IBRAGIMOV, I. A. and LINNIK, YU. V. (1971). *Independent and Stationary Sequences of Random Variables.* Wolters-Noordhoff, Groningen (English translation).

LIU, J., WONG, W. H. and KONG, A. (1994). Correlation structure and convergence rate of the Gibbs sampler with various scans. *J. Roy. Statist. Soc. Ser. B.* To appear.

MEYN, S. P. and TWEEDIE, R. T. (1993). *Markov Chains and Stochastic Stability.* Springer, New York.

RUDIN, W. (1991). *Functional Analysis*, 2nd ed. McGraw-Hill, New York.

THOMPSON, E. A. and GUO, S. W. (1991). Evaluation of likelihood ratios for complex genetic models. *IMA J. Math. Appl. Med. Biol.* **8** 149–169.

TWEEDIE, R. L. (1983). The existence of moments for stationary Markov chains. *J. Appl. Probab.* **20** 191–196.

DEPARTMENT OF STATISTICS
AND ACTUARIAL SCIENCES
MACLEAN HALL 225K
UNIVERSITY OF IOWA
IOWA CITY, IOWA 52242

SCHOOL OF STATISTICS
UNIVERSITY OF MINNESOTA
270A VINCENT HALL
206 CHURCH STREET SE
MINNEAPOLIS, MINNESOTA 55455

# REJOINDER

LUKE TIERNEY[1]

*University of Minnesota*

I would like to thank the discussants for their thoughtful comments, and I would also like to thank the editors of *The Annals of Statistics* for this opportunity to respond. My comments are organized by topics addressed in the discussions.

**Irreducibility.** For simplicity, I wrote Theorem 1 and other results in my presentation to use as their key assumption that $P$ is irreducible with respect to $\pi$. In most applications this is relatively easy to verify, but as Doss points out there are cases where it is not. The theory in Nummelin used to develop these results is actually more general. In particular, it is sufficient to verify irreducibility with respect to *any* $\sigma$-finite measure. Thus the following generalization of Theorem 1 is available.

THEOREM 1*. *Suppose $P$ is $\varphi$-irreducible for some $\sigma$-finite measure $\varphi$ on $E$ and $\pi P = \pi$. Then $\varphi$ is absolutely continuous with respect to $\pi$, $P$ is $\pi$-irreducible, $P$ is positive recurrent and $\pi$ is the unique invariant distribution of $P$. If $P$ is also aperiodic, then, for $\pi$-almost all $x$,*

$$\|P^n(x, \cdot) - \pi(\cdot)\| \to 0,$$

*with $\|\cdot\|$ denoting the total variation distance. If $P$ is Harris recurrent, then the convergence occurs for all $x$.*

[1]Research supported in part by NSF Grant DMS-93-03557.

This can be proved along the same lines as the original Theorem 1. While there are some other minor differences, this captures the main difference between Theorem 1 and the result of Athreya, Doss and Sethuraman (1992) mentioned by Doss.

One of the implications of this theorem is that $\varphi$-irreducibility with respect to some $\sigma$-finite $\varphi$ together with $\pi P = \pi$ for a proper distribution $\pi$ implies $\pi$-irreducibility. Therefore the assumption of $\pi$-irreducibility can be similarly weakened to $\varphi$-irreducibility in all other results where $\pi$ is assumed to be an invariant distribution.

**Harris recurrence.** Being able to show that a recurrent chain is Harris recurrent is useful since it rules out the possibility of starting points from which convergence is not assured. It may be worth adding a bit more on the nature of such points, if they exist. If an irreducible, aperiodic chain with invariant distribution $\pi$ is not Harris recurrent and $N$ is the set where convergence fails in Theorem 1, then the probability of remaining forever in $N$ is positive for any starting point in $N$. Moreover, using Nummelin's terminology, $N$ is a $\pi$-null set and is dissipative. This means it is a countable union of transient sets, that is, sets for which the last exit time is finite with probability 1. Under fairly mild topological regularity conditions, this requires that to remain in $N$ forever the chain must drift to $\infty$ in the sense of visiting every compact set only finitely often [Meyn and Tweedie (1993), Theorem 9.0.2].

Chan and Geyer give a useful sufficient condition for a variable-at-a-time Metropolis scheme to be Harris recurrent. A key part of their argument can be extracted as a more general version of Corollary 1.

COROLLARY 1*. *Suppose $P$ is $\pi$-irreducible and $\pi P = \pi$. For each $n$ and each $x \in E$, let $(P_s^n(x, \cdot), P_a^n(x, \cdot))$ be the Lebesgue decomposition of $P$ with respect to $\pi$. That is, $P_a^n(x, \cdot)$ is absolutely continuous with respect to $\pi$, $P_s^n(x, \cdot)$ is singular with respect to $\pi$ and*

$$P^n(x, \cdot) = P_s^n(x, \cdot) + P_a^n(x, \cdot).$$

*If $P_a^n(x, E) \to 1$, that is, the total mass of the absolutely continuous part converges to 1, for all $x \in E$, then $P$ is Harris recurrent.*

PROOF. Let $h$ be a bounded harmonic function for $P$. The assumptions imply that $P$ is recurrent, hence $h = \pi h$ $\pi$-almost everywhere. Thus, for any $x \in E$,

$$h(x) = (Ph)(x) = \left(P^n h(x)\right) = \left(P_s^n h\right)(x) + P_a^n(x, E)\pi h.$$

The second term on the right converges to $\pi h$ by assumption, and, since $h$ is bounded, the first term converges to 0. Thus $h(x) = \pi h$ for all $x \in E$; that is, $h$ is a constant. □

In the case considered by Chan and Geyer, the absolutely continuous part of $P^n(x, \cdot)$ is its restriction to the set $S_{x, \emptyset}$, and their argument shows that the total mass of this component tends to 1.

**Central limit theorems.** Chan and Geyer and Robert presented a number of useful results on central limit theorems. Chan and Geyer's Theorem 2 fits in between the two results I gave as Theorems 4 and 5. The requirement on the convergence rate, geometric ergodicity, is stronger than ergodicity of degree 2 and weaker than uniform ergodicity; the requirement on the function $f$ is stronger than finite second moments under $\pi$ but weaker than boundedness.

The most powerful central limit theorem available for Markov chains is probably the Kipnis–Varadhan theorem mentioned by Chan and Geyer. Its requirement of reversibility is not very restrictive in the Markov chain Monte Carlo setting, since most chains can be modified to be reversible. Unfortunately, verifying finiteness of the asymptotic variance can be quite difficult.

Robert and Chan and Geyer provide a number of valuable references to the stationary process and mixing approach to central limit theorems. There are some special features of mixing conditions in a Markov chain setting that are worth taking into account; these are summarized in the paper of Bradley referenced by Robert.

I would like to address a few points in Robert's discussion. In subsection 1.3 Robert describes central limit results for $\varphi$-mixing Markov chains. In a Harris recurrent chain the condition of $\varphi$-mixing is equivalent to the Doeblin condition. This, in turn, is equivalent to uniform ergodicity [Meyn and Tweedie (1993), 16.2.3]. Thus this result is equivalent to the result I gave as Theorem 5, and verifying $\varphi$-mixing is equivalent to verifying a minorization condition. It can be done in some cases, and sometimes it can be enforced, for example, by using an appropriate hybrid chain, but I do not think it is quite as widely available as Robert suggests.

In Section 1.4 Robert gives a result for $\rho$-mixing Markov chains. Since $\rho$-mixing in a Markov chain must be exponentially fast if it occurs at all, a second moment condition that $f \in L^2(\pi)$ alone is sufficient. The additional conditions listed by Robert are needed in the triangular array context considered in Rosenblatt (1971), Section 7.4, but not when considering a single $f$.

**Regenerative analysis.** Robert and Besag both point out the usefulness of identifying regeneration times in a Markov chain. Such embedded renewal processes are useful theoretical tools—they form the basis of the ergodic theory developed in Nummelin (1984). They can also be of practical use in the context of regenerative simulation analysis. The tours between renewals are independent and identically distributed. This allows the use of statistical methods for i.i.d. samples and also allows for parallel simulation, since there is no need to actually run the tours sequentially.

But it is important to remember that any Harris recurrent chain will have an embedded renewal process, even a chain that converges very slowly. A chain for sampling a bimodal posterior distribution that only rarely moves from one mode to another will typically have embedded renewal processes where the interrenewal times are extremely heavy tailed; an example of such a process is given in Mykland, Tierney and Yu (1993). To be of use for a regenerative analysis, an embedded renewal process must have a reasonably high renewal rate in order to produce a large enough sample size for variance estimation, but it must also have a tail that is thin enough for the expected interrenewal time to be estimated accurately. For such a sequence to exist, the chain must mix reasonably rapidly. But even when such sequences exist, they are not necessarily easy to find. Mykland, Tierney and Yu (1993) argue for the use of hybrid chains designed to induce easily identified renewal sequences.

**Some implementation issues.** Doss gives a warning about the use of importance weights to adjust a Markov chain sample from one distribution toward another. The point is well taken. This problem exists in i.i.d. sampling and can typically only be exacerbated in a Markov chain context. Unbounded importance weights need to be used with caution. The idea of drawing a Markov chain sample from an easily sampled distribution and reweighting it toward a harder one is very tempting; it is unfortunate that the easily sampled distributions tend to be rather light tailed in many problems.

Besag makes an important point by suggesting that comparing sample sizes needed to achieve a certain level of accuracy can lead to misleading conclusions. In Besag's example, one sampler $P_1$ requires a larger sample size than another $P_2$ to achieve a given level of accuracy, but observations are cheaper to generate from $P_1$, so that the accuracy obtained from $P_1$ for a given amount of CPU time is actually greater than for $P_2$. Similar considerations apply in several other contexts. As one example, it has been pointed out that the common practice of using only every $k$th observation from a chain of length $kn$ to compute a sample path average of a function $f$ produces larger asymptotic standard errors than to average $f$ over the entire sample path of $kn$ observations. This suggests that subsampling is never a good idea. However, once the cost of evaluating $f$ is taken into account, the conclusion is much less clear. If $f$ is expensive to evaluate, or if there are many different $f$'s, then it may make sense to run a longer chain with subsampling in order to reduce the total cost for achieving a given level of accuracy. Variance reduction methods are yet another area that must be considered carefully. Even when they reduce variances, which need not always be the case for dependent sequences, the reduction may not warrant the additional cost of performing the required computations when comapred to the cost of running a longer chain.

**Final comments.** At the time my paper was written, the book by Nummelin (1984) was the most complete treatment of general state space Markov chain theory available. Since then the book of Meyn and Tweedie (1993) has become available. This monograph provides a more extensive and more accessible development of general state space Markov chain theory than Nummelin (1984). It also provides a number of useful results that take advantage of topological properties of transition operators that are often available in statistical problems.

Markov chain Monte Carlo is an important new, or perhaps not so new, tool for the analysis of complex statistical problems. The wealth of possible applications is well illustrated by the examples given by the discussants and by many other papers in the recent literature. I hope that this paper and discussion can make a contribution toward a better understanding of these methods and to the range of possible methods that are available.

## REFERENCES

MEYN, S. P. and TWEEDIE, R. L. (1993). *Markov Chains and Stochastic Stability*. Springer, New York.

MYKLAND, P. A., TIERNEY, L. and YU, B. (1995). Regeneration in Markov chain samplers. *J. Amer. Statist. Assoc.* To appear.

ROSENBLATT, M. (1971). *Markov Processes. Structure and Asymptotic Behavior*. Springer, New York.

SCHOOL OF STATISTICS
UNIVERSITY OF MINNESOTA
270 VINCENT HALL
206 CHURCH STREET
MINNEAPOLIS, MINNESOTA 55455