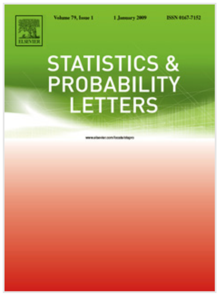
A new proof of convergence of MCMC via the ergodic theorem

Søren Asmussen, Peter W. Glynn

Accepted Manuscript

A new proof of convergence of MCMC via the ergodic theorem

Søren Asmussen, Peter W. Glynn



PII: S0167-7152(11)00175-1
DOI: 10.1016/j.spl.2011.05.004
Reference: STAPRO 6001

To appear in: *Statistics and Probability Letters*

Received date: 5 July 2010
Revised date: 7 May 2011
Accepted date: 9 May 2011

Please cite this article as: Asmussen, S., Glynn, P.W., A new proof of convergence of MCMC via the ergodic theorem. *Statistics and Probability Letters* (2011), doi:10.1016/j.spl.2011.05.004

A new proof of convergence of MCMC via the ergodic theorem

Søren Asmussen[a,*], Peter W. Glynn[b]

[a] Department of Mathematical Sciences, Aarhus University, Ny Munkegade, DK-8000 Aarhus C, Denmark
[b] Department of Management Science and Engineering, Stanford University, Stanford, CA 94305-4026, USA

Abstract

A key result underlying the theory of MCMC is that any η-irreducible Markov chain having a transition density with respect to η and possessing a stationary distribution π is automatically positive Harris recurrent. This paper provides a short self-contained proof of this fact using the ergodic theorem in its standard form as the most advanced tool.

Key words: Markov chain Monte Carlo, Harris recurrence, η-irreducibility

1. Introduction

The use of Markov chain Monte Carlo methods (MCMC) has become a fundamental numerical tool in modern statistics, as well as in the study of many stochastic models arising in mathematical physics; see Asmussen and Glynn (2007), Gilks et al. (1996), Kendall et al. (2005), and Robert and Casella (2004), for example. When applying this idea, one constructs a Markov chain $X = (X_n : n \geq 0)$ having a prescribed stationary distribution π. By simulating a trajectory of X over $\{0, 1, \ldots, n-1\}$, the hope is that the time-average $n^{-1} \sum_{j=0}^{n-1} f(X_j)$ will converge to $\pi f \triangleq \int_S f(x)\pi(\mathrm{d}x)$ where S is the state space. Thus, MCMC permits one to numerically investigate the distribution π.

If X is an irreducible discrete state space Markov chain with stationary distri-

*Corresponding author
Email addresses: asmus@imf.au.dk (Søren Asmussen)

bution π, it is well known that for all x

$$\frac{1}{n}\sum_{j=0}^{n-1} f(X_j) \to \pi f \quad \mathbb{P}_x\text{-a.s.} \tag{1}$$

as $n \to \infty$ for each $f : S \to \mathbb{R}_+$, where $\mathbb{P}_x(\cdot) \triangleq \mathbb{P}(\cdot|X_0 = x)$ for $x \in S$. Many statistical applications of MCMC involve, however, distributions π that are continuous. A central theoretical question in MCMC is therefore the extension of the above result to a general state space. Some key references for this are Tierney (1994), Roberts and Rosenthal (2004), Roberts and Rosenthal (2004) and Robert and Casella (2004); see also Chan and Geyer (1994). As in the discrete state space setting, some notion of irreducibility is required. The Markov chain X is said to be *η-irreducible* if η is a non-trivial (reference) measure for which $\eta(B) > 0$ implies that $K(x, B) > 0$ for all $x \in S$, where

$$K(x, \mathrm{d}y) \triangleq \sum_{n=1}^{\infty} 2^{-n}\mathbb{P}_x(X_n \in \mathrm{d}y)$$

for $x, y \in S$. Typically, the key step for the general state space MCMC setting is to establish results of the following spirit (note the 'π-a.a x' rather than 'all x'!):

Theorem 1. *Assume that X is an S-valued Markov chain having a stationary distribution π and being η-irreducible for some η. Then for $f : S \to \mathbb{R}_+$,*

$$\frac{1}{n}\sum_{j=0}^{n-1} f(X_j) \to \pi f \quad \mathbb{P}_x\text{-a.s.}$$

as $n \to \infty$, for π-a.a. $x \in S$.

Unfortunately, the existing proofs tend to rely on referencing a substantial body of advanced Markov chain theory (in particular, material from Nummelin (1984) or Meyn and Tweedie (1993), and/or harmonic functions, and/or decomposition into recurrent/transient classes which is a far more complicated topic when the state space is general rather than discrete). The contribution of this paper is to offer an alternative (short) proof that as background knowledge requires only graduate probability, with the most advanced result being the ergodic theorem in its standard form. In our view, the advantages of this approach is that it is self-contained, that short proofs of the ergodic theorem can be found in standard intermediate-level textbooks such as Breiman (1968), Durrett (2010), and that the ergodic theorem has a much wider scope and range of applications than the specialized Markov chain results referred to above.

Specializing to the MCMC setting, η-irreducibility is not quite strong enough to guarantee (1) (see Example 1 at the end of Section 3). However, (1) is known to hold subject to minor additional conditions. In particular:

Theorem 2. *Assume that X is an S-valued Markov chain satisfying*

$$\mathbb{P}_x(X_1 \in \mathrm{d}y) = p(x,y)\eta(\mathrm{d}y) \tag{2}$$

for each $(x,y) \in S\times S$ and some measure η and some jointly measurable transition density $p : S \times S \to \mathbb{R}_+$. If X has a stationary distribution π, is η-irreducible and $f : S \to \mathbb{R}_+$, then

$$\frac{1}{n}\sum_{j=0}^{n-1} f(X_j) \to \pi f \quad \mathbb{P}_x\text{-a.s.}$$

as $n \to \infty$, for all $x \in S$.

Note that if S is discrete and η assigns positive mass to each state, (2) is immediate. The η-irreducibility of X is then equivalent to the standard notion of irreducibility in the discrete setting. Note also that by specializing to functions f that are indicators, it follows that whenever $\pi(B) > 0$, $\mathbb{P}_x(X_n \in B$ infinitely often$) = 1$ for $x \in S$. This is precisely the definition of Harris recurrence. Thus, Theorem 2 implies that X is a positive recurrent Harris chain.

A nice feature of Theorem 2 is that it does not require construction of any Lyapunov functions to establish positive Harris recurrence. The assumed existence of a stationary distribution, which is natural in MCMC applications, dispenses with this need.

The proof of Theorem 1 is given in Section 2. Typical MCMC algorithms do not satisfy (2). Rather, the one-step transition kernel can often be written in the form

$$\mathbb{P}_x(X_1 \in \mathrm{d}y) = (1 - a(x))\delta_x(\mathrm{d}y) + a(x,y)q(x,y)\eta(\mathrm{d}y), \tag{3}$$

where $\delta_x(\cdot)$ is a unit mass at x and $a(x)$ and $a(x,y)$ are non-negative. For example, this arises in the context of the Metropolis-Hastings sampler with $q(x,y)$ being the proposal density at y for a given x and $a(x,y)$ representing the probability of accepting proposal y. The key result is then:

Corollary 1. *Assume that X is an S-valued Markov chain satisfying* (3) *for which $a(x) > 0$ for each $x \in S$. If X is η-irreducible and has a stationary distribution π, then X is a positive recurrent Harris chain.*

The proof is a simple translation of Theorem 2 and can be found in previous papers as well, but for the sake of self-containedness, it is given in Section 3 together with the proof of Theorem 2.

2. Proof of Theorem 1

Let $\mathbb{P}_\pi(\cdot) \triangleq \int_S \pi(\mathrm{d}x)\mathbb{P}_x(\cdot)$ and let $\mathbb{E}_\pi(\cdot)$ be the expectation operator corresponding to $\mathbb{P}_\pi$.

A: First, suppose that $\eta = \pi$. The ergodic theorem implies that for each $f : S \to \mathbb{R}_+$,

$$\frac{1}{n}\sum_{i=0}^{n-1} f(X_i) \to Z \quad \mathbb{P}_\pi\text{-a.s.}$$

as $n \to \infty$, where $Z = \mathbb{E}_\pi[f(X_0)|\mathscr{I}]$ and $\mathscr{I}$ is the invariant σ-field. We first establish that $Z = \mathbb{E}_\pi f(X_0)$.

Note that we may assume that $\mathbb{E}_\pi f(X_0) < \infty$ (for if this is not the case, we may work instead with $f_n = f \wedge n$ and then send $n \to \infty$). Put $h(x) = \mathbb{E}_x Z$. Note that

$$\mathbb{E}_\pi[Z|X_0, \ldots, X_n] \to Z \quad \mathbb{P}_\pi\text{-a.s.}$$

as $n \to \infty$. Since Z is invariant, the left-hand side equals $h(X_n)$ $\mathbb{P}_\pi$-a.s., so we may conclude that

$$h(X_n) \to Z \quad \mathbb{P}_\pi\text{-a.s.} \tag{4}$$

as $n \to \infty$. Suppose that $Z \neq \mathbb{E}_\pi f(X_0)$ $\mathbb{P}_\pi$-a.s.. Then, there exists $a, b \in \mathbb{R}_+$ (with $a < b$) for which $\pi(A_1) > 0$ and $\pi(A_2) > 0$, where $A_1 = \{x : h(x) \le a\}$ and $A_2 = \{x : h(x) \ge b\}$.

Let $\tau_1, \tau_2, \ldots$ be iid Geometric($\frac{1}{2}$) random variables (rv's) independent of X, and set $T_0 = 0$ and $T_n = \tau_1 + \cdots + \tau_n$ for $n \ge 1$. Note that $(X_{T_n} : n \ge 0)$ is an S-valued Markov chain having one-step transition kernel K and stationary distribution π. Then,

$$\frac{1}{n}\sum_{i=1}^{n} I(X_{T_i} \in A_1) = \frac{1}{n}\sum_{i=1}^{n}[I(X_{T_i} \in A_1) - \mathbb{P}_\pi(X_{T_i} \in A_1|X_{T_{i-1}})] + \frac{1}{n}\sum_{i=0}^{n-1} K(X_{T_i}, A_1) \quad \mathbb{P}_\pi\text{-a.s.} \tag{5}$$

Of course, since the rv's in [ ] form a bounded sequence of martingale differences,

$$\frac{1}{n}\sum_{i=1}^{n}[I(X_{T_i} \in A_1) - \mathbb{P}_\pi(X_{T_i} \in A_1|X_{T_{i-1}})] \to 0 \tag{6}$$

$\mathbb{P}_\pi$-a.s. as $n \to \infty$. Also, because $(X_{T_i} : i \geq 0)$ is a stationary sequence under $\mathbb{P}_\pi$, a second application of the ergodic theorem ensures that

$$\frac{1}{n}\sum_{i=0}^{n-1} K(X_{T_i}, A_1) \to \mathbb{E}_\pi[K(X_0, A_1)|\mathscr{I}] \quad \mathbb{P}_\pi\text{-a.s.} \tag{7}$$

as $n \to \infty$. Since $\pi(A_1) > 0$, the π-irreducibility of X guarantees that $K(x, A_1) > 0$ for each $x \in S$. Consequently, $\mathbb{E}_\pi[K(X_0, A_1)|\mathscr{I}] > 0$ $\mathbb{P}_\pi$-a.s., so that (5), (6), and (7) yield the conclusion

$$\varliminf_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} I(X_{T_i} \in A_1) > 0 \quad \mathbb{P}_\pi\text{-a.s.}$$

and hence $\mathbb{P}_\pi(h(X_n) \leq a$ infinitely often$) = 1$. Similarly, we conclude that $\mathbb{P}_\pi(h(X_n) \geq b$ infinitely often$) = 1$. Since this contradicts (4), it must be that $Z = \mathbb{E}_\pi f(X_0)$. Consequently, $\mathbb{P}_\pi(N) = 0$, where $N = \{n^{-1}\sum_{i=0}^{n-1} f(X_i) \nrightarrow \pi f$ as $n \to \infty\}$.

B: We now extend to a general η. According to step A, it suffices to show that if X is η-irreducible with an invariant probability π, then X is π-irreducible.

Step 1: Note that if X is η-irreducible, it is q-irreducible, where $q = \eta K$ with $K(x, \mathrm{d}y) = \sum_{n=1}^{\infty} 2^{-n} P^n(x, \mathrm{d}y)$.

Step 2: Suppose that $\pi(B) > 0$. Then, the ergodic theorem implies that the event that $X_n \in B$ i.o. has positive π-measure.

Step 3: We want to prove that $q(B) > 0$ (for then $K(x, B)$ is clearly positive for all x because of the q-irreducibility). Suppose, by way of contradiction, that $q(B) = 0$. Thus, for η-a.a. x, $\mathbb{P}_x(X_n \in B$ for some $n \geq 0) = 0$.

Step 4: Note that the η-irreducibility implies that for each x,

$$K(x, \mathrm{d}y) = r(x, y)\eta(\mathrm{d}y) + (1 - r(x))M(x, \mathrm{d}y) \;\text{ where }\; \int r(x, \mathrm{d}y)\eta(\mathrm{d}y) > 0\,.$$

We can use this to construct a randomized stopping time L for which $w(x) = \mathbb{P}_x(L < \infty) > 0$ for all x and $\mathbb{P}_x(X_L \in \cdot)$ is absolutely continuous w.r.t η.

Step 5: Let C_n be the indicator that $\theta_n L < \infty$ where $\theta_n L$ is L applied to the shifted post-n path of X. So, $\mathbb{E}[C_n \,|\, X_0, ...X_n] = w(X_n)$. As in the proof of Theorem 2,

$$\frac{1}{n}\sum_{i=1}^{n}(C_i - w(X_{i-1}))$$

is an average of bounded MG differences so it converges a.s. to zero under $\mathbb{P}_\pi$. On the other hand, the time-average of the $w(X_i)$ converges $\mathbb{P}_\pi$-a.s. to $\mathbb{E}[w(X_0)\,|\,\mathcal{I}]$. But $w(X_0)$ is a positive r.v., so it follows that the time-average of the C_i has $\mathbb{P}_\pi$-a.s. a positive $\liminf$. So, this implies that for π-a.a. x, the time-average of the C_i has $\mathbb{P}_x$-a.s. a positive $\liminf$. Consequently, for π-a.a. x,

$$\mathbb{P}_x(\text{there exists a finite randomized stopping time } N = N(x) \text{ s.t. } \theta_N L < \infty) \;=\; 1\,.$$

Step 6: For such an x, $L^* = N + \theta_N L < \infty$ $\mathbb{P}_x$-a.s. But

$$\mathbb{P}_x(X_n \in B \text{ for some } \; n \geq L^*) \;=\; \int \mathbb{P}_x(X_{L^*} \in \mathrm{d}y)\mathbb{P}_y(X_n \in B \text{ for some } n \geq 0) \;=\; 0$$

(by virtue of the fact that X_{L^*} is absolutely continuous w.r.t. η and Step 3). So, $\mathbb{P}_x(X_n \in B \text{ i.o.}) = 0$ for all π-a.a. x. This contradicts Step 2.

3. Remaining proofs and an example

To prove Theorem 2, it remains to prove that $\mathbb{P}_x(N) = 0$ for $x \in S$. Let $C = \{x : \mathbb{P}_x(N) > 0\}$. If $\eta(C) > 0$, the η-irreducibility of X ensures that $K(x, C) > 0$ for all $x \in S$. Since π is a stationary for K, an immediate implication would be

$$\pi(C) \;=\; (\pi K)(C) \;=\; \int_S K(x, C)\,\pi(\mathrm{d}x) \;>\; 0\,,$$

contradicting the fact that $\mathbb{P}_\pi(N) = 0$. So, $\eta(C) = 0$. But (2) implies that

$$\mathbb{P}_x(N) = \int_S p(x, y)\mathbb{P}_y(N)\eta(\mathrm{d}y) = \int_C p(x, y)\mathbb{P}_y(N)\eta(\mathrm{d}y) = 0$$

for each $x \in S$, proving the result.

Now consider the MCMC case (3). Put $\beta_0 = 0$ and let β_n be the first acceptance time after β_{n-1}. If $a(x) > 0$ for each $x \in S$, then $(X_{\beta_n} : n \geq 0)$ is itself a well-defined S-valued Markov chain. Note that the transition kernel is given by

$$\mathbb{P}_x(X_{\beta_1} \in \mathrm{d}y) = \frac{a(x, y)q(x, y)}{a(x)}\eta(\mathrm{d}y),$$

so that $(X_{\beta_n} : n \geq 0)$ has a one-step transition density with respect to η. Furthermore, it is trivial that $(X_n : n \geq 0)$ is η-irreducible if and only if $(X_{\beta_n} : n \geq 0)$ is η-irreducible. Finally, note that if π is a stationary distribution for X, then $\tilde{\pi}$ defined by $\tilde{\pi}(\mathrm{d}y) = a(y)\pi(\mathrm{d}y)/\int_S \pi(\mathrm{d}z)a(z)$ is a probability and

$$\begin{aligned}\int_S \tilde{\pi}(\mathrm{d}x)\mathbb{P}_x(X_{\beta_1} \in \mathrm{d}y) &= \frac{\int_S \pi(\mathrm{d}x)a(x,y)q(x,y)\eta(\mathrm{d}y)}{\int_S \pi(\mathrm{d}z)a(z)} \\ &= \frac{\int_S \pi(\mathrm{d}x)\left(\mathbb{P}_x(X_1 \in \mathrm{d}y) - (1-a(x))\delta_x(\mathrm{d}y)\right)}{\int_S \pi(\mathrm{d}z)a(z)} \\ &= \frac{(\pi(\mathrm{d}y) - (1-a(y))\pi(\mathrm{d}y))}{\int_S \pi(\mathrm{d}z)a(z)} \\ &= \tilde{\pi}(\mathrm{d}y),\end{aligned}$$

so that $\tilde{\pi}$ is stationary for $(X_{\beta_n} : n \geq 0)$. It follows that if $(X_n : n \geq 0)$ is η-irreducible and possesses a stationary distribution, Theorem 2 applies to $(X_{\beta_n} : n \geq 0)$, establishing the positive Harris recurrence of $(X_{\beta_n} : n \geq 0)$. It is then immediate that $(X_n : n \geq 0)$ is positive Harris recurrent.

Example 1. Let $S = \mathbb{R}$ and let X evolve as an autoregressive process with Gaussian increments in $\mathbb{R}/\mathbb{N}$, i.e. $X_{n+1} = \alpha X_n + \varepsilon_n$ for $X_n \notin \mathbb{N}$ where $0 < \alpha < 1$ and $\varepsilon_0, \varepsilon_1, \ldots$ are i.i.d. and standard Gaussian. If $X_n = x \in \mathbb{N}$, let $X_{n+1} = x + 1$ w.p. $1 - a_x$ and $X_{n+1} = \varepsilon_n$ w.p. $a_x > 0$. The Gaussian distribution π with mean 0 and variance $1/(1-\alpha^2)$ is stationary and the chain is π-irreducible. However, if $\sum_1^\infty a_x < \infty$, the Borel-Cantelli lemma implies that starting from any $X_0 = x \in \mathbb{N}$ there is positive probability that $X_n \in \mathbb{N}$ for all n and so (1) fails for such x.

Examples of similar spirit appear elsewhere, e.g. Example 3 of Roberts & Rosenthal (2004).

Acknowledgements

The authors would like to thank the Isaac Newton Institute for Mathematical Sciences at Cambridge University for its support of a joint visit during which this paper was completed. We are also grateful to the referee for comments that helped improve the presentation of the paper.

References

Asmussen, S., Glynn, P. W., 2007. Stochastic Simulation: Algorithms and Analysis. Springer-Verlag.

Breiman, L., 1968. Probability. Addison-Wesley.

Chan, K., Geyer, C., 1994. Comment on ”Markov chains for exploring posterior distributions” by L. Tierney. Ann. Statist 22, 1747–1758.

Durrett, R., 2010. Probability: Theory and Examples, 4th Edition. Cambridge University Press.

Gilks, W., Richardson, S., Spiegelhalter, D., 1996. Markov Chain Monte Carlo in Practice. Chapman & Hall.

Kendall, W., Liang, F., Wang, J.-S. (Eds.), 2005. Markov Chain Monte Carlo: Innovations and Applications. World Scientific.

Meyn, S., Tweedie, R., 1993. Markov Chains and Stochastic Stability. Springer-Verlag.

Nummelin, E., 1984. General Irreducible Markov Chains and Nonnegative Operatore. Cambridge University Press.

Robert, C., Casella, G., 2004. Monte Carlo Statistical Methods, 2nd Edition. Springer-Verlag.

Roberts, G., Rosenthal, J., 2004. General state space Markov chains and MCMC algorithms. Probability Surveys 1, 20–71.

Tierney, L., 1994. Markov chains for exploring posterior distributions (with discussion). Ann. Statist. 22, 1701–1786.