC Simulations

Let $\{X(t)\}_{t\in\Re^+}$ be a continuous time MC (CTMC) taking values on a finite state space E. Let $G = \{g_{ij}\}_{i,j\in E}$ be the generator of the MC. G is a matrix with row sums equal to zero, having negative entries along the main diagonal and positive entries otherwise. Assume that the MC is reversible, this condition, usually checked on the MC transition matrix, can also be checked on the generator by requiring that

$$\pi_i g_{ij} = \pi_j g_{ji} \qquad \forall i, j \in E.$$

Let I be the identity matrix, $c = \sup_i |g_{ii}|$ and $\nu \geq c$, then $P_\nu = I + (1/\nu)G$ is a stochastic matrix. Note that if G is reversible with respect to π, then so is $P_\nu, \forall\nu$. Such CTMC (based on P_ν) could be used for MCMC purposes in the following way. Assume, without loss of generality, that f has zero mean and finite variance under π, $f \in L_0^2(\pi)$, and furthermore assume that f belongs to the range of the generator, $R(G)$, of the CTMC. Suppose we are interested in estimating $\mu = \int f(x)\pi(dx)$. Construct a CTMC $\{X(t)\}_{t\in\Re^+}$ ergodic with respect to π, fix $t > 0$, and take

$$\hat{\mu}_{nt} = \frac{1}{\sqrt{n}} \int_0^{nt} f(X(s))ds$$

to be the MCMC estimator. By Theorem 2.1 in Bhattacharya (1982), $\hat{\mu}_{nt}$ converges weakly to the Wiener measure with zero drift and variance parameter

$$v(f, G) = -2 < f, g >= -2 \int f(x)g(x)\pi(dx) \geq 0,$$

where g is some element in the domain of the generator, $D(G)$ (i.e., the limit

$$\lim_{t\to 0} \frac{E[g(X_t)|X_0 = x]}{t}$$

exists), and $Gg = f$.

In Proposition 2.4, Bhattacharya (1982), proves that $v(f,G) > 0$ for all non constant (a.s. π) bounded f in the range of G provided, for some $t > 0$ and all x, the transition probability $P(t,x,dy)$ and the invariant measure π are mutually absolutely continuous. If, however, G is reversible, then $v(f,G) > 0$ for all nonzero $f \in R(G)$ without the additional assumption of boundedness and mutual absolute continuity.

4. Peskun Ordering for Continuous Time Finite State Space MCs

We now introduce the generalized version of Peskun ordering for CTMC.

Definition 4. Let $G_1 = \{g_{(1)ij}\}$ and $G_2 = \{g_{(2)ij}\}$ be two CTMCs, both stationary with respect to π. We say that G_1 dominates G_2 in the Peskun sense, and write $G_1 \succeq_P G_2$, if $g_{(1)ij} \geq g_{(2)ij}$, $\forall i \neq j$.

The following theorem mimics the one in Tierney (1998).

Theorem 3. *If $G_1 \succeq_P G_2$ and if the corresponding CTMCs are reversible, then $G_2 - G_1$ is a positive operator.*

Proof. Let $c_1 = \sup_i |g_{(1)ii}|$, $\quad c_2 = \sup_i |g_{(2)ii}|$ and $\nu \geq \max(c_1, c_2)$. Define

$$P_1(\nu) = I + \frac{1}{\nu} G_1 \text{ and } P_2(\nu) = I + \frac{1}{\nu} G_2.$$

We thus have that $G_1 = \nu(P_1(\nu) - I)$ and $G_2 = \nu(P_2(\nu) - I)$. If $G_1 \succeq_P G_2$ it follows that $P_1(\nu) \succeq_P P_2(\nu)$. By Lemma 3 in Tierney (1998), it then follows that $P_2(\nu) - P_1(\nu)$ is a positive operator, and $G_2 - G_1 = \nu(P_2(\nu) - P_1(\nu))$.

For $f, g \in L^2(\pi)$, write $< f, g >= \int f(x)g(x)\pi(dx)$, and recall that an operator P on $L^2(\pi)$ is said to be self-adjoint, if for all $f, g \in L^2(\pi)$, $< Pf, g >=< f, Pg >$.

Theorem 4. *If $G_1 \succeq_P G_2$ and if the corresponding CTMCs are reversible, then*

$$v(f, G_1) \leq v(f, G_2), \qquad \forall f \in R(G_1) \cap R(G_2),$$

where $v(f, G_1)$ and $v(f, G_2)$ are the asymptotic variances of estimators $\hat{\mu}_n$ obtained by simulating the CTMCs that have G_1 and G_2, respectively, as generators.

Proof. From Bhattacharya (1982), for all functions $f \in R(G_1) \cap R(G_2)$, we have

$$v(f, G_i) = -2 < f, g_i >, \qquad i = 1, 2, \tag{4.1}$$

where $g_i \in D(G_i)$ and is such that

$$G_i g_i = f, \qquad i = 1, 2. \tag{4.2}$$

It follows that $v(f,G_i) = -2 < G_i g_i, g_i >, i = 1,2$. Define $H_\beta = G_1 + \beta(G_2 - G_1)$ and $g_\lambda = g_1 + \lambda(g_2 - g_1)$, where $0 \leq \beta \leq 1$ and $0 \leq \lambda \leq 1$. Let $h_\lambda(\beta) = -2 < H_\beta g_\lambda, g_\lambda >$. Then $h_\lambda^{'}(\beta) = -2 < (G_2 - G_1)g_\lambda, g_\lambda >$ and the derivative is non-positive for every λ because $G_2 - G_1$ is a positive operator. It follows that $h_\lambda(\beta)$ is a decreasing function in β for any λ. We thus have that $h_\lambda(0) \geq h_\lambda(1), \forall \lambda \in [0,1]$. Taking $\lambda = 0$ we get $h_0(0) = v(f,G_1)$ and $h_0(1) = -2 < G_2 g_1, g_1 >$. We thus have

$$\begin{aligned}
v(f,G_1) &\geq -2 < G_2 g_1, g_1 > \\
&= -2 < G_2(g_1 - g_2 + g_2), (g_1 - g_2 + g_2) > \\
&= -2 < G_2(g_1 - g_2), (g_1 - g_2) > -2 < G_2 g_1, g_2 > +2 < G_2 g_2, g_2 > \\
&\quad -2 < G_2 g_2, g_1 > +2 < G_2 g_2, g_2 > -2 < G_2 g_2, g_2 > \\
&= -2 < G_2(g_1 - g_2), (g_1 - g_2) > -2 < G_1 g_1, g_1 > +2 < G_2 g_2, g_2 > \\
&\quad -2 < G_1 g_1, g_1 > +2 < G_2 g_2, g_2 > -2 < G_2 g_2, g_2 > \\
&= -2 < G_2(g_1 - g_2), (g_1 - g_2) > +v(f,G_1) - v(f,G_2) \\
&\quad +v(f,G_1) - v(f,G_2) + v(f,G_2). \qquad (4.3)
\end{aligned}$$

The third equality in (4.3) follows from the fact that G is a self-adjoint operator and from (4.2). The last equality in (4.3) follows from (4.1). As a result we obtain

$$v(f,G_2) - v(f,G_1) \geq -2 < G_2(g_1 - g_2), (g_1 - g_2) > \geq 0.$$

This results can easily be extended to countable state spaces provided the Markov chain is uniformizable (see Kijima (1997, p.195)), which guarantees the existence of P_ν.

5. Tierney Ordering for Continuous Time General State Space Markov Chains

Let E be a general state space and $\mathcal{E}$ the associated sigma-algebra. Consider an homogeneous continuous time MC $\{X_t\}_{t \in \Re^+}$ taking values on E, with transition kernel $P(t,x,dy)$ and generator $G : D(G) \to R(G)$. If the generator of the process can be written as an operator:

$$Gf(x) = \int f(y) Q(x,dy), \qquad (5.1)$$

where the kernel Q is defined in terms of the transition kernel P as

$$Q(x,dy) = \frac{\partial}{\partial t} P(t,x,dy) \mid_{t=0},$$

we can extend Tierney ordering to CTMCs in the following way. Let G_1 and G_2 be the generators of two CTMCs with kernels Q_1 and Q_2, respectively, both stationary with respect to a common distribution π taking values on E. Assume $\sup_x Q_i(x, E \setminus \{x\}) < \infty$, $i = 1, 2$.

Definition 5. G_1 dominates G_2 in the Tierney ordering, $G_1 \succeq_P G_2$, if $Q_1(x, A \setminus \{x\}) \geq Q_2(x, A \setminus \{x\})$, $\forall A \in \mathcal{E}$.

In this setting Theorems 3 and 4 can be easily extended to general state spaces.

Theorem 5. *If $G_1 \succeq_P G_2$ and if the corresponding kernels are reversible with respect to a common distribution π on $\mathcal{E}$, then $G_2 - G_1$ is a positive operator.*

Proof. Let $c_1 = \sup_x Q_1(x, E \setminus \{x\}) < \infty$, $c_2 = \sup_x Q_2(x, E \setminus \{x\}) < \infty$ and $\nu \geq \max(c_1, c_2)$. Then $P_{1\nu}(x, dy) = \delta_x(dy) + (1/\nu)Q_1(x, dy)$ and $P_{2\nu}(x, dy) = \delta_x(dy) + (1/\nu)Q_2(x, dy)$ are transition kernel of CTMCs, reversible with respect to π, and such that $P_{1\nu} \succeq_P P_{2\nu}$. By Lemma 3 in Tierney (1998), it then follows that $P_{2\nu} - P_{1\nu} = (1/\nu)(Q_2 - Q_1)$ is a positive operator and, as a consequence, $Q_2 - Q_1$ is also a positive operator.

Theorem 6. *If $G_1 \succeq_P G_2$ and the corresponding kernels are reversible with respect to a common distribution π on $\mathcal{E}$, then*

$$v(f, G_1) \leq v(f, G_2), \qquad \forall f \in R(G_1) \cap R(G_2),$$

where $v(f, G_1)$ and $v(f, G_2)$ are the asymptotic variances of MCMC estimators obtained by simulating the processes having generators G_1 and G_2 respectively.

The proof is the same as the one of Theorem 4.

Unfortunately not all CTMCs admit the representation in (5.1), for example the generator of the multidimensional Brownian motion is the Laplacian. On the other hand, the generator of a Markov jump process is such that

$$Gf(x) = \lambda(x) \int [f(y) - f(x)]\mu(x, dy), \tag{5.2}$$

where $\mu(x, dy)$ is the transition kernel and $\lambda(x)$ is a non-negative bounded real function on E. Let $\overline{\lambda} = \sup_{x \in E} \lambda(x)$ (assume $\overline{\lambda} > 0$). We can then rewrite (5.2) as

$$Gf(x) = \overline{\lambda} \int [f(y) - f(x)]\mu'(x, dy),$$

where $\mu'(x, dy) = [1 - \lambda(x)/\overline{\lambda}]\delta_x(dy) + (\lambda(x)/\overline{\lambda})\mu(x, dy)$. Let $Q(x, dy) = \overline{\lambda}[\mu'(x, dy) - \delta_x(dy)]$, so that, as required, $Gf(x) = \int f(y)Q(x, dy)$.

6. An Example of Peskun Ordering

Let $\{X(t)\}_{t\in\Re^+}$ be a birth and death process taking values in $E = \{0,\ldots,N\}$, and having infinitesimal generator

$$G_1 = \begin{pmatrix} -\lambda_0 & \lambda_0 & 0 & 0 & \cdots & 0 \\ \mu_1 & -\lambda_1-\mu_1 & \lambda_1 & 0 & \cdots & 0 \\ 0 & \mu_2 & -\lambda_2-\mu_2 & \lambda_2 & \cdots & 0 \\ \vdots & & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \mu_{N-1} & -\lambda_{N-1}-\mu_{N-1} & \lambda_{N-1} \\ 0 & \cdots & 0 & 0 & \mu_N & -\mu_N \end{pmatrix},$$

where $\lambda_i, \mu_i > 0, \forall i$ are the birth and death rates respectively. The invariant distribution of the process is, up to a proportionality constant, $\pi_0 = 1$ and

$$\pi_i = \frac{\lambda_0\lambda_1\cdots\lambda_{i-1}}{\mu_1\mu_2\cdots\mu_i}, \qquad i = 1,\ldots,N.$$

Furthermore the process is reversible with respect to this invariant distribution.

Without loss of generality take $N = 4$ and fix the birth rates $\lambda_0 = 1, \lambda_1 = 2, \lambda_2 = 3, \lambda_3 = 2$ and the death rates $\mu_1 = 1, \mu_2 = 3, \mu_3 = 2, \mu_4 = 1$. The invariant distribution is, up to a proportionality constant, $\pi_0 = 1, \pi_1 = 1, \pi_2 = 2/3, \pi_3 = 1, \pi_4 = 2$.

Consider now a different birth and death process, G_2 with rates $\lambda_1 = 2, \lambda_2 = 6, \lambda_3 = 15, \lambda_4 = 4$ and $\mu_1 = 2, \mu_2 = 9, \mu_3 = 10, \mu_4 = 2$. It is easy to check that G_2 is also reversible with respect to the same stationary distribution. Furthermore we have $G_2 \succeq_P G_1$.

More general of, if we take any other birth and death process, $G_{(k)}$, where $k = (k_1, k_2, \ldots, k_N), \forall k > 0$, with rates given by $\lambda_{(k)i-1} = \lambda_{i-1}k_i$ and $\mu_{(k)i} = \mu_i k_i, i = 1,\ldots,N$, then this process is also reversible with respect to π and $G_{(k)} \succeq_P G_1$ if $k_i > 1, \forall i$.

6.1. A possible use of Peskun ordering for MCMC purposes

Suppose we are interested in sampling from a distribution defined on E, $\pi \propto (\pi_0,\ldots,\pi_N)$. For the sake of simplicity, assume that $\pi_0 = 1$. The idea is to use the birth and death process defined in Section 6 to sample from π. We first construct a birth and death process invariant with respect to π. This is done by letting the birth and death rates $\{\lambda_0, \lambda_1, \ldots, \lambda_{N-1}, \mu_1, \ldots, \mu_N\}$ be $\lambda_0/\mu_1 = \pi_1$ and $\lambda_{i-1}/\mu_i = \pi_i/\pi_{i-1}, i = 2,\ldots,N$. If $P(t) = \{p_{ij}\}_{i,j\in E}$ is the t-step transition matrix of the process, e.g. $p_{ij}(t) = P(X(t) = j|X(0) = i)$, we can use $P(\Delta)$ to sample from π, where $\Delta > 0$, is a predefined time-step. The intuition is that, for Δ sufficiently small, the asymptotic variance of the MCMC estimator will be close to $v(f,G)/\Delta$ (Mira and Leisen (2007)). We know, from Section 6, that the choice of the birth and death process is not unique, so by choosing a process which is better in the Peskun ordering, the resulting MCMC estimator should have a smaller asymptotic variance. Note that $P(\Delta)$ is reversible with respect to π, since

$$P(t) = \exp\{Gt\} = \sum_{i=0}^{\infty} \frac{G^i t^i}{i!}. \tag{6.1}$$

To calculate $P(\Delta)$, for sufficiently small Δ, we use a first order approximation of (6.1); i.e., we take the first two terms in the infinite sum on the right hand side to get $P(\Delta) = I + \Delta G + o(\Delta)$. Then $p_{i,i+1}(\Delta) = \lambda_i \Delta + o(\Delta)$ as $\Delta \to 0$, for $i \geq 0$; $p_{i,i-1}(\Delta) = \mu_i \Delta + o(\Delta)$ as $\Delta \to 0$, for $i \geq 1$; $p_{i,i}(\Delta) = 1 - (\lambda_i + \mu_i)\Delta + o(\Delta)$ as $\Delta \to 0$, for $i \geq 0$. In the limit, when $\Delta = 0$, we have $p_{ij}(0) = \delta_{ij}$. We note that also this first order approximation to $P(\Delta)$ (and any other higher order approximation) is reversible with respect to π.

7. Remarks

We plan to investigate the extension of the covariance ordering from discrete to continuous time Markov chains in further research. At this stage not much can be said about the performance, in terms of efficiency, of MCMC algorithms like Langevin diffusions, because we have been able to extend Tierney ordering from discrete to continuous time only under the assumption that the generator of the process can be written as an integral operator involving the time-derivative of the transition kernel. This is an open research line.

References

Baddeley, A. J. (2000). Time-invariance estimating equations. *Bernoulli* **6**, 783-808.

Besag, J. E. (1994). Comments on “Representations of knowledge in complex systems” by U. Grenander and M. I. Miller. *J. Roy. Statist. Soc. Ser. B* **56**, 549-603.

Bhattacharya, R. (1982). On the functional central limit theorem and the law of the iterated logarithm for Markov processes. *Z. Wahrsch. Verw. Gebiete* **60**, 185-201.

Doll, J. D., Rossky, P. J. and Friedman, H. L. (1978). Brownian dynamics as smart Monte Carlo simulation. *J. Chemical Physics* **69**, 4628-4633.

Godambe, V. P. and Kale, B. K. (1991). Estimating functions: an overview. *Oxford Statistical Science series* **7**, 3-20.

Hastings, W. K. (1970). Monte Carlo Sampling Methods using Markov chains and their applications. *Biometrika* **57**, 97-109.

Kijima, M. (1997). *Markov processes for stochastic modeling.* Chapman & Hall.

Mira, A. (2001). Ordering and Improving MCMC performances. *Statist. Sci.* **16**, 340-350.

Mira, A. and Geyer, C. J. (1999). Ordering Monte Carlo Markov Chains. School of Statistics University of Minnesota, Technical Report **632**.

Mira, A. and Leisen, F. (2007). Covariance Ordering for discrete and continuous time Markov chain. Accepted for publication on *Statist. Sinica*.

Mira, A. and Baddeley, A. (2001). Performance of time-invariance estimators. Department of Mathematics and Statistics, University of Western Australia (Perth, Australia), Technical Report **15**.

Mira, A. and Baddeley, A. (2007). Deriving Bayesian and frequentist estimators from time-invariance estimating equations: a unifying approach (with discussion). In *Bayesian Statistics* **8** (Edited by J. M. Bernardo, et al.), 325-348. Oxford University Press.

Peskun, P. H. (1973). Optimum Monte Carlo sampling using Markov Chains. *Biometrika* **60**, 607-612.

Roberts, G. O. and Tweedie, R. L. (1996a). Exponential convergence of Langevin diffusions and their discrete approximations. *Bernoulli* **2**, 341-364.

Roberts, G. O. and Tweedie, R. L. (1996b). Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms. *Biometrika* **83**, 95-110.

Stramer O. and Tweedie, R. L. (1999a). Langevin-type models I: Diffusions with given stationary distributions and their discretizations. *Methodology and Computing in Applied Probability* **1**, 283-306.

Stramer O. and Tweedie, R. L. (1999b). Langevin-type models II: Diffusions with given stationary distributions and their discretizations. *Methodology and Computing in Applied Probability* **1**, 307-328.

Tierney, L. (1994). Markov chains for exploring posterior distributions. *Ann. Statist.* **22**, 1701-1762.

Tierney, L. (1998). A note on Metropolis Hastings kernels for general state spaces. *Ann. Appl. Probab.* **8**, 1-9.

Università dell'Insubria, Dipartimento di Economia, Via Monte Generoso 71, 21100 Varese, Italy.

E-mail: fabrizio.leisen@gmail.com

Università dell'Insubria, Dipartimento di Economia, Via Monte Generoso 71, 21100 Varese, Italy.

E-mail: antonietta.mira@uninsubria.it

(Received October 2006; accepted April 2007)