Statistica Sinica **19** (2009), 651-666

COVARIANCE ORDERING FOR DISCRETE AND CONTINUOUS TIME MARKOV CHAINS

Antonietta Mira and Fabrizio Leisen

University of Insubria and Universidad de Navarra

Abstract: The covariance ordering, for discrete and continuous time Markov chains, is defined and studied. This partial ordering gives a necessary and sufficient condition for MCMC estimators to have small asymptotic variance. Connections between this ordering, eigenvalues, and suprema of the spectrum of the Markov transition kernel, are provided. A representation of the asymptotic variance of MCMC estimators in terms of eigenvalues and eigenvectors is extended to continuous time. This representation is used to establish convergence of the asymptotic variance of MCMC estimators derived from the discretization of a continuous time Markov chain.

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

1. Introduction

The basic idea of Markov chain Monte Carlo (MCMC) is that of approximating an expectation $\mu = E_\pi\{f(X)\} = \int f(x)\pi(dx)$ by an empirical average $\hat{\mu}_n = (1/n)\sum_{i=1}^n f(X_i)$ over the sample path of a discrete time Markov chain $X_1, \ldots, X_n$ having π as its unique stationary and limiting distribution. If the Markov chain and the function f are "well behaved" (Tierney (1994)), then $\hat{\mu}_n$ will obey the Central Limit Theorem (CLT): $\sqrt{n}\,(\hat{\mu}_n - \mu) \xrightarrow{\mathcal{D}} N(0, \sigma^2)$. Typically, for a fixed probability distribution π, the asymptotic variance, σ^2, depends on both the function f and the structure of the Markov chain through its transition operator P. Thus we denote it by $v(f, P, \pi)$. If, for a particular function f and transition kernel P, the CLT does not hold, then we define $v(f, P, \pi)$ to be ∞. There are often many different Markov chains with a specified stationary distribution π. Which is best? Or, a simpler question, given just two chains to consider, which is better? Efficiency is the relevant criterion here, as everywhere else in statistics. The chain P is better than Q for estimating the expectation of the function f, if $v(f, P, \pi) < v(f, Q, \pi)$, (assuming both chains are stationary with respect to π so that $\hat{\mu}_n$ is an asymptotically unbiased estimator for μ). Applications in which one is interested in the expectation of a single function f are rare. Usually, expectations of several functions are of interest, sometimes of many different functions. For example, the posterior mean and variance of a Bayesian image reconstruction involve expectations for millions of pixels and trillions of pairs of pixels. A likelihood function calculated by MCMC involves an expectation depending on a continuous parameter, that is, an uncountable family of expectations. Thus, contrary to what is often done in classical statistical inference when looking for minimum variance estimates, we do not assume any prior knowledge of the function whose expectation we want to evaluate. So, given two Markov chains P and Q stationary with respect to π, we say that P is more efficient than Q if $v(f,P,\pi) \leq v(f,Q,\pi)$ for all functions f that obey the CLT (efficiency partial ordering).

In Section 2, we recall two partial orderings for discrete time Markov chains that imply the efficiency ordering. One is Peskun ordering (1973), extended by Tierney (1998) to general state spaces, and the other is the covariance ordering introduced by Mira and Geyer (1999). Ordering Markov chains, is also important in the study of time invariance estimating equations (abbreviated TIEE), a general framework to construct estimators for a generic model (Baddeley (2000)). A criterion to study the performance of time invariance estimators is the Godambe-Heyde asymptotic variance, that is strictly connected with ordering Markov chains. Indeed, Mira and Baddeley (2001), have shown that Peskun ordering is a necessary condition for the Godambe-Heyde ordering. All the results in the literature regarding orderings of Markov Chains for MCMC or TIEE purposes (to our knowledge) are for discrete time Markov chains, and nothing has been said about continuous time. Only recently, Leisen and Mira (2008) have extended the Peskun ordering to continuous time Markov chains and, in Section 3, we recall the basic definition and theorems. Theoretically this result is important in the TIEE framework to study the performance of estimators, and could open new simulation strategies in the MCMC contest. How can a continuous time Markov chain be used to simulate a probability distribution? Leisen and Mira (2008) have intuitively answered this question in finite state state spaces by using a result that is formally proved in Section 4 of this paper. To distinguish the asymptotic variance of a continuous time Markov chains by the asymptotic variance of the discrete time Markov Chains, we use the notation $v(f,Q)$, instead of $v(f,Q,\pi)$. Relevant facts about continuous time Markov chains, the CLT, and a rigorous definition of asymptotic variance will be given in Section 3. Moreover, in Section 3, we extend covariance ordering to continuous time Markov chains and establish the equivalence between covariance ordering and efficiency of continuous time Markov chains.

2. Ordering Discrete Time Markov Chains

Let $L^2(\pi)$ be the Hilbert space of measurable functions that have finite second moment with respect the measure π, and let $L_0^2(\pi)$ be the subset of $L^2(\pi)$ of functions having zero mean under π. We define the inner product on $L^2(\pi)$ by $\langle f, g\rangle = \int f(x)g(x)\pi(dx)$. In classical statistics, estimates are compared in terms of their asymptotic relative efficiency, likewise here we prefer a Markov chain if it produces estimators that are asymptotically more efficient on a sweep-by-sweep basis.

Definition 1. If P and Q are Markov chains with stationary distribution π, then P is at least as efficient as Q, $P \succeq_E Q$, if $v(f, P, \pi) \leq v(f, Q, \pi), \forall f \in L_0^2(\pi)$.

2.1. Peskun and Tierney ordering

Throughout the paper we consider Markov chains with values in a space E that can be finite or general. The following ordering was introduced by Peskun (1973) for finite state spaces.

Definition 2. Given two Markov chains Q_1, Q_2, stationary with respect to π, $Q_1 = \{q_{(1)ij}\}_{i,j\in E}$, $Q_2 = \{q_{(2)ij}\}_{i,j\in E}$, we say that Q_1 is better than Q_2 in the Peskun sense and write $Q_1 \succeq_P Q_2$, if $q_{(1)ij} \geq q_{(2)ij}, \forall i \neq j$.

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

Theorem 1. *Given two Markov chains Q_1, Q_2, reversible with respect to π, if Q_1 dominates Q_2 in the Peskun sense, then $v(f, Q_1, \pi) \leq v(f, Q_2, \pi), \forall f \in L_0^2(\pi)$.*

The first use of Peskun ordering appears in Peskun (1973), where the author shows that the Metropolis-Hastings algorithm (Tierney (1994)), the main algorithm used in MCMC, dominates a class of competitors reversible with respect to some π. The competitor algorithms considered by Peskun (1973) are all algorithms with the same propose/accept updating structure, and with symmetric acceptance probability (see also Baddeley (2000)). Tierney (1998) extended Peskun ordering to a general state space $(E, \mathcal{E})$, where $\mathcal{E}$ is the associated σ-algebra. We identify Markov chains with the corresponding transition kernels $Q(x, A) = \Pr(X_n \in A|X_{n-1} = x)$ for every set $A \in \mathcal{E}$, and let Qf be the operator on $L_2(\pi)$ induced by Q: $(Qf)(x) = \int Q(x, dy)f(y)$.

Definition 3. Let Q_1, Q_2 be 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 $Q_1(x, B \setminus \{x\}) \geq Q_2(x, B \setminus \{x\})$, $\forall B \in \mathcal{E}$.

The next theorem, due to Tierney (1998), extends Theorem 2.1.1 by Peskun (1973) from finite to general state spaces.

Theorem 2. *Given two Markov chains Q_1, Q_2, reversible with respect to π, if Q_1 dominates Q_2 in the Tierney sense, $v(f, Q_1, \pi) \leq v(f, Q_2, \pi), \forall f \in L^2(\pi)$.*

The proof of the last theorem uses the following result.

Theorem 3. *If $Q_1 \succeq_T Q_2$ then $Q_2 - Q_1$ is a positive operator.*

2.2. Covariance ordering

The Peskun criterion and the generalization given by Tierney order only a limited number of Markov chains. For example, the ordering does not allow a comparison between two distinct transition matrices having all zeros on the main diagonal, or two transition kernels for which $P(x, \{x\}) = 0$ for every x in the state space. The latter includes all Gibbs samplers with continuous full conditional distributions. Furthermore, if only one of the off-diagonal entries of $P - Q$ is "out of order" then P and Q are incomparable. A natural way to define a weaker ordering for comparing more Markov chains is given in the following definition.

Definition 4. P dominates Q in the covariance ordering, $P \succeq_C Q$, if $Q - P$ is a positive operator on $L_0^2(\pi)$, that is, if $\langle f, (Q-P)f \rangle \geq 0$, for every $f \in L_0^2(\pi)$.

Restricting ourselves to $L_0^2(\pi)$ does not reduce the generality of the previous definition (see Mira (2001)). The binary relation $\succeq_C$ defines a partial ordering on the space of reversible Markov chains with respect to π, since it is symmetric, anti-reflexive and transitive (see the Appendix). By Theorem 3 we have the following.

Theorem 4. *Let P, Q be two Markov Chains reversible with respect to π, then*

$$P \succeq_P Q \Rightarrow P \succeq_C Q.$$

The covariance ordering is equivalent to the Löwner partial ordering, ($\succeq_L$), on positive, bounded, linear operators on a Hilbert space, Löwner (1934). Löwner ordering is defined on positive operators, therefore we need to consider the Laplacian of P, $l_P = I - P$, instead of P. Since $P \succeq_P I$ for every P stationary with respect to π, we have that $l_P \geq 0$.

Definition 5. Let l_P, l_Q be positive, bounded, self-adjoint, linear operators on a Hilbert space.Then l_P dominates l_Q in the Löwner sense, $l_P \succeq_L l_Q$, if $l_P - l_Q \geq 0$.

The following conditions are equivalent:

1. $P \succeq_C Q$ i.e., $Q - P \geq 0$;
2. $l_P \succeq_L l_Q$ i.e., $l_P - l_Q \geq 0$.

A variety of inequalities are obtainable, for any partial ordering, once the order-preserving functions are identified. For the Löwner ordering or, better, for a generalization of it that does not require the operators to be positive, the following theorem characterizes the class of order preserving functions, see Löwner (1934). Let f be a bounded real-valued function of a real variable, x, defined in an interval. Consider a bounded self-adjoint operator, A, on a Hilbert space, H, whose spectrum lies in the domain of f. Then by $f(A)$ we mean the self-adjoint operator defined as

$$f(A) = \int f(\lambda) E_A(d\lambda), \tag{2.1}$$

where $E_A(\cdot)$ is the spectral measure defined on the Borel subset of $\sigma(A)$, the spectrum of A (see Theorem 2.2, p. 269 of Conway (1985)). Moreover, if g is a complex function, $(Im\, g)$ means the imaginary part of g.

Theorem 5. *A necessary and sufficient condition for a continuous real-valued function f on the interval (I_1, I_2) to have the property that $f(A) \leq f(B)$ for all pairs of bounded, self-adjoint operators A and B with $\sigma(A), \sigma(B) \subseteq (I_1, I_2)$ and $A \leq B$, is that f is analytic in (I_1, I_2) and can be analytically continued into the whole upper half-plane with $(Im\, f) \geq 0$.*

Further characterizations of such classes of functions can be found in Korányi (1956). A function that satisfies the conditions of Theorem 5 is $h(x) = (ax + b) /(cx + d)$ with $ad - bc > 0$ either in $x > -d/c$ or $x < -d/c$. For example, take $a = b = d = 1$ and $c = -1$ and then $ad - bc = 2 > 0$, so $h(x) = (1 + x)/(1 - x)$ preserves the ordering for $x < 1$. Thus

$$P \succeq_C Q \quad \text{if and only if} \quad Q \geq P \quad \text{if and only if} \quad \frac{I + Q}{I - Q} \geq \frac{I + P}{I - P}.$$

We use this fact to prove that the covariance ordering is equivalent to the efficiency ordering. This provides a characterization of the efficiency ordering.

Theorem 6. *Let P and Q be reversible and irreducible transition kernels with stationary distribution π. Then $P \succeq_E Q$ if and only if $P \succeq_C Q$.*

For proving the theorem we need some technical lemmas and propositions. We denote the *domain* and *range* of an operator A by $D(A)$ and $R(A)$, respectively. An operator on $L_0^2(\pi)$ is said to be *densely defined* if $D(A)$ is dense in $L_0^2(\pi)$. We recall that an operator is *positive*, $A \geq 0$, if $\langle g, Ag \rangle \geq 0$ $\forall g \in L_0^2(\pi)$, and that $A^{-1/2}$ has been defined in (2.1).

Lemma 1. *Let A be a positive, self-adjoint, injective, bounded operator. Then, for every $g \in D(A)$, $\langle g, Ag\rangle = \sup_{f\in D(A^{-1/2})}\big[2\langle f,g\rangle - \big\langle A^{-1/2}f, A^{-1/2}f\big\rangle\big]$.*

Proof. Since A is positive, A^{-1} is also positive. This allows us to take square roots of both A and A^{-1}. Let $h = Ag$ so $g = A^{-1}h$. Clearly $D(A^{-1}) \subset D(A^{-1/2})$, and for every $f \in D(A^{-1/2})$,

$$\begin{aligned}0 &\le \Big\langle A^{-\frac{1}{2}}(f-h), A^{-\frac{1}{2}}(f-h)\Big\rangle\\ &= \Big\langle A^{-\frac{1}{2}}f, A^{-\frac{1}{2}}f\Big\rangle - 2\Big\langle A^{-\frac{1}{2}}f, A^{-\frac{1}{2}}h\Big\rangle + \Big\langle A^{-\frac{1}{2}}h, A^{-\frac{1}{2}}h\Big\rangle.\end{aligned}$$

Now substitute $h = Ag$ and use the fact that $\langle f,g\rangle = \langle g,f\rangle$, true in a real Hilbert space but not true in complex Hilbert spaces. Thus

$$\langle g, Ag\rangle \ge \big[2\langle f,g\rangle - \Big\langle A^{-\frac{1}{2}}f, A^{-\frac{1}{2}}f\Big\rangle\big], \qquad \forall f \in D(A^{-\frac{1}{2}}) \tag{2.2}$$

and the supremum is achieved by taking $f = h$ since, in this case, the right hand side equals the left hand side in (2.2).

Corollary 1. *Suppose A and B are positive, self-adjoint, injective, bounded operators. If $\big\langle B^{-1/2}f, B^{-1/2}f\big\rangle \le \big\langle A^{-1/2}f, A^{-1/2}f\big\rangle, \forall f \in D(A^{-1/2})$, and $D(A^{-1/2}) \subset D(B^{-1/2})$, then $A \le B$.*

Proof. By Lemma 1 we have, for every $g \in D(A) = D(B)$,

$$\begin{aligned}\langle g, Bg\rangle &= \sup_{f\in D(B^{-\frac{1}{2}})}\big[2\langle f,g\rangle - \Big\langle B^{-\frac{1}{2}}f, B^{-\frac{1}{2}}f\Big\rangle\big]\\ &\ge \sup_{f\in D(A^{-\frac{1}{2}})}\big[2\langle f,g\rangle - \Big\langle A^{-\frac{1}{2}}f, A^{-\frac{1}{2}}f\Big\rangle\big] = \langle g, Ag\rangle.\end{aligned}$$

Lemma 2. *For a transition kernel P with stationary distribution π, the asymptotic variance can be written as $v(g,P) = \big\langle g, [2l_P^{-1} - I]g\big\rangle, \forall g \in D(l_P^{-1})$.*

Proof. For any $g \in D(l_P^{-1})$ there exists an $f \in L_0^2(\pi)$ such that $g = l_P f$ so that $Pf = f - g$. Using a result in Gordin and Lifsic (1978) we can write the asymptotic variance as

$$\begin{aligned}v(g,P) &= \|f\|^2 - \|Pf\|^2 = \|f\|^2 - \|f-g\|^2\\ &= \langle f,f\rangle - \langle f-g, f-g\rangle = 2\langle g,f\rangle - \langle g,g\rangle\\ &= 2\big\langle g, l_P^{-1}g\big\rangle - \langle g,g\rangle = \big\langle g, [2l_P^{-1} - I]g\big\rangle.\end{aligned}$$

The previous result generalizes the representation of the asymptotic variance given in Kemeny and Snell (1969) for finite state spaces. Notice that the transition kernel does not need to be reversible for Lemma 2 to hold.

Proof of Theorem 6. Let us consider two cases depending on whether the Laplacian is an invertible operator on $L_0^2(\pi)$.

Case (1) Suppose l_P is invertible. Let $h(l_P) = 2/l_P - I = (I+P)/(I-P)$. Using Lemma 2, $P \succeq_E Q$ holds if and only if, for all $f \in L_0^2(\pi)$, $\langle f, h(l_P)f\rangle \leq \langle f, h(l_Q)f\rangle$ which, by definition, is equivalent to

$$h(l_P) \leq h(l_Q) \tag{2.3}$$

and, by Theorem 5, this is true if and only if

$$Q - P \geq 0. \tag{2.4}$$

Case (2) If l_P is not invertible, we have to prove the equivalence of (2.3) and (2.4) without using Theorem 5 on any non-invertible operator.

First we prove $P \succeq_C Q$ implies $P \succeq_E Q$. Assume $P \succeq_C Q$, and let $K_{\epsilon P} = I-(1-\epsilon)P$ for $0 < \epsilon < 1$. $K_{\epsilon P}$ is invertible since its spectrum $\sigma(K_{\epsilon P}) \subseteq (\epsilon, 2-\epsilon)$ does not contain zero. Furthermore, $h(K_{\epsilon P})$ is also invertible since its spectrum is $\sigma(h(K_{\epsilon P})) = h(\sigma(K_{\epsilon P})) \subseteq (\epsilon/(2-\epsilon), (2-\epsilon)/\epsilon)$. Then, for all $0 < \epsilon < 1$, $Q - P \geq 0$ implies $K_{\epsilon Q} \leq K_{\epsilon P}$ and, from case (1), this is true if and only if

$$\langle f, h(K_{\epsilon,Q})f\rangle \geq \langle f, h(K_{\epsilon,P})f\rangle, \qquad \forall f \in L_0^2(\pi). \tag{2.5}$$

We now want to take the limit as $\epsilon \to 0$. Consider

$$\langle f, h(K_{\epsilon,P})f\rangle = \int \frac{1+(1-\epsilon)\lambda}{1-(1-\epsilon)\lambda} E_{fP}(d\lambda).$$

The derivative of the integrand with respect to ϵ is $-2\lambda/[1-(1-\epsilon)\lambda]^2$ thus, for $\lambda \in [-1, 0)$, the integrand is increasing in ϵ while for $\lambda \in [0, +1)$, the integrand is decreasing. This suggests that we write

$$\langle f, h[K_{\epsilon,P}]f\rangle = \int_{-1}^{0} \frac{1+(1-\epsilon)\lambda}{1-(1-\epsilon)\lambda} E_{fP}(d\lambda) + \int_{0}^{1} \frac{1+(1-\epsilon)\lambda}{1-(1-\epsilon)\lambda} E_{fP}(d\lambda).$$

For every $\lambda \in \sigma(P)$ and every $\epsilon \in (0,1)$, the integrals are finite by construction, therefore a modified version of the standard monotone convergence theorem (Fristed and Gray (1997)) can be used to take the limit inside the integral and we get that (2.5) implies (2.3). Hence $P \succeq_C Q$ implies $P \succeq_E Q$.

Now we prove the implication in the other direction: $P \succeq_E Q$ implies $P \succeq_C Q$. Assuming $P \succeq_E Q$ we have that (2.3) holds. We now use the properties of the Laplacian and, in particular, the fact that the range of $l_Q^{1/2}$ is the set of functions that have a finite asymptotic variance, see Kipnis and Varadhan (1986); i.e., $v(f,P) \leq v(f,Q) < \infty, \forall f \in R(l_Q^{1/2})$ and $R(l_Q^{1/2}) \subseteq R(l_P^{1/2})$. It follows that

$$\left\langle l_P^{-\frac{1}{2}} f, l_P^{-\frac{1}{2}} f \right\rangle \leq \left\langle l_Q^{-\frac{1}{2}} f, l_Q^{-\frac{1}{2}} f \right\rangle \qquad \forall f \in R(l_Q^{\frac{1}{2}}) = D(l_Q^{-\frac{1}{2}}) \tag{2.6}$$

and, by Corollary 1, we have $l_Q \le l_P$, hence $P \succeq_C Q$.

The final part of this subsection is devoted to two examples where Peskun ordering fails while the covariance ordering holds. The first is a toy example, the second refers to data augmentation algorithms.

Toy example. Let P and A be the matrices

$$P = \begin{pmatrix} 0.3 & 0.3 & 0.2 & 0.2 \\ 0.3 & 0.3 & 0.2 & 0.2 \\ 0.2 & 0.2 & 0.3 & 0.3 \\ 0.2 & 0.2 & 0.3 & 0.3 \end{pmatrix} \qquad A = \begin{pmatrix} 0.1 & 0.1 & -0.1 & -0.1 \\ 0.1 & 0.1 & -0.1 & -0.1 \\ -0.1 & -0.1 & 0.1 & 0.1 \\ -0.1 & -0.1 & 0.1 & 0.1 \end{pmatrix}$$

and let $Q = P+A$. Then both P and Q are reversible with respect to the uniform distribution, but are not comparable in the Peskun sense while $P \succeq_C Q$.

Data augmentation. By using the covariance ordering, Hobert and Marchev (2008), prove that a class of data augmentation algorithms is better than the usual data augmentation algorithm (DA) of Tanner and Wong (1987). This class contains the PX-DA algorithm of Liu and Wu (1999) and the marginal data augmentation algorithm (MA) of Meng and van Dyk (1999). Suppose that we want to sample from $f_X(x)$ on a space $\mathcal{Y}$, and that a joint density $f(x, y)$ having $f_X(x)$ as its marginal is available. Furthermore, assume that it is straightforward to sample from $f_{X|Y}(x|y)$ and $f_{Y|X}(y|x)$. Then the DA reversible kernel to sample from $f_X(x)$ is given by $P(x|x') = \int_{\mathcal{Y}} f_{X|Y}(x|y) f_{Y|X}(y|x')dy$. If R is a Markov kernel reversible with respect to $f_Y(y)$, one can build another reversible algorithm (wrt to $f_X(x)$) as $P_R(x|x') = \int_{\mathcal{Y}} \int_{\mathcal{Y}} f_{X|Y}(x|y')R(y, dy')f_{Y|X}(y|x')dy$. Hobert and Marchev (2008) prove that $P_R \succeq_E P$ by showing that $P_R \succeq_C P$. In the sequel we construct a specific example where P and P_R are not comparable in the Peskun sense. Let $f_X = [3/8, 3/8, 1/4]$ and take $f(x, y) = \begin{pmatrix} 3/10 & 1/20 & 1/40 \\ 2/10 & 1/10 & 3/40 \\ 1/10 & 1/20 & 1/10 \end{pmatrix}$.
The DA kernel is $P = f_{Y|X} f_{X|Y} = \begin{pmatrix} 0.4416667 & 0.3583333 & 0.2 \\ 0.3583333 & 0.3861111 & 0.2555556 \\ 0.3 & 0.3833333 & 0.3166667 \end{pmatrix}$ and if $R = f_{X|Y} f_{Y|X}$, then R is reversible w.r.t. $f_Y(y)$ and $P_R = f_{Y|X} R f_{X|Y} = \begin{pmatrix} 0.3834722 & 0.3732870 & 0.2432407 \\ 0.3732870 & 0.3754475 & 0.2512654 \\ 0.3648611 & 0.3768981 & 0.2582407 \end{pmatrix}$. P and P_R are not comparable in Peskun sense but $P - P_R$ is positive semidefinite, so $P_R \succeq_C P$.

2.3. Orderings and eigenvalues

Let us first consider finite state spaces. Let $\{\lambda_{0P}, \lambda_{1P}, \ldots\}$ be the eigenvalues of P, arranged in decreasing order, and let $\{e_{0P}, e_{1P}, \ldots\}$ be the corresponding normalized right eigenvectors, so that $Pe_{jP} = \lambda_{jP}e_{jP}$, $j = 0, 1, \ldots$. For P stationary with respect to π, there is an eigenvalue equal to one, λ_{0P}, which is associated with the constant eigenvector. Since this is always the case let us restrict our attention to the eigenvalues associated with non-constant eigenvectors. Reversibility of a transition kernel ensures that the eigenvalues and eigenvectors are real. The following theorem is proved in Mira (2001).

Theorem 7. *For P, Q reversible with respect to π, $Q - P \geq 0$ if and only if $\lambda_{iP} \leq \lambda_{iQ}$ for all i.*

The previous theorem is a known fact for symmetric matrices. In our setting neither P nor Q need to be symmetric but, if we consider them as operators on $L^2(\pi)$, they are indeed self-adjoint operators, provided that the detailed balance condition holds. By Theorem 3, $P \succeq_P Q$ implies that $Q - P \geq 0$, thus Peskun ordering induces an ordering on all the eigenvalues of the two transition matrices. This proof can be generalized to compact operators on Hilbert spaces since their spectra are either empty, finite, or countable with zero as the only limit point, (Conway (1985)). But, as noticed in Chan and Geyer (1994), not many Markov chains used for MCMC purposes have compact transition operators.

Let us now move to general state spaces. While in finite state spaces we have a finite number of eigenvalues and it makes sense to compare and order them, in general state spaces we cannot talk about eigenvalues anymore, but need to introduce the concept of a spectrum. Let $\sigma(P)$ be the *spectrum* of P considered as an operator on $L^2(\pi)$, that is, the set of λ's such that $\lambda I - P$ is not invertible, where I denotes the identity operator on $L^2(\pi)$. The spectrum includes the eigenvalues, the λ's for which $\lambda I - P$ is not one-to-one, but it also includes the values λ such that $\lambda I - P$ is not onto. For linear operators on finite dimensional vector spaces, one-to-one and onto are equivalent so that $\sigma(P)$ is the set of the eigenvalues of P. The *norm* of a linear operator on $L^2(\pi)$ is defined by $\|P\| = \sup_{\substack{u \in L^2(\pi) \\ u \neq 0}} \|Pu\|/\|u\|$, where $\|u\|^2 = \langle u, u \rangle$. The spectrum is a non-empty closed subset of the interval $[-1, +1]$ since the norm of P is less than or equal to one, by Jensen's inequality, and the norm of an operator bounds the spectrum (Conway (1985, Proposition 1.11 (e), p.239)). On general state spaces it does not make sense to say that the spectrum of one operator is smaller than the spectrum of another operator; we can, at most, compare the suprema of the spectra and this is what we will do. For reversible geometrically ergodic chains, all the eigenvalues but the principal one, $\lambda_{0P} = 1$, are bounded away from ± 1, see Roberts and Rosenthal (1997).

When considering a transition kernel as an operator on $L_0^2(\pi)$ we eliminate from its spectrum the eigenvalue equal to one associated with constant functions. Unless otherwise stated, in the sequel a transition kernel will be considered as an operator on $L_0^2(\pi)$. Let $\lambda_{\max,P} = \sup\{\, \lambda : \lambda \in \sigma(P)\, \}$, then the following theorem is the analogue, for a general state space, of Theorem 7:

Theorem 8. *Let P, Q be reversible with respect to π, and assume $P \succeq_C Q$, then*

$$\lambda_{\max,P} \leq \lambda_{\max,Q}. \tag{2.7}$$

Proof. It follows directly from Theorem X.4.2 of Dunford and Schwartz (1963) that, for any bounded self-adjoint operator A on a Hilbert space, we have $\lambda_{\max,A} = \sup_{\|f\|=1} \langle f, Af\rangle$. Thus (2.7) holds whenever $Q - P \geq 0$, and Theorem 3 finishes the proof.

3. Continuous Time Markov Chains and Their Orderings

Let $\{X(t)\}_{t\in\Re^+}$ be a continuous time MC (CTMC) taking values on a finite state space E. Let $G = \{g_{ij}\}_{i,j\in\mathrm{E}}$ 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 $\pi_i g_{ij} = \pi_j g_{ji}$ $\forall i, j \in \mathrm{E}$. Let I be the identity matrix, $c = \sup_i |g_{ii}|$, and $\nu \geq c$, then $P_\nu = I + G/\nu$ is a stochastic matrix. Note that if G is reversible with respect to π, then so is P_ν, $\forall \nu$. We could use such CTMC for MCMC purposes in the following way. Assume without loss of generality, that f has zero mean and finite variance under π, $f \in L_0^2(\pi)$. 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 $\hat{\mu}_{nt} = (1/\sqrt{n}) \int_0^{nt} f(X(s))ds$ 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

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

where g belongs to the domain of the generator and is such that $Gg = f$.

In Proposition 2.4 of Bhattacharya (1982), it is proved 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 the range of G, without the additional assumption of boundedness and mutual absolute continuity.

3.1. Peskun ordering for continuous time Markov chains

Let E be a finite state space. The following ordering has been introduced by Leisen and Mira (2008).

Definition 6. Suppose that $G_1 = \{g_{(1)ij}\}$ and $G_2 = \{g_{(2)ij}\}$ are the generators of CTMCs stationary with respect to π (i.e., $\pi G_1 = 0$, $\pi G_2 = 0$). We say that G_1 dominates G_2 in the Peskun sense, and write $G_1 \succeq_{EP} G_2$, if $g_{(1)ij} \geq g_{(2)ij}, \forall i \neq j$.

Now, let E be a general state space and $\mathcal{E}$ the associated sigma-algebra. We begin by recalling some definitions and results from Leisen and Mira (2008) on Peskun ordering and then extend, to general state spaces, the covariance ordering. Consider an homogeneous continuous time Markov chain, $\{X_t\}_{t\in\Re^+}$, taking values on E, with transition kernel $P(t, x, dy)$ and generator $G : D(G) \to R(G)$, where $D(G)$ and $R(G)$ are the domain and range of G, respectively. If the generator of the process can be written as an operator

$$Gf(x) = \int f(y)Q(x, dy), \tag{3.1}$$

where the kernel Q is defined in terms of the transition kernel P, $Q(x, dy) = \frac{\partial}{\partial t}P(t, x, dy) \mid_{t=0}$, then, in the general case, Peskun ordering has been extended, in Leisen and Mira (2008), in the following way.

Definition 7. Let G_1 and G_2 be the generators of two CTMCs admitting the representation (3.1), with kernels Q_1 and Q_2 respectively, both stationary with respect to a common distribution π, taking values on E. Assume $\sup_x Q_i(x, E \setminus \{x\}) < \infty, i = 1, 2$. Then G_1 dominates G_2 in the Tierney ordering, $G_1 \succeq_{EP} G_2$, if $Q_1(x, A \setminus \{x\}) \geq Q_2(x, A \setminus \{x\})\ \forall A \in \mathcal{E}$.

Then, for E finite or general, and in the hypothesis of the previous definitions, two results are available in Leisen and Mira (2008):

Theorem 9. *If $G_1 \succeq_{EP} G_2$ and if the corresponding CTMCs are reversible, then $G_2 - G_1$ is a positive operator.*

Theorem 10. *If $G_1 \succeq_{EP} G_2$ and if the corresponding CTMCs are reversible, then $v(f, G_1) \leq v(f, G_2)\ \forall f \in R(G_1) \cap R(G_2)$, 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.*

3.2. Covariance ordering for continuous time Markov chains

In this section, the covariance ordering for continuous time is introduced. We start with a few definitions.

Definition 8. Let E be a finite state space and let G_1, G_2 be stationary with respect to π. We say that G_1 dominates G_2 in the covariance ordering, and write $G_1 \succeq_{EC} G_2$, if $G_2 - G_1$ is a positive operator on $D(G_1) \cap D(G_2)$.

If E is a general state space and $\mathcal{E}$ the associated sigma-algebra, the covariance ordering is defined in the following way.

Definition 9. Let G_1 and G_2 be the generators of two CTMCs admitting the representation (3.1), with kernels Q_1 and Q_2, respectively, both stationary with respect to a common distribution π taking values on E. Assume $\sup_x Q_i(x, E \setminus \{x\}) < \infty, i = 1, 2$. Then G_1 dominates G_2 in the covariance ordering, $G_1 \succeq_{EC} G_2$, if $G_2 - G_1$ is a positive operator on $D(G_1) \cap D(G_2)$.

It is easy to show that the continuous covariance ordering is a partial ordering.

Theorem 11. *Given two CTMCs on a state space E, with generators G_1 and G_2 reversible w.r.t. a distribution π, with the representation (3.1) and* $\sup_x Q_i(x, E \setminus \{x\}) < \infty, i = 1, 2$, *in the general case, the following are equivalent.*

1. $G_1 \succeq_{EC} G_2$,
2. $v(f, G_1) \leq v(f, G_2)$ *for all functions* $f \in R(G_1) \cap R(G_2)$.

Proof. "(1) $\Rightarrow$ (2)" is a little modification of the proof of Theorem 10 in Leisen and Mira (2008).
"(2) $\Rightarrow$ (1)" For all functions $f \in R(G_1) \cap R(G_2)$, we have:

$$v(f, G_i) = -2\langle f, g_i \rangle, \qquad i = 1, 2, \tag{3.2}$$

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

$$G_i g_i = f, \qquad i = 1, 2. \tag{3.3}$$

We have that

$$\begin{aligned} v(f, G_1) &= -2\langle G_1 g_1, g_1 \rangle = -2\langle G_1(g_1 - g_2 + g_2), (g_1 - g_2 + g_2) \rangle \\ &= -2\langle G_1(g_1 - g_2), (g_1 - g_2) \rangle - 2\langle G_1 g_1, g_2 \rangle \\ &\quad -2\langle G_1 g_2, g_1 \rangle + 2\langle G_1 g_2, g_2 \rangle \\ &\geq -2\langle G_2 g_2, g_2 \rangle - 2\langle G_2 g_2, g_2 \rangle + 2\langle G_1 g_2, g_2 \rangle, \end{aligned}$$

where the last inequality follows from the self-adjontness of G_1 and G_2, by (3.3) and from the fact that $-2\langle G_1(g_1 - g_2), (g_1 - g_2) \rangle \geq 0$. So, from the hypothesis, $-2\langle G_2 g_2, g_2 \rangle - 2\langle G_2 g_2, g_2 \rangle + 2\langle G_1 g_2, g_2 \rangle \leq v(f, G_1) \leq v(f, G_2) = -2\langle G_2 g_2, g_2 \rangle$, which gives $\langle (G_2 - G_1) g_2, g_2 \rangle \geq 0$, and concludes the proof.

3.3. Continuous orderings and eigenvalues

In this section, we give, for continuous time, analogous theorems as the ones given in Section 2.3. As in Section 2.3, first consider finite state spaces.

Theorem 12. *For G_1, G_2 generators of Markov chains reversible with respect to π, if $G_2 - G_1 \geq 0$, then $\lambda_{iG_1} \leq \lambda_{iG_2}$ for all i.*

Proof. Let $c_1 = \sup_i |g_{(1)ii}|, c_2 = \sup_i |g_{(2)ii}|$, and $\nu \geq \max(c_1, c_2)$. Define $P_1(\nu) = I + G_1/\nu$ and $P_2(\nu) = I + G_2/\nu$ We have that $G_1 = \nu(P_1(\nu) - I)$ and $G_2 = \nu(P_2(\nu) - I)$. If $G_2 - G_1 \geq 0$, it follows that $P_2(\nu) - P_1(\nu) \geq 0$; i.e., for Theorem 7, $\lambda_{iP_1(\nu)} \leq \lambda_{iP_2(\nu)}$ for all i. But $\lambda_{iP_1(\nu)} = 1 + \lambda_{iG_1}/\nu$ and $\lambda_{iP_2(\nu)} = 1 + \lambda_{iG_2}/\nu$, and so $\lambda_{iG_1} \leq \lambda_{iG_2}$ for all i.

Let us now move to general state spaces. For a generator G that admits the representation (3.1), let $\lambda_{\max,G} = \sup\{\, \lambda : \lambda \in \sigma(G) \,\}$.

Theorem 13. *Given two Markov chains with generators G_1 and G_2 reversible with respect to π, suppose $G_1 \succeq_{EC} G_2$. Then*

$$\lambda_{\max,G_1} \leq \lambda_{\max,G_2}. \tag{3.4}$$

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

$$P_{1\nu}(x, dy) = \delta_x(dy) + \frac{1}{\nu} Q_1(x, dy) \text{ and } P_{2\nu}(x, dy) = \delta_x(dy) + \frac{1}{\nu} Q_2(x, dy) \tag{3.5}$$

are transition kernel of CTMCs reversible with respect to π, and such that $P_{1\nu} \succeq_P P_{2\nu}$. By Theorem 3, it then follows that $P_{2\nu} - P_{1\nu} = (Q_2 - Q_1)/\nu$ is a positive operator. So, from Theorem 8, we have that $\lambda_{\max,P_{1\nu}} \leq \lambda_{\max,P_{2\nu}}$. From (3.5) and from the fact that $\lambda_{\max,P_{1\nu}} = \sup_{||f||=1} \langle f, P_{1\nu} f\rangle$ and $\lambda_{\max,P_{2\nu}} = \sup_{||f||=1} \langle f, P_{2\nu} f\rangle$, the conclusion follows.

4. Asymptotic Variance: From Discrete to Continuous Time

Throughout this section we consider a finite state space $E = \{1, \ldots, N\}$. We recall some known facts on discrete and continuous time Markov chains.

Let $\{X(t)\}_{t \in \mathbb{R}}$ be a Markov chain on E with generator $Q = \{q_{ij}\}_{i,j \in E}$ reversible with respect a probability distribution π. Let $0 = \beta_1 > \cdots > \beta_N$, be the eigenvalues of Q, and let u_i and v_i be the eigenvectors, respectively left and right, of Q; i.e., $u_i^T Q = \beta_i u_i^T$ and $Qv_i = \beta_i v_i$, $i = 1, \ldots, N$. Then, the t-step transition matrix of the CTMC that has Q as generator, has the following properties.

1. $P(t)$ is reversible with respect to π,
2. u_i, v_i are, respectively, left and right eigenvectors of $P(t)$ with eigenvalues: $1 = e^{\beta_1 t} > \cdots > e^{\beta_N t}$.

A function $f : E \to \mathbb{R}$ can be represented as

$$f = \sum_{i=1}^{N} \langle f, v_i \rangle v_i. \tag{3.6}$$

Moreover, we recall a representation of the asymptotic variance of a discrete time Markov chain in terms of eigenvalues (Bremaud (1998, p.235)):

Theorem 14. *Let P be the transition matrix of a discrete time Markov chain, $\{Y_n\}_{n\in\mathbb{N}}$ on E, reversible with respect to π. Let $v(f,P,\pi)$ be the asymptotic variance of the estimator $\hat{\mu}_n$. Then if $1=\lambda_1>\cdots>\lambda_N$ are the eigenvalues of P with right eigenvectors v_i, the asymptotic variance $v(f,P,\pi)$ admits the representation*

$$v(f,P,\pi)=\sum_{i=2}^{N}\frac{1+\lambda_i}{1-\lambda_i}|\langle f,v_i\rangle|^2.$$

We now give a continuous time analogous of Theorem 14.

Theorem 15. *Let $\{X(t)\}_{t\in\mathbb{R}}$ be a CTMC on E, reversible with respect to π with generator Q. Let $0=\beta_1>\cdots>\beta_N$ and v_i, $i=1,\ldots,N$, be the eigenvalues of Q with corresponding right eigenvectors. Then the asymptotic variance $v(f,Q)$ admits the representation $v(f,Q)=-2\sum_{i=2}^{N}\beta_i|\langle g,v_i\rangle|^2$.*

Proof. From (3.6) we have that $g=\sum_{i=1}^{N}\langle g,v_i\rangle v_i$. Hence,

$$\begin{aligned}
v(f,Q)&=-2\langle f,g\rangle=-2\langle Qg,g\rangle=-2\left\langle Qg,\sum_{i=1}^{N}\langle g,v_i\rangle v_i\right\rangle\\
&=-2\sum_{i=1}^{N}\left(\langle g,v_i\rangle\cdot\langle Qg,v_i\rangle\right)=-2\sum_{i=1}^{N}\left(\langle g,v_i\rangle\cdot\langle g,Qv_i\rangle\right)\\
&=-2\sum_{i=1}^{N}\left(\langle g,v_i\rangle\cdot\langle g,\beta_i v_i\rangle\right)=-2\sum_{i=2}^{N}\beta_i|\langle g,v_i\rangle|^2,
\end{aligned}$$

where the fifth equality follows from the self-adjointness of Q and the last from the fact that $\beta_1=0$.

4.1. A connection between discrete and continuous time Markov chains

The following theorem provides an interesting connection between the asymptotic variances of estimators obtained by running a continuous time Markov chain and a related discretization.

Theorem 16. *Let $\{X(t)\}_{t\in\mathbb{R}}$ be a CTMC on E, reversible with respect to π with generator Q. Let $0=\beta_1>\cdots>\beta_N$ and v_i, $i=1,\ldots,N$, be the eigenvalues with corresponding right eigenvectors. Let $P(\Delta)$ be the Δ-step matrix of the CTMC, $\Delta>0$ fixed, and $v(f,P(\Delta),\pi)$ be the asymptotic variance of the discrete time Markov chain that has $P(\Delta)$ as transition matrix. If $v(f,Q)$ is the asymptotic variance of the CTMC $X(t)$ and $f\in R(Q)$, then*

$$\Delta v(f,P(\Delta),\pi)\to v(f,Q)\quad \text{as } \Delta\to 0.$$

Proof. The eigenvalues of $P(\Delta)$ are $1 = e^{\beta_1 \Delta} > \cdots > e^{\beta_N \Delta}$, with corresponding left eigenvectors v_i. So, by Theorem 14,

$$v(f, P(\Delta), \pi) = \sum_{i=2}^{N} \frac{1 + e^{\beta_i \Delta}}{1 - e^{\beta_i \Delta}} |\langle f, v_i \rangle|^2.$$

But if $f \in R(A)$, there exists $g \in D(A)$ such that $Qg = f$. We have that

$$\langle f, v_i \rangle = \langle Qg, v_i \rangle = \langle g, Qv_i \rangle = \langle g, \beta_i v_i \rangle = \beta_i \langle g, v_i \rangle,$$

$$v(f, P(\Delta), \pi) = \sum_{i=2}^{N} \frac{1 + e^{\beta_i \Delta}}{1 - e^{\beta_i \Delta}} \beta_i^2 |\langle g, v_i \rangle|^2.$$

Thus

$$\Delta v(f, P(\Delta), \pi) = \sum_{i=2}^{N} \frac{\Delta}{1 - e^{\beta_i \Delta}} (1 + e^{\beta_i \Delta}) \beta_i^2 |\langle g, v_i \rangle|^2.$$

From $\lim_{\Delta \to 0} (1 + e^{\beta_i \Delta}) = 2$ and

$$\lim_{\Delta \to 0} \frac{\Delta}{1 - e^{\beta_i \Delta}} = \lim_{\Delta \to 0} \frac{\Delta}{1 - (1 + \beta_i \Delta + o(\Delta))} = -\frac{1}{\beta_i},$$

$$\Delta v(f, P(\Delta), \pi) \to \sum_{i=2}^{N} -\frac{1}{\beta_i} 2 \beta_i^2 |\langle g, v_i \rangle|^2 = -2 \sum_{i=2}^{N} \beta_i |\langle g, v_i \rangle|^2 = v(f, Q),$$

where the last equality follows from Theorem 15.

References

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

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.

Bremaud, P. (1998). *Markov Chains.* Springer Verlag, New York.

Chan, K. S. and Geyer, C. (1994). Discussion of the paper of Tierney: Markov chains for exploring posterior distributions. *Ann. Statist.* **22**, 1747-1758.

Conway, J. B. (1985). *A course in Functionals Analysis.* Springer Verlag.

Dunford, N. and Schwartz, J. (1963). *Linear Operator Part II.* First edition. John Wiley and Sons, New York.

Fristed, B. and Gray, L. (1997). *A Modern Approach to Probability Theory*, Birkhäuser, Boston.

Gordin, M. I. and Lifsic, B. A. (1978). The central limit theorem for stationary Markov processes. *Soviet Mathematics Doklady* **19**, 392-394.

Hobert, J. P. and Marchev, D. (2008). A theoretical comparison of the data augmentation, marginal augmentation and PX-DA algorithms. *Ann. Statist.* **32**, 532-554.

Kipnis, C. and Varadhan, S. R. S. (1986). Central limit theorem for additive functionals of reversible Markov processes and applications to simple exclusions. *Commun. Math. Phys.* **104**, 1-19.

Kemeny, J. G. and Snell, J. L. (1969). *Finite Markov Chain.* Princeton: Van Nostrand.

Korányi, A. (1956). On a theorem by Löwner and its connections with resolvents of selfadjoint transformations. *Acta Scientiarum Mathematicarum* **7**, 63-70.

Leisen, F. and Mira, A. (2008). An extension of Peskun and Tierney orderings to continuous time Markov chains. *Statist. Sinica* **18**, 1641-1651.

Liu, J. S. and Wu, Y. N. (1999). Parameter expansion for data augmentation. *J. Amer. Statist. Assoc.* **94**, 1264-1274.

Löwner, K. (1934). Uber monotone matrixfunktionen. *Mathematische Zeitschrift* **38**, 177-216.

Meng, X. L. and van Dyk, D. A. (1999). Seeking efficient data augmentation schemes via conditional and marginal augmentation. *Biometrika* **86**, 301-320.

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

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

Roberts, G. O. and Rosenthal, J. S. (1997). Geometric ergodicity and hybrid Markov chains. *Electronic Communications in Probability* **2**, 13-25.

Tanner, M. A. and Wong, W. H. (1987). The calculation of posterior distributions by data augmentation (with discussion). *J. Amer. Statist. Assoc.* **82**, 528-550.

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.

Department of Economics, Insubria University, Via Monte Generoso 71, 21100 Varese, Italy.

E-mail: amira@eco.uninsubria.it

Faculty of Economics, Universidad de Navarra, Campus Universitario, edificio de biblioteca (entrada este), 31008, Pamplona, Spain.

E-mail: fabrizio.leisen@gmail.com

(Received May 2007; accepted December 2007)