THIELE CENTRE
FOR APPLIED MATHEMATICS IN NATURAL SCIENCE

Harris Recurrence and MCMC: A Simplified Approach

Søren Asmussen and Peter W. Glynn

10

Research Report

No. 06 | July 2010

Harris Recurrence and MCMC: A Simplified Approach

Søren Asmussen[a,*] and 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 is automatically positive Harris recurrent. This paper provides a short self-contained proof of this fact.

Keywords: 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 $[0, n)$, the hope is that the time-average $n^{-1}\sum_{j=0}^{n-1} f(X_j)$ will converge to $\pi f \triangleq \int_S f(x)\pi(\mathrm{d}x)$. Thus, MCMC permits one to numerically investigate the distribution π.

If $\pi = (\pi(x) : x \in S)$ is a discrete distribution, X is a finite or countably infinite state Markov chain. In general, if the dynamics of X are not chosen carefully, S may be reducible and/or contain transient states, in which case the time-averages may not converge to πf. However, if X is an irreducible discrete state space Markov chain with stationary distribution π, then it is well known that

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

as $n \to \infty$ for each $f : S \to \mathbb{R}_+$, where $\mathbb{P}_x(\cdot) \triangleq \mathbb{P}(\cdot | X_0 = x)$ for $x \in S$. Furthermore, it is known that such a Markov chain automatically contains embedded positive

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

recurrent regenerative structure, so that a path of X consists of independent and identically distributed (iid) cycles; see, for example, Asmussen (2003).

Many statistical applications of MCMC involve distributions π that are continuous. A central theoretical question in MCMC is therefore the extension of the above result to general state space. 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 $x, y \in S$. This irreducibility concept is not quite strong enough to guarantee (1). Instead, we further require existence of a jointly measurable transition density $p : S \times S \to \mathbb{R}_+$ for which

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

for $x, y \in S$. 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.

Our main contribution in this paper is to provide a simple and self-contained proof of the following known result.

Theorem 1 *Assume that X is an S-valued Markov chain satisfying (2). If X is η-irreducible having a stationary distribution π 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 each $x \in S$.

Note that by specializing to functions f that are indicators, it follows that whenever $\eta(B) > 0$, $\mathbb{P}_x(X_n \in B \text{ infinitely often}) = 1$ for $x \in S$. This is precisely the definition of Harris recurrence. Thus, Theorem 1 implies that X is a positive recurrent Harris chain. When the σ-algebra on S is countable generated, it is known that positive recurrent Harris chains contain embedded positive recurrent regenerative structure; see, for example, Athreya and Ney (1978) and Nummelin (1978). Hence, Theorem 1 is a natural generalization of discrete state space theory to the general state space context.

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. Put $\beta_0 = 0$ and $\beta_n = \inf\{j > \beta_{n-1} : X_j \neq X_{j-1}\}$, so that $(X_{\beta_n} : n \geq 0)$ is the Markov chain X sampled at acceptance epochs. 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 1 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. Hence the corollary below is a consequence of Theorem 1.

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.*

Results essentially equivalent to Corollary 1 appear in Tierney (1994), Roberts and Rosenthal (2004), Roberts and Rosenthal (2006), and Robert and Casella (2004). However, the proofs tend to rely on referencing a substantial body of advanced Markov chain theory (in particular, Nummelin (1984)). By contrast, the alternative (short) proof that we offer here is self-contained and as background knowledge requires only graduate probability (with the most advanced result being the ergodic theorem in its standard form).

A nice feature of these recurrence results is that they do 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.

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 $\mathbb{P}_\pi$. 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 $A_1 = \{x : h(x) \leq a\}$ and $A_2 = \{x : h(x) \geq b\}$.

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 $n \geq 1$. Note that $(X_{T_n} : n \geq 0)$ is an S-valued Markov chain having one-step transition kernel K and stationary distribution π. Then,

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

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.s., a second application of the ergodic theorem ensures that

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

as $n \to \infty$. The stationarity of π for X implies that $\pi(B) = \int_S \pi(\mathrm{d}x) \int_B p(x, y)\eta(\mathrm{d}y)$, so that π is absolutely continuous with respect to η. It follows that $\eta(A_1) > 0$. The η-irreducibility of X then 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 $\mathbb{P}_\pi(h(X_n) \leq a \text{ infinitely often}) = 1$. Similarly, we conclude that $\mathbb{P}_\pi(h(X_n) \geq b \text{ infinitely often}) = 1$. 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 $n \to \infty\}$.

The proof is complete if we can show 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 $x \in S$. By virtue of the fact that π is a stationary distribution of K, an immediate implication would be that $\pi(C) > 0$, 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 $x \in S$, finishing the proof.

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.

References

Asmussen, S., 2003. Applied Probability and Queues, 2nd Edition. Springer-Verlag.

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

Athreya, K. B., Ney, P., 1978. A new approach to the limit theory of recurrent Markov chains. Trans. Amer. Math. Soc. 245, 493–501.

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.

Nummelin, E., 1978. A splitting technique for Harris recurrent chains. Z. Wahrscheinlichkeitstheorie Verw. Geb. 43, 309–318.

Nummelin, E., 1984. General Irreducible Markov Chains and Non-Negative Operators. 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.

Roberts, G., Rosenthal, J., 2006. Harris recurrence of Metropolis-within-Gibbs and trans-dimensional Markov chains. Ann. Appl. Probab. 16, 2123–2139.

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