CHARLOT, F. (1991). Régénération, chaînes de Markov et files d'attente. *Séminaire de Mathématique de Rouen* **2** 143–154.
DAVYDOV, Y. A. (1973). Mixing conditions for Markov chains. *Theory Probab. Appl.* **18** 312–328.
DIEBOLT, J. and ROBERT, C. P. (1992). Discussion of Smith and Roberts and Besag and Green's papers. *J. Roy. Statist. Soc. Ser. B* To appear.
DIEBOLT, J. and ROBERT, C. P. (1994). Estimation of finite mixture distributions by Bayesian sampling. *J. Roy. Statist. Soc. Ser. B.* **56** 363–376.
GELFAND, A. E., SMITH, A. F. M. and LEE, T. M. (1992). Bayesian analysis of constrained parameter and truncated data problems. *J. Amer. Statist. Assoc.* **87** 523–532.
HEITJAN, D. F. and RUBIN, D. B. (1991). Ignorability and coarse data. *Ann. Statist.* **19** 2244–2253.
MALINOVSKII, V. K. (1986). Limit theorems for Harris Markov chains. I. *Theory Probab. Appl.* **31** 269–285.
MALINOVSKII, V. K. (1989). Limit theorems for Harris Markov chains. II. *Theory Probab. Appl.* **34** 252–265.
MEYN, S. P. and TWEEDIE, R. L. (1992). Stability of Markovian processes. I. *Adv. in Appl. Probab.* **24** 542–574.
PELIGRAD, M. (1986). Recent advances in the central limit theorem and its weak invariance principle for mixing sequences of random variables. In *Dependence in Probability and Statistics* (E. Ebberlein and M. Taqqu, eds.) 192–223. Birkhäuser, Boston.
ROBERT, C. P., CELEUX, G. and DIEBOLT, J. (1993). Bayesian estimation of hidden Markov chains: a stochastic implementation. *Statist. Probab. Lett.* **16** 77–83.
ROSENBLATT, M. (1971). *Markov Processes. Structure and Asymptotic Behavior*. Springer, New York.
TANNER, M. and WONG, W. (1987). The calculation of posterior distributions (with discussion). *J. Amer. Statist. Assoc.* **82** 528–550.

LABORATOIRE DE STATISTIQUE THÉORETIQUE
ET APPLIQUÉE
CNRS–URA 1321
UNIVERSITÉ PIERRE ET MARIE CURIE,
PARIS VI
TOUR 45–55, BOÎTE 158
4, PLACE JUSSIEU
75252 PARIS CEDEX 05
FRANCE

KUNG SIK CHAN[1] AND CHARLES J. GEYER[2]

University of Iowa and University of Minnesota

We congratulate Luke Tierney for this paper, which even before its appearance has done a valuable service in clarifying both theory and practice in this important area. For example, the discussion of combining strategies in Section 2.4 helped researchers break away from pure Gibbs sampling in 1991; it was, for example, part of the reasoning that lead to the "Metropolis-coupled" scheme of Geyer (1991) mentioned at the end of Section 2.3.3.

Harris Recurrence. The discussion of Harris recurrence in Section 3.1 has been very helpful. Harris recurrence essentially says that there is no measure-

[1]Supported in part by NSF Grant DMS-91-18626.
[2]Supported in part by NSF Grant DMS-90-07833.

theoretic pathology. Hence it seems that it should always be present in any sampler that can be run on a computer, even in the continuous approximation that treats the computer's real numbers as if they were the analyst's real numbers.

The main point of Harris recurrence is that asymptotics do not then depend on the starting distribution because of the "split chain" construction [Nummelin (1984), Proposition 4.8], something that was pointed out to us by Tierney. More precisely, the argument of the paragraph at the bottom of the page 135 in Nummelin (1984) shows that the first regeneration time is a.s. finite (for any initial distribution) hence the sum up to the first regeneration time is also a.s. finite and negligible when divided by $\sqrt{n}$.

Typically irreducibility implies Harris recurrence. Corollaries 1 and 2 of Tierney's paper show this for Gibbs samplers and non-hybrid Metropolis–Hastings algorithms, respectively. The following theorem shows this for the other important case, variable-at-a-time Metropolis–Hastings algorithms. These are hybrid algorithms in the terminology of Tierney's Section 2.4 that update one coordinate at a time using a Metropolis–Hastings update as in the original example of Metropolis, Rosenbluth, Rosenbluth, Teller and Teller (1953).

Consider a variable-at-a-time Metropolis–Hastings algorithm on a subset of $\mathbb{R}^d$. We are trying to simulate from the distribution with density proportional to $h(\mathbf{x}) = h(x_1, \dots, x_d)$ with respect to Lebesgue measure on $\mathbb{R}^d$. The sampler proceedes through the variables in order updating one at a time. It updates the ith variable by proposing a new value x_i^* from a univariate density with respect to Lebesgue measure $q_i(\mathbf{x}, x_i^*)$ that depends on the current position $\mathbf{x} = (x_1, \dots, x_d)$ and accepts the proposal moving to the new position $\mathbf{x}^* = (x_1, \dots, x_{i-1}, x_i^*, x_{i+1}, \dots, x_d)$ with probability

$$a(\mathbf{x}, x_i^*) = \min\left(1, \frac{h(\mathbf{x}^*)}{h(\mathbf{x})} \frac{q_i(\mathbf{x}^*, x_i)}{q_i(\mathbf{x}, x_i^*)}\right) \tag{1}$$

and otherwise remains at the point $\mathbf{x}$, that is, the usual Metropolis–Hastings update described in Tierney's Section 2.3.1 applied to the variable x_i.

We want to find conditions under which π-irreducibility of such a chain implies Harris recurrence. The conditions involve the "conditional samplers" for subsets of the d variables. These are the chains obtained by fixing k of the variables at some possible value and updating the remaining $d - k$ variables one at a time. The variables being updated are updated using the same basic update step as in the original sampler for the unconditional distribution; the proposals are generated from $q_i(\mathbf{x}, \cdot)$ and the acceptance probability is given by (1). In order for these conditional samplers to be well defined we need to assume that if $I = \{i_1, \dots, i_k\}$ and $I^c = \{i_{k+1}, \dots, i_d\}$ that the function

$$g(x_{i_{k+1}}, \dots, x_{i_d}) = h((x_1, \dots, x_d)) \tag{2}$$

is not almost everywhere zero and is integrable with respect to Lebesgue measure on $\mathbb{R}^{d-k}$ for any possible values of $x_{i_1}, \dots, x_{i_k}$. Then (2) is proportional to a version of the conditional density of $(X_{i_{k+1}},\dots,X_{i_d})$ given $X_{i_1},\dots,X_{i_k}$.

THEOREM 1. *A variable-at-a-time Metropolis–Hastings algorithm on $\mathbb{R}^d$ with proposal distributions that are absolutely continuous with respect to Lebesgue measure is Harris recurrent if all of the conditional samplers (including the unconditional sampler which conditions on the empty set of variables $I = \emptyset$) are irreducible for any values of the fixed variables, the irreducibility measure and stationary distribution having density* (2).

PROOF. Let $L(\mathbf{x},A)$ denote the probability that the chain started at a point $\mathbf{x}$ ever hits a set A. By Theorem 9.1.5 in Meyn and Tweedie (1993) the state space can be decomposed as $H \cup N$ where H is a maximal absorbing Harris set and N is π-null and transient, where "maximal absorbing" means that $L(\mathbf{x},H) = 1$ implies $\mathbf{x} \in H$. The chain is Harris if and only if N is empty. Suppose to get a contradiction that N is nonempty and fix $\mathbf{x} \in N$. For each subset I of $\{1,\dots,d\}$ consider the set

$$S_{\mathbf{x},I} = \{y : x_i = y_i,\ i \in I \text{ and } x_i \neq y_i,\ i \notin I\}.$$

By the continuity of the proposal distributions, once the chain leaves $S_{\mathbf{x},I}$ is never returns, since from outside $S_{\mathbf{x},I}$ there is probability zero of a proposal in $S_{\mathbf{x},I}$. Hence for a chain $\mathbf{X}_n$ started at $\mathbf{X}_0 = \mathbf{x}$, the number of coordinates in which $\mathbf{X}_n$ differs from $\mathbf{x}$ is almost surely nondecreasing in time. Moreover, the distribution of $\mathbf{X}_n$ conditional on being in $S_{\mathbf{x},I}$ is almost surely absolutely continuous with respect to Lebesgue measure on $S_{\mathbf{x},I}$ by Fubini and induction on the dimension of the set I. If the chain eventually hits

$$S_{\mathbf{x},\emptyset} = \{y : y_i \neq x_i, \text{ for all } i\}$$

with probability 1, then it hits H with probability 1, which would imply $\mathbf{x} \in H$ contrary to the assumption. Hence the only way the chain can fail to be Harris is if for some I the probability of the chain eventually leaving $S_{\mathbf{x},I}$ is strictly less than 1.

Let $a(z)$ denote the probability of leaving $S_{\mathbf{x},I}$ in one step of the Markov chain given the current position $z \in S_{\mathbf{x},I}$. The probability of never leaving $S_{\mathbf{x},I}$ is given by

$$\mathbb{E}\prod_{i=1}^{\infty}\left[1 - a(\mathbf{Z}_i)\right] \tag{3}$$

where $\{\mathbf{Z}_i\}$ is the conditional chain obtained by starting at the first point where $\{\mathbf{X}_i\}$ enters $S_{\mathbf{x},I}$ and then only updating the variables x_i for $i \notin I$. This chain is irreducible by assumption, hence by Birkhoff's ergodic theorem

$$\frac{1}{n}\sum_{i=1}^{n} a(\mathbf{Z}_i) \xrightarrow{\text{a.s.}} \mathbb{E}a(\mathbf{Z}) = \alpha \tag{4}$$

where the expectation here is taken over the stationary distribution of the conditional chain. Note that we do not need to assume that this chain is Harris, since Birkhoff's theorem implies a.s. convergence from almost all starting points and distribution of $\mathbf{Z}_1$ is absolutely continuous with respect to Lebesgue measure on $S_{\mathbf{x},I}$ and hence gives probability 0 to the bad null set of starting points for which convergence fails. Also note that $\alpha > 0$ because otherwise we would have $a = 0$ almost everywhere and the chain could never leave $S_{\mathbf{x},I}$ for almost all starting points in $S_{\mathbf{x},I}$ and hence could not be irreducible.

Because of $\log(1+x) \le x$ and (4),

$$\prod_{i=1}^{\infty} [1 - a(\mathbf{Z}_i)] = \exp\left(\sum_{i=1}^{n} \log[1 - a(\mathbf{Z}_i)] \right) \le \exp\left(- \sum_{i=1}^{n} a(\mathbf{Z}_i) \right) \to 0.$$

Hence the integrand in (3) is almost surely 0, and the expectation is 0. Hence it is impossible for the chain to remain forever on a lower dimensional subset and the chain is Harris recurrent. □

Central limit theorems. Our main comments apply to central limit theorems. Before delving into the complexities of general state spaces, we should mention one elementary case: there is always a CLT if the chain is irreducible and the state space is finite. This follows from Tierney's Theorem 5, because a finite state space is small, hence the chain is uniformly ergodic. There seem to be several approaches to establishing a CLT for a Markov chain on a general state space. Roughly these can be divided into two categories. One uses Markov chain theory. The other uses functional analysis and martingales.

The Markov chain approach, based on the book of Nummelin (1984) and related work, is followed by Tierney to arrive at his Theorems 4 and 5. We would like to make readers aware of two other theorems along the same lines, but more useful. To state the theorems, we need some preliminary discussion of the variance in the CLT. Let $L_0^2(\pi)$ denote the subspace of $L^2(\pi)$ orthogonal to the constants

$$L_0^2(\pi) = \left\{ g \in L^2(\pi) \colon \int g \, d\pi = 0 \right\}.$$

Then the *transition operator* of the chain is defined by

$$(Tg)(x) = \int g(y) P(x, dy), \qquad g \in L_0^2(\pi).$$

For any $f \in L_0^2(\pi)$ let

$$\gamma_k(f) = (f, T^k f) = \int f(x) f(y) P^k(x, dy) \pi(dx)$$

denote the lag k autocovariance of the functional f for the stationary chain. Let $\bar{f}_n$ be as in Tierney's subsection 3.3. Then for the stationary chain

$$n \operatorname{var} \bar{f}_n = \gamma_0(f) + 2 \sum_{k=1}^{n-1} \frac{n-k}{n} \gamma_k(f).$$

If the sequence $\bar{f}_n$ has a CLT and is uniformly integrable, then the variance in the CLT is

$$\lim_{n \to \infty} n \operatorname{var} \bar{f}_n = \lim_{n \to \infty} \left(\gamma_0(f) + 2 \sum_{k=1}^{n-1} \frac{n-k}{n} \gamma_k(f) \right), \tag{5}$$

that is, the Cesáro sum of the autocovariances. If the sequence of autocovariances is summable, then this agrees with the ordinary sum

$$\gamma_0(f) + 2 \sum_{k=1}^{\infty} \gamma_k(f). \tag{6}$$

Tierney's subsection 3.2 discusses the types of ergodicity. In order of increasing strength, these are ergodicity, ergodicity of degree 2, geometric ergodicity and uniform ergodicity. The consequences of ergodicity of degree 2 and uniform ergodicity are given in Tierney's Theorems 4 and 5. The following theorem gives the consequence of geometric ergodicity.

THEOREM 2. *For a geometrically ergodic chain, suppose that* $\int |f|^{2+\varepsilon} d\pi < +\infty$ *for some* $\varepsilon > 0$. *Then* $\sqrt{n}(\bar{f}_n - \pi f)$ *converges weakly to a normal distribution with mean* 0 *and variance* $\sigma(f)^2$ *given by* (6) *for any initial distribution.*

PROOF. Integrating the definition of geometric ergodicity in Tierney's subsection 3.2 gives the following: for all Borel sets A and B

$$\left| \int_B \left[P^n(x,A) - \pi(A) \right] \pi(dx) \right| = \left| \Pr(X_n \in A \text{ and } X_0 \in B) - \pi(A)\pi(B) \right| \le \|M\|_1 r^n.$$

The supremum over all A and B is the so-called strong mixing coefficient $\alpha(n)$. Hence geometric ergodicity implies exponentially fast α-mixing. The result then follows from a well-known stationary process central limit theorem [Ibragimov and Linnik (1971), Theorem 18.5.3]. Since the chain is Harris recurrent, the asymptotics do not depend on the starting distribution. □

This gives us a conclusion almost as strong as Tierney's Theorem 5 but with a much weaker hypothesis. Typically, a chain with a continuous, unbounded state space will not be uniformly ergodic, but many such chains are geometrically ergodic. Moreover, geometric ergodicity can often be verified using a drift condition (Tierney's Proposition 1).

The next theorem uses only the weakest type of ergodicity and establishes a CLT only for a single function. Recall the definition of *small set* from Tierney's subsection 3.2.

THEOREM 3 [Chan (1993)]. *For a Harris ergodic chain, suppose that there is a nonnegative function g in* $L^2(\pi)$ *and a small set C such that*:

(a) $g(x)$ *and* $E[g(X_{k+1}) | X_k = x]$ *are both bounded on* C,

(b) $g(x) \geq |f(x) - \pi f|$ *for all* $x \in C$ *and*
(c) $g(x) \geq E[g(X_{k+1}) \mid X_k = x] + |f(x) - \pi f|$ *for all* $x \notin C$.

Then $\sqrt{n}(\bar{f}_n - \pi f)$ *converges weakly to a normal distribution with mean* 0 *and variance* $\sigma(f)^2$ *given by* (5) *for any initial distribution.*

It should be pointed out that the statement of the theorem in Chan (1993) contains an error. The formula for $\sigma(f)^2$ there is incorrect but is correct as stated here with the variances referring to the stationary chain.

It is not always known which functions are in $L^2(\pi)$. Then the following theorem is useful in establishing whether a function f is integrable.

THEOREM 4 [Tweedie (1983)]. *For a Harris ergodic chain, suppose that* f *and* g *are nonnegative measurable functions and for some set* A *with* $\pi(A) > 0$ *and* $\int_A f\,d\pi < \infty$:

(a) $g(x) \geq f(x)$ *for all* $x \in A^c$,
(b) $\int_{A^c} g(y)P(x, dy) \leq g(x) - f(x)$ *for all* $x \in A^c$ *and*
(c) $\sup_{x \in A} \int_{A^c} g(y)P(x, dy) < \infty$.

Then $\int f\,d\pi < \infty$.

Note that A is not required to be a small set, though, of course, any small set A does satisfy $\pi(A) > 0$.

To finish the Markov chain approach, we should mention the theorem in Nummelin (1984), which may be sharper than Theorem 3 (it cannot be weaker since Theorem 3 is proved as a consequence of Nummelin's theorem). The condition for Nummelin's theorem, however, seems difficult to verify.

The martingale–functional analytic approach seems to give the sharpest results for reversible chains. Requiring reversibility causes no essential loss of generality. Since the basic Metropolis–Hastings update step is reversible, in order to produce a reversible sampler it is only necessary that the basic update steps be combined in a reversible way. This observation has been made by several people, the earliest perhaps being Besag (1986) where it is attributed to Peter Green.

For a reversible chain the transition operator is a self-adjoint operator on $L_0^2(\pi)$ that is obviously a contraction, $\|T\| \leq 1$ (since conditioning reduces variance). The spectrum of T is the set of complex numbers λ such that $T - \lambda I$ is not invertible. Because T is self-adjoint, the spectrum is real, and from the spectral radius formula, it is bounded by $\|T\|$. Hence the spectrum is a closed subset of $[-1, 1]$. The spectral theory of bounded operators on Hilbert spaces [Rudin (1991), page 316 ff.] associates with T a spectral measure E mapping Borel subsets of the spectrum to self-adjoint projections on $L_0^2(\pi)$. A point λ in the spectrum is an atom $E(\{\lambda\}) \neq 0$ if and only if λ is an eigenvalue. The following lemma shows, for a Harris ergodic chain, 1 and -1 cannot be eigenvalues, and hence the spectral measure is concentrated on the open interval $(-1, 1)$.

LEMMA 5. *For a Harris ergodic chain, any eigenvalues of the transition operator have modulus strictly less than* 1.

PROOF. For any $f \in L_0^2(\pi)$ with $\|f\| = 1$, there is a bounded function g such that $\|f - g\| \leq \frac{1}{3}$. By Cauchy–Schwarz, $\pi(g) \leq \|f - g\| \leq \frac{1}{3}$. Now

$$\|T^m f\| \leq \pi(g) + \|T^m g - \pi(g)\| + \|T^m(f - g)\|$$

and the first and third terms on the right-hand side are less than $\frac{1}{3}$. The middle term is

$$\int \left\{ \int g(y)\big[P^m(x, dy) - \pi(dy)\big] \right\}^2 \pi(dx) \leq \|g\|_\infty^2 \int \|P^m(x, \cdot) - \pi(\cdot)\|^2 \pi(dx),$$

where on the right-hand side $\|\cdot\|$ indicates total variation norm. Since the integrand converges to 0 (Tierney's Theorem 1), this term converges to 0 by dominated convergence. Hence $\|T^m f\|$ is eventually strictly less than 1, and it cannot satisfy $Tf = \lambda f$ with $|\lambda| = 1$. □

Spectral theory simplifies questions about the variance in the CLT. It follows from dominated and monotone convergence that the limit in (5) always exists and is equal to the ordinary sum (6) and also to the spectral integral

$$\sigma(f)^2 = \int \frac{1 + \lambda}{1 - \lambda} E_f(d\lambda) \tag{7}$$

though the limit (and the integral) may be $+\infty$, where E_f is the positive Borel measure defined by

$$E_f(B) = \big(f, E(B)f\big).$$

If (7) is finite, there is a CLT.

THEOREM 6 [Kipnis and Varadhan (1986)]. *For a reversible Harris ergodic chain, if* (7) *is finite, then* $\sqrt{n}(\bar{f}_n - \pi f)$ *converges weakly to a normal distribution with mean* 0 *and variance* $\sigma(f)^2$ *given by* (5), (6) *or* (7) *for any initial distribution.*

Kipnis and Varadhan give two conditions, equivalent to (7) being finite. The first is that f lies in the range of $(I - T)^{1/2}$, and the second is that there be a constant C such that

$$(f, g) \leq C\sqrt{\big(g, (I - T)g\big)}, \qquad g \in L_0^2(\pi). \tag{8}$$

It is not clear how useful this is in practice. We know of no examples where this has been used to establish a CLT for a Markov chain Monte Carlo sampler. The main benefit of reversibility and the Kipnis–Varadhan CLT to date seems to have been in simplifying theoretical questions. The calculation of $\sigma(f)^2$ by the spectral integral (7) is one example. Other examples are given by Geyer (1992).

Let us now turn from consideration of the existence of a CLT for a single function to the question of whether every square-integrable function has a CLT. Kipnis and Varadhan did not show that finiteness of (7) was necessary for a CLT (this appears to be an open question), but we can say that the variance of $\bar{f}_n$ is $O(n^{-1})$ for all $f \in L_0^2(\pi)$ if and only if $(I - T)^{1/2}$ is a surjection, in which case $I - T$ is also surjective. From Lemma 5 the only $L^2(\pi)$ solutions of $Tf = f$ are constant, so $I - T$ is also one to one on $L_0^2(\pi)$. That is, we have a CLT for all square-integrable functions if $I - T$ is invertible, and if $I - T$ is not invertible, there is some $f \in L_0^2(\pi)$ for which (5) is not finite.

For this criterion, reversibility is superfluous.

THEOREM 7 [Gordin and Lifšic (1978)]. *For a Harris ergodic chain, suppose that f is in the range of $I - T$, say $f = (I - T)g$. Then $\sqrt{n}(\bar{f}_n - \pi f)$ converges weakly to a normal distribution with mean* 0 *and variance given by*

$$\sigma(f)^2 = \|g\|^2 - \|Tg\|^2 \tag{9}$$

for any initial distribution.

Gordin and Lifšic remark that the theorem could conceivably hold with $\sigma(f)^2 = 0$ in which case $\sqrt{n}(\bar{f}_n - \pi f)$ converges in probability to 0 (the limit is degenerate). This cannot happen in the reversible case. If the chain is aperiodic, -1 is not an eigenvalue, and the formula (7) for $\sigma(f)^2$ is strictly positive.

It is not clear how the invertibility of $I - T$ can be easily checked. It is, of course, equivalent to

$$0 < \inf_{g \in L_0^2(\pi)} \frac{\|(I - T)g\|^2}{\|g\|^2}. \tag{10}$$

In the reversible case, it is enough to check that the right-hand side of (8) is bounded away from 0 when g is bounded away from 0, that is,

$$0 < \inf_{g \in L_0^2(\pi)} \frac{\left\|(I - T)^{1/2} g\right\|^2}{\|g\|^2} = \inf_{g \in L_0^2(\pi)} \frac{\big(g, (I - T)g\big)}{\|g\|^2}. \tag{11}$$

But neither condition seems easy to check in general.

To finish the functional analytic-martingale approach, we should mention the theorem of Tóth (1986) which extends the methods of Kipnis and Varadhan to the nonreversible case, but at the cost of imposing a condition that is difficult to verify. In the reversible case Tóth's theorem reduces to that of Kipnis and Varadhan.

Schervish and Carlin (1992) and Liu, Wong and Kong (1994) have followed another path to geometric convergence, establishing that T is Hilbert–Schmidt.

This also implies a CLT. Any Hilbert–Schmidt operator is compact and hence has a countable spectrum having no point of accumulation other than 0. By Lemma 5 a Harris ergodic chain has no eigenvalues of modulus 1, hence if T is compact, $\|T\| < 1$. Thus $I - T$ is an invertible operator on $L_0^2(\pi)$. Hence T Hilbert–Schmidt implies T compact implies $\|T\| < 1$ implies $I - T$ invertible implies a CLT by Theorem 7.

It is clear from the theorem that a Hilbert–Schmidt transition operator is not essential, since not all compact operators are Hilbert–Schmidt and $I - T$ can be invertible when T is not compact. In general, a Metropolis–Hastings sampler for a continuous distribution will not have a compact transition operator unless, like the Gibbs sampler, all proposals are accepted with probability 1. From Tierney's Section 2.3 we see that the general form of the transition operator is

$$(Tf)(x) = r(x)f(x) + \int f(y)p(x,y)\mu(dy)$$

so $T = M + S$, where $(Mf)(x)$ is the first term on the right-hand side and $(Sf)(x)$ is the second term. If M is a compact operator, then $r(x)$ is 0 almost surely except on a countable set, and hence is 0 almost surely if π has no atoms [cf. Exercise 5, Section 2.4 in Conway (1985)]. If M is noncompact, then T and S cannot both be compact, and T can be compact only if there is some rather strange cancellation occurring.

Examples. With all of this theory, there is some question as to whether there is much hope of obtaining a CLT for practical problems. In our somewhat limited experience, this does seem to be fairly easy using drift conditions. Geyer and Møller (1994) proposed a Metropolis sampler for spatial point processes and obtained a CLT for one particular sampler, an unconditional Strauss process, using Theorem 3, though while this discussion was being written it was discovered that the drift condition of Tierney's Proposition 1 and our Theorem 2 imply that the sampler is geometrically ergodic and hence has a CLT for all functions in $L^{2+\varepsilon}(\pi)$.

A somewhat different example is the Gibbs sampler for the mixed model in genetics [Thompson and Guo (1991) and Guo and Thompson (1994)]. A quantitative trait y that depends additively on unobserved discrete genotypes g, continuous genotypes z (thought of as approximating the combined effect of a large number of genes of small effect) and environmental influence e is assumed to have the form

$$y = \mu(g) + z + e,$$

where g, z and e are independent, g has a finite number of possible values and z and e both have nondegenerate (multivariate) normal distributions with mean 0. We want to sample from the conditional distribution of g and z given y (everything that follows is conditional on y but this will be suppressed in the notation) in the case where z and g have high dimension and it is convenient to Gibbs sample them one component at a time. First we consider the case when the variables are sampled in fixed order, sampling the z's in a block and then the g's.

Let g, z and g', z' denote consecutive steps of the stationary chain. Then g' is updated from a kernel $k_1(g' \mid g, z')$, which is discrete, and z' is updated from a kernel $k_2(z' \mid g, z)$, which is normal with mean $a(g) + Bz$ and variance Λ. Let $f(z \mid g)$ denote the stationary distribution of z given g, which is normal with mean $m(g)$ and variance Σ, and $p(g)$ the marginal stationary distribution of g. Here B, Λ and Σ are constant matrices, and Λ and Σ are positive definite by the nondegeneracy assumption.

What must be shown to establish that the transition operator is Hilbert–Schmidt [Schervish and Carlin (1992)] is that

$$\sum_{g,g'} \int \int \left[\frac{k_1(g' \mid g, z') k_2(z' \mid g, z)}{f(z' \mid g') p(g')} \right]^2 f(z' \mid g') p(g') f(z \mid g) p(g) \, dz \, dz'$$

is finite. This happens if each term of the sum for which $p(g)$ and $p(g')$ are nonzero is finite, which since $k_1 \le 1$ happens if

(12) $$\int \int \frac{k_2(z' \mid g', z)^2 f(z \mid g)}{f(z' \mid g')} \, dz \, dz'$$

is finite for all g and g'. But this is finite for the same reasons that a Gibbs sampler for the multivariate normal is Hilbert–Schmidt [Schervish and Carlin (1992) and Schervish (personal communication)]. Since every basic update step of the Gibbs sampler preserves the stationary distribution, z' and g have the same distribution as z and g. Hence

$$\Sigma = \operatorname{var}(z' \mid g) = \Lambda + B \Sigma B^T.$$

Now $\int k_2(z' \mid g', z)^2 f(z \mid g) \, dz$ is seen to be proportional to a normal density with mean $a(g') + Bm(g)$ and variance $\frac{1}{2}\Lambda + B\Sigma B^T = \Sigma - \frac{1}{2}\Lambda$. Hence, when this is divided by $f(z' \mid g')$ and integrated with respect to z', the integrand is proportional to a normal density with precision matrix $\Sigma^{-1} - (\Sigma - \frac{1}{2}\Lambda)^{-1}$, hence integrable (this matrix is positive definite because $0 < A < B$ implies $B^{-1} < A^{-1}$ for any matrices A and B).

For a general scan (12) becomes

(13) $$\int \int \frac{\prod_i k_{2,i}(z'_i \mid g, g', z, z')^2 f(z \mid g)}{f(z' \mid g')} \, dz \, dz',$$

where $k_{2,i}$ is the one-dimensional normal density of the Gibbs update for z'_i. It is clear that (13) is integrable if (12) is because all the normal kernels have the same variances and dependence on other components of z, only the dependence on the g's differs and this only changes means. The result extends to a random sequence scan (each variable is updated exactly once per scan, but the order is random), because then the overall transition operator is a convex combination of the operators for each of the scan orders, and a linear combination of Hilbert–Schmidt operators is Hilbert–Schmidt.

Conclusion. Using the methods discussed here and in Tierney's subsection 3.3, it should be possible in many cases to establish whether the CLT holds for a Markov chain sampler.

The strongest conclusion is that $I - T$ is invertible, which implies a CLT for all functionals in $L^2(\pi)$. One would like to establish this whenever it is true and disprove it whenever it is false. Our methods for doing so, however, are rather weak. $I - T$ is invertible if T is Hilbert–Schmidt [Schervish and Carlin (1992) and Liu, Wong and Kong (1994)], but this applies only to Gibbs samplers not to general Metropolis–Hastings samplers and even for Gibbs samplers will not always work. The invertibility of $I - T$ could be disproved by finding a sequence of functions g that drive the right-hand side of (10) or (11) to 0, but that also seems difficult.

The next strongest conclusion is that the chain is geometrically ergodic, which implies a CLT for all functionals in $L^{2+\varepsilon}(\pi)$. This is established using the drift condition argument (Tierney's Proposition 1 and our Theorem 2). It is not clear to us how one establishes that a chain is not geometrically ergodic.

For a sampler that is believed not to be geometrically ergodic, the CLT can still be established for specific functionals using the drift condition argument [Chan (1993), repeated as Theorem 3 here]. Another approach would be to attempt to verify (8) directly (reversible chains only).

It should also be mentioned, just for the sake of completeness, that the normalized partial sums of a square-integrable functional can converge in law to a stable distribution with exponent strictly between 1 and 2 (of course, then the normalizing constants cannot be $n^{-1/2}$). Davydov (1973) gives examples. It is also possible for the CLT to hold with normalizing constants other than $n^{-1/2}$ (for a trivial example, consider i.i.d. sampling from a distribution without finite variance but still in the domain of attraction of the normal distribution).

REFERENCES

BESAG, J. (1986). The statistical analysis of dirty pictures (with discussion). *J. Roy. Statist. Soc. Ser. B* **48** 259–302.

CHAN, K.-S. (1993). On the central limit theorem for an ergodic Markov chain. *Stochastic Process. Appl.* **47** 113–117.

CONWAY, J. B. (1985). *A Course in Functional Analysis.* Springer, New York.

DAVYDOV, YU. A. (1973). Mixing conditions for Markov chain. *Theory Probab. Appl.* **18** 312–328 (English translation).

GEYER, C. J. and MØLLER, J. (1994). Simulation and likelihood inference for spatial point processes. *Scand. J. Statist.*, To appear.

GORDIN, M. I. and LIFŠIC, B. A. (1978). The central limit theorem for stationary Markov processes. *Soviet Math. Dokl.* **19** 392–394 (English translation).

GUO, S. W. and THOMPSON, E. A. (1994). Monte Carlo estimation of mixed models for large complex pedigrees. *Biometrics* **50** 417–432.

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 π. 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* σ-finite measure. Thus the following generalization of Theorem 1 is available.

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