VELUT ARBOR ÆVO

Variance Bounding Markov Chains

by

Gareth O. Roberts*
Department of Mathematics and Statistics
Lancaster University

and

Jeffrey S. Rosenthal**
Department of Statistics
University of Toronto

Technical Report No. 0612 September 29, 2006

TECHNICAL REPORT SERIES

University of Toronto

Department of Statistics

Variance Bounding Markov Chains

by

Gareth O. Roberts* and Jeffrey S. Rosenthal**

(September, 2006.)

Abstract. We introduce a new property of Markov chains, called *variance bounding*. We prove that, for reversible chains at least, variance bounding is weaker than, but closely related to, geometric ergodicity. Furthermore, variance bounding is equivalent to the existence of central limit theorems for all L^2 functionals. Also, variance bounding (unlike geometric ergodicity) is preserved under the Peskun order. We close with some applications to Metropolis-Hastings algorithms.

1. Introduction.

Markov chain Monte Carlo (MCMC) algorithms are widely used in statistics, physics, and computer science. Measures of how good an MCMC algorithm is include quantitative bounds on convergence to stationarity (e.g. [27], [28], [11], [12]), qualitative convergence rates such as geometric ergodicity (e.g. [31], [30], [25], [23]), the existence of central limit theorems (e.g. [5], [31], [2], [8], [10]), and bounds on asymptotic variance of estimators (e.g. [32], [5], [19]).

In this paper, we introduce a new notion, *variance bounding*. Roughly, a Markov chain is variance bounding if the asymptotic variances for functionals with unit stationary variance are uniformly bounded (precise definitions are given below). We shall show that, for reversible chains at least, variance bounding is implied by geometric ergodicity, and conversely if P is variance bounding then $aI + (1-a)P$ is geometrically ergodic for all $0 < a < 1$. More importantly, we shall prove that a reversible Markov chain is variance bounding if and only if all L^2 functionals satisfy a central limit theorem, indicating that variance bounding is in some sense the "right" definition to use. We also prove that variance bounding is preserved under the Peskun partial ordering on Markov chains. Finally, applications to Metropolis-Hastings algorithms are presented.

*Department of Mathematics and Statistics, Fylde College, Lancaster University, Lancaster, LA1 4YF, England. Email: g.o.roberts@lancaster.ac.uk.

**Department of Statistics, University of Toronto, Toronto, Ontario, Canada M5S 3G3. Email: jeff@math.toronto.edu. Web: http://probability.ca/jeff/ Supported in part by NSERC of Canada.

Hence,

$$\sum_{n=0}^{\infty} \mathbf{Cov}(h(X_0), h(X_n)) \geq \sum_{i=1}^{\infty} 2^{-i} \sum_{n=0}^{\infty} (1-2^{-i})^n$$

$$= \sum_{i=1}^{\infty} 2^{-i}[1-(1-2^{-i})]^{-1} = \sum_{i=1}^{\infty} 2^{-i} 2^{i} = \sum_{i=1}^{\infty} (1) = \infty .$$

It then follows as in [21] that h does not satisfy a CLT for P. ∎

Proof of Theorem 6. Lemma 3 of Tierney [32] says that since $P_1 \succeq P_2$, therefore $P_2 - P_1$ is a positive operator. It follows that $\sup(\sigma(P_2)) \geq \sup(\sigma(P_1))$. Hence, using Theorem 12 twice, if P_2 is variance bounding, then $\sup(\sigma(P_2)) < 1$, so $\sup(\sigma(P_1)) < 1$, so P_1 is variance bounding. (Alternatively, by Theorem 4 of [32], $\mathbf{Var}(h, P_1) \leq \mathbf{Var}(h, P_2) \leq K\, \mathbf{Var}_\pi(h)$.) ∎

Remark. The above theorems have all been proven for reversible chains only. However, it seems likely that analogs of some of them (e.g. Theorem 1) carry over in some form to non-reversible chains, about which various facts about convergence are known (see e.g. [4], [15]). We leave this as an open problem for future work.

References

[1] J.R. Baxter and J.S. Rosenthal (1995), Rates of convergence for everywhere-positive Markov chains. Stat. Prob. Lett. **22**, 333–338.

[2] K.S. Chan and C.J. Geyer (1994), Discussion to reference [31]. Ann. Stat. **22**, 1747–1758.

[3] J.B. Conway (1985), A course in functional analysis. Springer, New York.

[4] J.A. Fill (1991), Eigenvalue bounds on convergence to stationarity for non-reversible Markov chains, with an application to the exclusion process. Ann. Appl. Prob. **1**, 62–87.

[5] C. Geyer (1992), Practical Markov chain Monte Carlo. Stat. Sci., Vol. **7**, No. **4**, 473-483.

[6] W.R. Gilks and G.O. Roberts (1995), Strategies for improving MCMC.R. Gilks). In *MCMC in Practice* (eds. W.R. Gilks, D.J. Spiegelhalter and S. Richardson), Chapman and Hall, 89–114.

[7] W.K. Hastings (1970), Monte Carlo sampling methods using Markov chains and their applications. Biometrika **57**, 97–109.

[8] J.P. Hobert, G.L. Jones, B. Presnell, and J.S. Rosenthal (2002), On the Applicability of Regenerative Simulation in Markov Chain Monte Carlo. Biometrika **89**, 731–743.

[9] I.A. Ibragimov and Y.V. Linnik (1971), Independent and Stationary Sequences of Random Variables. Wolters-Noordhoff, Groningen (English translation).

[10] G.L. Jones (2004), On the Markov chain central limit theorem. Prob. Surv. **1** (2004), 299–320.

[11] G.L. Jones and J.P. Hobert (2001), Honest exploration of intractable probability distributions via Markov chain Monte Carlo. Stat. Sci. **16**, 312–334.

[12] G.L. Jones and J.P. Hobert (2004), Sufficient burn-in for Gibbs samplers for a hierarchical random effects model. Ann. Stat. **32**, 784–817.

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

[14] G.F. Lawler and A.D. Sokal (1988), Bounds on the L^2 spectrum for Markov chains and Markov processes: A generalization of Cheeger's inequality. Trans. Amer. Math. Soc. **309**, 557–580.

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

[16] K.L. Mengersen and R.L. Tweedie (1996), Rates of convergence of the Hastings and Metropolis algorithms. Ann. Stat. **24**, 101–121.

[17] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller (1953), Equations of state calculations by fast computing machines. J. Chem. Phys. **21**, 1087–1091.

[18] S.P. Meyn and R.L. Tweedie (1993), Markov chains and stochastic stability. Springer-Verlag, London. Available at probability.ca/MT.

[19] A. Mira (2001), Ordering and improving the performance of Monte Carlo Markov chains. Stat. Sci. **16**, 340–350.

[20] P.H. Peskun (1973), Optimum Monte Carlo sampling using Markov chains. Biometrika **60**, 607–612.

[21] G.O. Roberts (1999), A note on acceptance rate criteria for CLTs for Metropolis-Hastings algorithms. J. Appl. Prob. **36**, 1210–1217.

[22] G.O. Roberts and J.S. Rosenthal (1997), Geometric ergodicity and hybrid Markov chains. Electronic Comm. Prob. **2**, Paper no. 2, 13–25.

[23] G.O. Roberts and J.S. Rosenthal (1998), Markov chain Monte Carlo: Some practical implications of theoretical results (with discussion). Canadian J. Stat. **26**, 5–31.

[24] G.O. Roberts and J.S. Rosenthal (2006), Examples of Adaptive MCMC. Preprint.

[25] G.O. Roberts and R.L. Tweedie (1996), Geometric Convergence and Central Limit Theorems for Multidimensional Hastings and Metropolis Algorithms. Biometrika **83**, 95–110.

[26] G. O. Roberts and R. L. Tweedie (1996), Exponential Convergence of Langevin Diffusions and Their Discrete Approximations. Bernoulli **2**, 341–364.

[27] J.S. Rosenthal (1995), Minorization conditions and convergence rates for Markov chain Monte Carlo. J. Amer. Stat. Assoc. **90**, 558–566.

[28] J.S. Rosenthal (2002), Quantitative convergence rates of Markov chains: A simple account. Electronic Comm. Prob. **7**, 123–128.

[29] J.S. Rosenthal (2003), Asymptotic Variance and Convergence Rates of Nearly-Periodic MCMC Algorithms. J. Amer. Stat. Assoc. **98**, 169–177.

[30] A.F.M. Smith and G.O. Roberts (1993), Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods (with discussion). J. Roy. Stat. Soc. Ser. B **55**, 3–24.

[31] L. Tierney (1994), Markov chains for exploring posterior distributions (with discussion). Ann. Stat. **22**, 1701–1762.

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