F. Leisen, A. Mira

An extension of Peskun ordering to continous time Markov chains

2006/10

UNIVERSITÀ DELL'INSUBRIA
FACOLTÀ DI ECONOMIA

http://eco.uninsubria.it

In questi quaderni vengono pubblicati i lavori dei docenti della Facoltà di Economia dell’Università dell’Insubria. La pubblicazione di contributi di altri studiosi, che abbiano un rapporto didattico o scientifico stabile con la Facoltà, può essere proposta da un professore della Facoltà, dopo che il contributo sia stato discusso pubblicamente. Il nome del proponente è riportato in nota all'articolo. I punti di vista espressi nei quaderni della Facoltà di Economia riflettono unicamente le opinioni degli autori, e non rispecchiano necessariamente quelli della Facoltà di Economia dell'Università dell'Insubria.

These Working papers collect the work of the Faculty of Economics of the University of Insubria. The publication of work by other Authors can be proposed by a member of the Faculty, provided that the paper has been presented in public. The name of the proposer is reported in a footnote. The views expressed in the Working papers reflect the opinions of the Authors only, and not necessarily the ones of the Economics Faculty of the University of Insubria.

© Copyright F. Leisen, A. Mira
Printed in Italy in July 2006
Università degli Studi dell'Insubria
Via Monte Generoso, 71, 21100 Varese, Italy

All rights reserved. No part of this paper may be reproduced in any form without permission of the Author.

An extension of Peskun ordering to continuous time Markov chains

Fabrizio Leisen*and Antonietta Mira†

July 27, 2006

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-invariance estimating equations in that it provides a necessary condition for ordering the asymptotic variances of the resulting estimators.

In this paper Peskun ordering is extended from discrete time to continuous time Markov chains.

Key words and phrases: Peskun ordering, Covariance ordering, Efficiency ordering, MCMC, time-invariance estimating equations, asymptotic variance, continuous time Markov chains.

1 Introduction

The class of Markov chains (MC) that are stationary with respect to a specified distribution, π, play an important role in two separate but connected fields, namely Markov chain Monte Carlo methods (MCMC), [7] and time-invariance estimating equations (TIEE), [1].

In MCMC we are interested in estimating the expected value of a function f with respect to a distribution π: $E_\pi f$. If we cannot compute such integral analytically either because the state space is too large or because π and/or f are too complicated, we can construct a Markov chain that has π as its unique stationary and limiting distribution. We then run the Markov chains for n time-steps, and produce a simulated path: $x_1, x_2, \cdots x_n$, possibly after a burn-in period that allows the Markov chain

*Università di Modena e Reggio Emilia, Dipartimento di Matematica, Via Campi 213/b Modena, E-mail: leisen.fabrizio@unimore.it

†Università dell'Insubria, Dipartimento di Economia, Via Monte Generoso 71, 21100 Varese, E-mail: antonietta.mira@uninsubria.it

to forget its initial distribution and to reach the stationary regime. We then estimate $\mu = E_\pi f$, by

$$\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} f(x_i).$$

Under regularity conditions, the strong law of large numbers and the central limit theorem, [16], ensure that $\hat{\mu}$ is asymptotically unbiased and give an expression for its asymptotic variance.

Time-invariance 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 can construct a Markov chain that has the model of interest as its stationary distribution. If we equate to zero the generator of the Markov chain, applied to some function, S, defined on the sample space and evaluated at the data, x, we obtain an unbiased estimating equation. A natural way to evaluate the performance of time-invariance estimators is the Godambe-Heyde asymptotic variance [6]. In [10] the authors show that Peskun ordering is a necessary condition for Godambe-Heyde ordering.

Both in the MCMC and the TIEE framework we have some degrees of freedom on how to choose the Markov chain since, given the distribution or the model of interest, there are many Markov chains that are stationary with respect to it. In the MCMC context this rises 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 similar question in the TIEE framework is the following: given two Markov chains stationary with respect to π_θ, which one produces time-invariance estimators of θ with smaller Godambe-Heyde asymptotic variance?

The first one who addressed the above question in the MCMC framework was Peskun, [11], who defined a partial ordering on the space of discrete time Markov chains defined on finite state spaces. The ordering was later extended by Tierney [17] to general state space Markov chain 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 [9]. While Peskun ordering gives a sufficient condition for efficiency ordering, the covariance ordering is both necessary and sufficient. Thus the covariance ordering is equivalent to the efficiency ordering.

The related question in the TIEE framework was first addressed by Mira and Baddeley [10]. The Authors show that Peskun ordering gives a necessary but not sufficient condition to 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 more than one successful attempt to construct an efficient proposal distribution in a 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 [5] and the idea was only later brought into the mainstream statistical literature in a discussion by Besag [2]. Theoretical convergence properties, in terms of speed of convergence to stationarity, of these type of MCMC algorithms have been extensively studied, see for example [12, 13, 14]. On the other hand, up 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 [1]. In order to study the performance of the resulting time-invariance estimators in this context the extension of Peskun ordering proposed in this paper would 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 will mainly focus on the MCMC context which is the original context where Peskun ordering was introduced. Furthermore we will consider finite state spaces (our results can be easily extended to countable state spaces). General state space Markov chains will be studied in further research.

2 Ordering of Markov chains relevant for MCMC purposes

2.1 Efficiency ordering

We begin by giving some definitions and setting up the notation.

Let $Q = \{q_{ij}\}_{ij \in \mathcal{E}}$ 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 $\mathcal{E}$.

We identify Markov chains with the corresponding transition matrices.

Let $\mathcal{S}$ the class of Markov chains stationary with respect to some given distribution of interest, say π; $\mathcal{R}$ 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 $v(f, Q)$, 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.

Definition 2.1 *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 $f \in L^2(\pi)$.*

3 Peskun ordering for discrete time Markov chains

Assume that $Q \in \mathcal{S}$ i.e.

$$\pi Q = \pi.$$

Definition 3.1 *Given two Markov chains $Q^1, Q^2 \in \mathcal{S}$, we say that Q^1 is better than Q^2 in the Peskun sense and write $Q^1 \succeq_P Q^2$ if*

$$q^1_{ij} \geq q^2_{ij}, \qquad \forall i \neq j$$

Theorem 3.1 *Given two Markov chains $Q^1, Q^2 \in \mathcal{R}$ if Q^1 dominates Q^2 in the Peskun sense then Q^1 dominates Q^2 in the efficiency ordering, i.e. $Q^1 \succeq_P Q^2 \Rightarrow Q^1 \succeq_E Q^2$.*

The first use of Peskun ordering appears in Peskun [11] where the author shows that the Metropolis-Hastings algorithm, [7], 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 [4]).

4 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 $\mathcal{E}$. Let

$$G = \{g_{ij}\}_{i,j\in\mathcal{E}}$$

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

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

Let I be the identity matrix, $c = \sup_i |g_{ii}|$ and $v \geq c$, then

$$P_\nu = I + \frac{1}{\nu} G$$

is a stochastic matrix. Note that, if G is reversible with respect to π, then so is P_ν, $\quad \forall \nu$. 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, $f \in L^2_0(\pi)$, and furthermore assume that f belongs to the range of the generator of the CTMC, G. 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 [3], $\hat{\mu}_{nt}$ 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 belongs to the domain of the generator and is such that: $Gg = f$.

In Proposition 2.4 of [3], Bhattacharya, proves that $v(f,G) > 0$ for all nonconstant (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.

5 Peskun ordering for continuous time Markov chains

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

Definition 5.1 *Let $G^1 = \{g^1_{ij}\}$ and $G^2 = \{g^2_{ij}\}$ be two CTMC. We say that G^1 dominates G^2 in the Peskun sense and write $G^1 \succeq_P G^2$ if*

$$g^1_{ij} \geq g^2_{ij}, \qquad \forall i \neq j$$

The following theorem mimics the one in [16]:

Theorem 5.1 *If $G^1 \succeq_P G^2$ and if the corresponding CTMC are reversible, then $G^2 - G^1$ is a positive operator.*

Proof:
Let

$$c_1 = \sup |g^1_{ij}|, \qquad c_2 = \sup |g^2_{ij}| \text{ and } \nu \geq \max(c_1, c_2).$$

Define

$$P^1_\nu = I + \frac{1}{\nu} G^1 \text{ and } P^2_\nu = I + \frac{1}{\nu} G^2.$$

Then we have that

$$G^1 = \nu(P^1_\nu - I) \text{ and } G^2 = \nu(P^2_\nu - I).$$

If $G^1 \succeq_P G^2$ it follows that $P^1_\nu \succeq_P P^2_\nu$ by the same definition of P^1_ν and P^2_ν. By Lemma 3 in Tierney [16] it then follows that $P^2_\nu - P^1_\nu$ is a positive operator. After observing that

$$G^2 - G^1 = \nu(P^2_\nu - P^1_\nu)$$

we immediately get the result we want. □

We are now ready to prove the main result of the paper:

Theorem 5.2 *If $G^1 \succeq_P G^2$ and if the corresponding CTMC are reversible, then*

$$v(f, G^1) \leq v(f, G^2)$$

for all functions f in the range of the generators, where $v(f,G^1)$ and $v(f,G^2)$ are the asymptotic variances of estimators $\hat{\mu}_n$ obtained by simulating the CTMC that have G^1 and G^2, respectively, as generators.

Proof:

We know (from [3]) that, for all functions f in the range of the generators, we have:

$$v(f,G^1) = -2 < f, g^1 > \text{ and } v(f,G^2) = -2 < f, g^2 >$$

where g^1 and g^2 belong to the domain of the generators and are such that:

$$G^1 g^1 = f, \text{ and } G^2 g^2 = f. \tag{1}$$

Therefore we have:

$$v(f,G^1) = -2 < G^1 g^1, g^1 > \text{ and } v(f,G^2) = -2 < G^2 g^2, g^2 > .$$

Define:

$$H_\beta = G^1 + \beta(G^2 - G^1) \text{ and } g_\lambda = g^1 + \lambda(g^2 - g^1)$$

where $0 \le \beta \le 1$ and $0 \le \lambda \le 1$. 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

$$h_\lambda(0) \ge h_\lambda(1), \qquad \forall \lambda \in [0,1]$$

and if we take $\lambda = 0$ we get:

$$h_0(0) = v(f,G^1) \text{ and } h_0(1) = -2 < G^2 g^1, g^1 > .$$

We thus have:

$$\begin{aligned} v(f,G^1) \ge& -2 < G^2 g^1, g^1 > \\ =& -2 < G^2(g^1 - g^2 + g^2), (g^1 - g^2 + g^2) > \\ =& -2 < G^2(g^1 - g^2), (g^1 - g^2) > \\ & -2 < G^2(g^1 - g^2), g^2 > \\ & -2 < G^2 g^2, (g^1 - g^2) > -2 < G^2 g^2, g^2 > \\ =& -2 < G^2(g^1 - g^2), (g^1 - g^2) > \\ & -2 < G^2 g^1, g^2 > +2 < G^2 g^2, g^2 > \\ & -2 < G^2 g^2, g^1 > +2 < G^2 g^2, g^2 > \\ & -2 < G^2 g^2, g^2 > \\ =& -2 < G^2(g^1 - g^2), (g^1 - g^2) > \\ & -2 < G^1 g^1, g^1 > +2 < G^2 g^2, g^2 > \\ & -2 < G^1 g^1, g^1 > +2 < G^2 g^2, g^2 > \\ & -2 < G^2 g^2, g^2 > \\ =& -2 < G^2(g^1 - g^2), (g^1 - g^2) > \\ & + v(f,G^1) - v(f,G^2) \\ & + v(f,G^1) - v(f,G^2) + v(f,G^2). \end{aligned} \tag{2}$$

The fourth equality in (2) follows from

$$< G^2 g^1, g^2 >=< g^1, G^2 g^2 >$$

because G is a self-adjoint operator. Furthermore

$$< g^1, G^2 g^2 >=< g^1, G^1 g^1 >$$

because of (1) and finally we have

$$< g^1, G^1 g^1 >=< G^1 g^1, g^1 > .$$

Also,

$$< G^2 g^2, g^1 >=< G^1 g^1, g^1 >$$

again because of (1). The last equality in (2) follows from

$$v(f, G^1) = -2 < G^1 g_1, g_1 >$$

and similarly for G^2. As a result we obtain:

$$v(f, G^2) - v(f, G^1) \geq -2 < G^2(g^1 - g^2), (g^1 - g^2) > \geq 0$$

and therefore:

$$v(f, G^2) \geq v(f, G^1).$$

□

6 Conclusions

For Markov chains taking values on finite state spaces, we have extended Peskun ordering from discrete to continuous time. The extension we propose has, potentially, many applications both in the MCMC context and in the TIEE framework. Indeed, in these settings, Peskun ordering for discrete time MC has been extensively used to give necessary and/or sufficient conditions for efficiency of the resulting MCMC and time-invariance estimators.

We plan to investigate the extension of Peskun ordering to continuous time general state space Markov chains, in further research.

References

[1] A.J. Baddeley, Time-invariance estimating equations. *Bernoulli*, 6(5):783–808, 2000.

[2] J. E. Besag. Comments on "Representations of knowledge in complex systems" by U. Grenander and M.I. Miller. *Journal of the Royal Statistical Society, Series B*, 56, 549–603, 1994.

[3] R. Bhattacharya, On the functional central limit theorem and the law of the iterated logarithm for Markov processes. *Z. Wahrsch. Verw. Gebiete*, 60(2): 185–201, 1982.

[4] L. J. Billera, and P. Diaconis, A geometric interpretation of the Metropolis algorithm. *Statistical Science*, 16(4): 335–339, 2001.

[5] J. D. Doll, P. J. Rossky, and H. L. Friedman. Brownian dynamics as smart Monte Carlo simulation. *Journal of Chemical Physics*, 69:4628–4633, 1978.

[6] V. P. Godambe and B.K. Kale, Estimating functions: an overview. In V.P. Godambe, editor, *Estimating Functions*, volume 7 of *Oxford Statistical Science series*, pp. 3–20. Oxford Science Publications, 1991.

[7] W. K. Hastings, *Monte Carlo Sampling Methods using Markov chains and their applications*, Biometrika, 57: 97-109, 1970.

[8] A. Mira, *Ordering and Improving MCMC performances*, Statistical Science, 16(4): 340–350, 2001.

[9] A. Mira and C. J. Geyer, *Ordering Monte Carlo Markov Chains*, School of Statistics University of Minnesota Tech. Report 632, 1999.

[10] A. Mira and A. Baddeley, *Performance of time-invariance estimators*, Department of Mathematics and Statistics, University of Western Australia (Perth, Australia), *Technical Report* No. 2001/15, 2001.

[11] P.H. Peskun, *Optimum Monte Carlo sampling using Markov Chains*, Biometrika, 60: 607–612, 1973.

[12] G. O. Roberts and R. L. Tweedie. Exponential convergence of Langevin diffusions and their discrete approximations. *Bernoulli*, 2: 341–364, 1996.

[13] G. O. Roberts and R. L. Tweedie. Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms. *Biometrika*, 83, 1996.

[14] O. Stramer and R. L. Tweedie. Langevin-tyoe models I: Diffusions with given stationary distributions and their discretizations. *Methodology and Computing in Applied Probability*, 1: 283–306, 1999.

[15] O. Stramer and R. L. Tweedie. Langevin-tyoe models II: Diffusions with given stationary distributions and their discretizations. *Methodology and Computing in Applied Probability*, 1: 307–328, 1999.

[16] L. Tierney, Markov chains for exploring posterior distributions. *Annals of Statistics*, 22: 1701–1762, 1994.

[17] L. Tierney, *A Note on Metropolis Hastings Kernels for General state spaces*, Ann. of App. Prob., 8: 1–9, 1998.