On Markov Chains with Continuous State Space

December 1997

by Persi Diaconis, Mathematics Department, Cornell University, Ithaca, NY 14853
and David Freedman, Statistics Department, U.C. Berkeley, CA 94720

Abstract

In this expository paper, we prove the following theorem, which may be of some use in studying Markov chain Monte Carlo methods like hit and run, the Metropolis algorithm, or the Gibbs sampler. Suppose a discrete-time Markov chain is aperiodic, irreducible, and there is a stationary probability distribution. Then from almost all starting points the distribution of the chain at time n converges in norm to the stationary distribution. This known theorem is a special case of more general results due to Doeblin, and the paper concludes with a brief review of the literature.

1. Introduction

This paper is largely expository. We develop in a fairly self-contained way part of Doeblin's theory for Markov chains in discrete time with a continuous state space, which provides a framework for demonstrating convergence of algorithms like "hit and run," the Gibbs sampler, and the Metropolis algorithm. Generally, we follow Orey (1971) and Harris (1955).

Let $(\mathcal{X}, \mathcal{B})$ be a measurable space; we assume that $\mathcal{B}$ is countably generated. Let $P(x, A)$ be a "kernel." In other words, $A \to P(x, A)$ is a probability on $\mathcal{B}$ for each $x \in \mathcal{X}$, while $x \to P(x, A)$ is $\mathcal{B}$-measurable for each $A \in \mathcal{B}$. If μ is a probability and $P = P(x, A)$ is a kernel, the probability μP is defined by the relation $\mu P(A) = \int P(x, A)\, \mu(dx)$. If $\mu P = \mu$, then μ is "invariant" or "stationary." If P and Q are kernels, then $(PQ)(x, A) = \int P(x, dy) Q(y, A)$, the integral being over $y \in \mathcal{X}$; thus, PQ is a kernel. Multiplication of kernels is associative. If μ, ν are probabilities, then $\|\mu - \nu\| = \sup_A |\mu(A) - \nu(A)| \leq 1$.

We will consider a Markov chain starting from $x \in \mathcal{X}$ at time 0 and moving according to P; the position at time n has distribution $P^n(x, \bullet)$. If x is itself chosen at random from a stationary μ, the Markov chain will be a stationary stochastic process with values in $\mathcal{X}$. Doeblin's theory of these processes involves φ, an auxiliary probability on $(\mathcal{X}, \mathcal{B})$.

Definitions. (i) P is "φ-irreducible" iff for all $x \in \mathcal{X}$ and all $A \in \mathcal{B}$ with $\varphi(A) > 0$, there is a positive integer $n = n_{xA}$ such that $P^n(x, A) > 0$.

(ii) P is "strongly φ-irreducible" iff for all $x \in \mathcal{X}$ and all $A \in \mathcal{B}$ with $\varphi(A) > 0$, there is a positive integer $n = n_{xA}$ such that $P^m(x, A) > 0$ for all $m \geq n$.

(iii) P is "φ-recurrent" iff for all $x \in \mathcal{X}$ and all $A \in \mathcal{B}$ with $\varphi(A) > 0$, a Markov chain which starts from x at time 0 hits A at some positive time, a.s.; of course, this time is random.

Usually, φ is taken as a σ-finite measure rather than a probability; the extra generality is more apparent than real; φ-recurrence is often called "Harris recurrence" in the literature. Typically, the chain will have some periodic structure that needs to be taken into account, because there are "cyclically moving classes." "Strong irreducibility" precludes such cycles. The condition does not seem to be standard, but it is useful; indeed, for some applications of interest, it is fairly easy to establish the existence of a stationary probability π, relative to which the chain is strongly irreducible. In that setting, the main result is as follows.

Theorem 1. Let P be a kernel on a separable measurable space, and let π be a stationary probability. Suppose P is strongly π-irreducible. Then there exists a measurable set $\mathcal{X}_1$ such that (i) $\pi(\mathcal{X}_1) = 1$, (ii) $P(x, \mathcal{X}_1) = 1$ for all $x \in \mathcal{X}_1$, (iii) P retracted to $\mathcal{X}_1$ is π-irreducible, and (iv) for any probability μ on $(\mathcal{X}_1, \mathcal{B})$,

$$\|\mu P^n - \pi\| \to 0 \text{ as } n \to \infty.$$

This theorem will be proved in Section 2; in other expositions, strong irreducibility is derived from other conditions that look weaker. In Diaconis and Freedman (1997), we proved a theorem like Theorem 1, assuming absolute continuity rather than strong irreducibility. That condition is stronger, but gives more precise results, by identifying the exceptional null set of starting points from which convergence to stationarity does not obtain. For ease of reference, several intermediate results are presented in both papers.

The next theorem establishes the existence and uniqueness of a stationary distribution, with a geometric rate of convergence to stationarity; the conditions of the theorem are rather strong. This theorem is well known; there is a self-contained proof in Diaconis and Freedman (1997). Under the conditions of Theorem 1, geometric convergence need not obtain; see Athreya, Doss and Sethuraman (1996) or Diaconis and Freedman (1997).

Theorem 2. Let P be a kernel, and φ an auxiliary probability. Suppose that for all $x \in \mathcal{X}$ and $A \in \mathcal{B}$,

$$P(x, A) \geq \epsilon\varphi(A).$$

(a) There is a unique stationary probability μ.
(b) $\mu \geq \epsilon\varphi$.
(c) $\|P^n(x, \bullet) - \mu\| \leq (1 - \epsilon)^n$ for all $x \in \mathcal{X}$.

Informally, the condition in Theorem 2 puts a positive lower bound on the transition densities. In general, this lower bound does not exist (and neither do the densities). However, a lower bound can be constructed on a measurable rectangle of positive measure—if not for P, then for a power P^k; part of the trick is showing that P^k has an absolutely continuous component with respect to φ, at least under suitable regularity conditions. That will all be done in Section 2. From this perspective, the two key ideas in the proof of Theorem 1 will be (i) deriving irreducibility from recurrence, and (ii) getting a positive lower bound on the k-step transition density, on a measurable rectangle of positive φ-measure. After that, Theorem 2 can be proved by a coupling argument. This section closes with Example 1 to indicate the difficulty in getting lower bounds, and Example 2 to indicate why retraction to $\mathcal{X}_1$ is needed in Theorem 1; the construction for Example 1 is discussed in Diaconis and Freedman (1997); Example 2 is quite easy.

Example 1. Let $\mathcal{X} = [0, 1]$ and let $\mathcal{B}$ be the σ-field of Borel sets in $\mathcal{X}$. Let φ be Lebesgue measure on $\mathcal{B}$. There is a positive measurable function f on $\mathcal{X}^2$ with the following property. If (i) $\delta > 0$, (ii) A and B are Borel sets, and (iii) $A \times B \subset \{f \geq \delta\}$ up to a φ^2-null set, then $\varphi(A) \times \varphi(B) = 0$.

Example 2. Let $\mathcal{X}_1$ be a finite set, and P a stochastic matrix on $\mathcal{X}_1$, with all entries strictly positive. Under the circumstances, there is a unique stationary probability π, and P is π-recurrent. Adjoin a sequence of states $1, 2, \ldots$, each with π-probability 0, and the following transition rules: $i \to i + 1$ with probability $1/2^i$; with the remaining probability, i goes to a point in $\mathcal{X}_1$, chosen at random from π. The resulting kernel is strongly π-irreducible, but not π-recurrent—due to the adjoined states.

2. The Proof of Theorem 1

The proof of Theorem 1 is a bit intricate. We begin with some lemmas which are variations on Fubini's theorem and Markov's inequality. Let $(\mathcal{X}_i, \mathcal{B}_i)$ be measurable spaces. Suppose $\mathcal{B}_i$ is countably generated and let φ_i be a probability measure on $(\mathcal{X}_i, \mathcal{B}_i)$. The setting for Lemma 1 is the product space $(\mathcal{X}_1 \times \mathcal{X}_2,\ \mathcal{B}_1 \times \mathcal{B}_2,\ \varphi_1 \times \varphi_2)$; Lemma 2 involves a three-fold product. The leading special case, of course, is Lebesgue measure on the unit square or unit cube. If $A \subset \mathcal{X}_1 \times \mathcal{X}_2$, then $A_{x\bullet}$ is the vertical section of A through x, namely, $\{\, y : y \in \mathcal{X}_2 \text{ and } (x, y) \in A \,\}$. Likewise, $A_{\bullet y}$ is the horizontal section through y. We use ϵ for the generic small positive number, and N for the generic large positive number.

Lemma 1. Let f be a non-negative, measurable function on $\mathcal{X}_1 \times \mathcal{X}_2$, with $\int f \, d\varphi_1 d\varphi_2 \leq \epsilon$. Then $\varphi_1\{\, x : x \in \mathcal{X}_1 \text{ and } \int f(x, y)\varphi_2(dy) \geq \sqrt{\epsilon} \,\} \leq \sqrt{\epsilon}$.

Proof. Let $U(x) = \int f(x, y)\, \varphi_2(dy)$; consider U as a random variable on the probability triple $(\mathcal{X}_1, \mathcal{B}_1, \varphi_1)$. By Fubini's theorem, $E(U) = \int f \, d\varphi_1 d\varphi_2$. This is at most ϵ by the conditions of the lemma, so the chance that $U \geq N\epsilon$ is at most $1/N$ by Markov's inequality. Put $N = 1/\sqrt{\epsilon}$ to complete the proof of Lemma 1.

Corollary 1. If $A \in \mathcal{B}_1 \times \mathcal{B}_2$ and $(\varphi_1 \times \varphi_2)(A) \leq \epsilon$, then

$$\varphi_1\{\, x : x \in \mathcal{X}_1 \text{ and } \varphi_2(A_{x\bullet}) \geq \sqrt{\epsilon} \,\} \leq \sqrt{\epsilon}.$$

Proof. Use Lemma 1, with $f = 1_A$.

Corollary 2. If $B \in \mathcal{B}_1 \times \mathcal{B}_2$ and $(\varphi_1 \times \varphi_2)(B) \geq 1 - \epsilon$, then

$$\varphi_1\{\, x : x \in \mathcal{X}_1 \text{ and } \varphi_2(B_{x\bullet}) > 1 - \sqrt{\epsilon} \,\} \geq 1 - \sqrt{\epsilon}.$$

Proof. Set $A = \mathcal{X}_1 \times \mathcal{X}_2 - B$. Then $(\varphi_1 \times \varphi_2)(A) \leq \epsilon$. By Corollary 1,

$$\varphi_1\{\, x : x \in \mathcal{X}_1 \text{ and } \varphi_2(A_{x\bullet}) \geq \sqrt{\epsilon} \,\} \leq \sqrt{\epsilon}.$$

so that

$$\varphi_1\{\, x : x \in \mathcal{X}_1 \text{ and } \varphi_2(A_{x\bullet}) < \sqrt{\epsilon}\,\} \geq 1 - \sqrt{\epsilon}.$$

But $\varphi_2(A_{x\bullet}) < \sqrt{\epsilon}$ iff $\varphi_2(B_{x\bullet}) > 1 - \sqrt{\epsilon}$, which completes the proof.

Lemma 2. Suppose $C \in \mathcal{B}_1 \times \mathcal{B}_2$ and $D \in \mathcal{B}_2 \times \mathcal{B}_3$. Let $C \otimes D$ be the set of triples (x_1, x_2, x_3) such that $x_i \in \mathcal{X}_i$ and $(x_1, x_2) \in C$ and $(x_2, x_3) \in D$. Suppose

$$(\varphi_1 \times \varphi_2 \times \varphi_3)\{C \otimes D\} \geq 1 - \epsilon.$$

Let $A = \left\{\, x : x \in \mathcal{X}_1 \text{ and } \varphi_2(C_{x\bullet}) > 1 - \sqrt{\epsilon}\,\right\}$ and $B = \left\{\, z : z \in \mathcal{X}_3 \text{ and } \varphi_2(D_{\bullet z}) > 1 - \sqrt{\epsilon}\,\right\}$. Then

$$\varphi_1(A) \geq 1 - \sqrt{\epsilon} \quad \text{and} \quad \varphi_3(B) \geq 1 - \sqrt{\epsilon}.$$

Proof. Since $1_D \leq 1$,

$$\begin{aligned}(\varphi_1 \times \varphi_2 \times \varphi_3)\{C \otimes D\} &= \int 1_C(x_1x_2)1_D(x_2x_3)\,\varphi_1(dx_1)\varphi_2(dx_2)\varphi_3(dx_3)\\ &\leq \int 1_C(x_1x_2)\,\varphi_1(dx_1)\varphi_2(dx_2).\end{aligned}$$

In particular, $(\varphi_1 \times \varphi_2)(C) \geq 1 - \epsilon$, and Corollary 2 completes the proof for A. The argument for B is similar. This completes the proof of Lemma 2.

Our next topic is "C-sets"; we follow Orey (1971) and Harris (1955). Recall that $(\mathcal{X}, \mathcal{B})$ is a measurable space, $\mathcal{B}$ is countably generated, and φ is a probability on $\mathcal{B}$. Let $f(x, y)$ be a non-negative measurable function on $\mathcal{X} \times \mathcal{X}$, with $\int f(x, y)\,\varphi(dy) \leq 1$; this is a "transition sub-density." There is a corresponding "sub-kernel" $P(x, dy) = f(x, y)\,dy$. If f and g are transition sub-densities, so is

$$(f \star g)(x, y) = \int f(x, u)g(u, y)\,\varphi(du).$$

Now $f^{\star n}$ can be defined in the obvious way. Generally, we write fg for $f \star g$ and f^n for $f^{\star n}$: transition sub-densities will seldom be multiplied in this paper.

Even if f is positive everywhere, $f \geq \delta$ may include no positive rectangle—see Example 1. However, $f^2 \geq \delta$ does include positive rectangles; see Proposition 1 in Diaconis and Freedman (1997). Such results can be extended to the singular case; the proof uses some ideas from differentiation theory, which we review here; a reference is Doob (1953). Suppose μ and ν are σ-finite measures on $(\mathcal{X}, \mathcal{B})$. Write $\mu << \nu$ iff $\nu(A) = 0$ implies $\mu(A) = 0$; $\mu \equiv \nu$ iff $\mu << \nu$ and $\nu << \mu$. In the opposite case, $\mu \perp \nu$ when $\mu(A) = 0$ and $\nu(\mathcal{X} - A) = 0$ for some A. The corresponding terminology: "μ is absolutely continuous with respect to ν" if $\mu << \nu$; "μ is equivalent to ν" if $\mu \equiv \nu$; μ is "orthogonal to" or "singular with respect to" ν if $\mu \perp \nu$.

Suppose $\mathcal{B}$ is countably generated. For each n, we can partition $\mathcal{X}$ into a finite collection $\mathcal{C}_n$ of sets, such that

(i) $\mathcal{C}_n \subset \mathcal{B}$;
(ii) If $A \in \mathcal{C}_n$ then A is a union of sets in $\mathcal{C}_{n+1}$, so the partitions are "refining";
(iii) $\mathcal{B}$ is the smallest σ-field which includes $\bigcup_n \mathcal{C}_n$.

For each n, define the function ϕ_n as follows: $\phi_n = \mu(C)/\nu(C)$ on $C \in \mathcal{C}_n$; for now, leave $0/0$ indeterminate. Let $\mathcal{B}_n$ be the (finite) field generated by $\mathcal{C}_n$. These fields are increasing; easy martingale arguments show that $\phi_n \to \phi$ as $n \to \infty$, almost surely with respect to μ and ν.

The space $\mathcal{X}$ now divides into three parts;

(i) $\{\phi = 0\}$, where $\mu = 0$ but $\nu > 0$;
(ii) $\{0 < \phi < \infty\}$, where $\mu \equiv \nu$;
(iii) $\{\phi = \infty\}$, where $\mu > 0$ but $\nu = 0$.

Of course, any of these sets may have measure 0 with respect to both measures. It is always the case that $\mu\{\phi = 0\} = 0$ and $\nu\{\phi = \infty\} = 0$. On sets (i) and (ii), $\mu << \nu$; retracting μ to these two sets gives the part of μ that is absolutely continuous with respect to ν, and $\phi = d\mu/d\nu$. On sets (i) and (iii), $\mu \perp \nu$.

Technical difficulties aside, ϕ_n is the Radon Nikodym derivative—the "density"—of μ with respect to ν on the field $\mathcal{B}_n$; ϕ_n converges to $d\mu/d\nu$ as n gets large. Generally, ν will be a probability, φ being the typical choice; typically, all sets in $\mathcal{C}_n$ will have positive ν-measure. In Lemma 3 below, μ has mass 1 or less: later, μ may have infinite mass. The first result we need from this theory is easy, but not vacuous: the $\mathcal{C}_n$ are specified, while A is generic.

Lemma 3. Let $A \in \mathcal{B}$ with $\varphi(A) > 0$. Given $\epsilon > 0$, there is a positive integer n and a set $C \in \mathcal{C}_n$ such that $\varphi(C) > 0$ and $\varphi(A \mid C) > 1 - \epsilon$.

Proposition 1. Let $P(x, \bullet)$ be a kernel and let φ be an auxiliary probability measure on a separable measurable space $(\mathcal{X}, \mathcal{B})$. Suppose that P is φ-irreducible. Then there is a set $C \in \mathcal{B}$, a positive integer n, and a positive real number δ, such that (i) $\varphi(C) > 0$, and (ii) $P^n(x, D) \geq \delta\varphi(D)$ for all $x \in C$ and all measurable sets $D \subset C$.

Proof. We use i, j, k for generic positive integers. Let $f^k(x, y) = P^k(x, dy)/\varphi(dy)$. As indicated by our review of differentiation theory, this function can be constructed in a jointly measurable way, with values in $[0, \infty]$; for each $x \in \mathcal{X}$, the function will be defined and finite for φ-almost all y, giving the density of the part of $P^k(x, \bullet)$ which is absolutely continuous with respect to φ. More explicitly, fix k for now; let $\{\mathcal{C}_j\}$ be a refining sequence of partitions that generates $\mathcal{B}$. For each j, let $f_j^k(x, y)$ be the derivative of $P^k(x, dy)$ with respect to $\varphi(dy)$, relative to the partition $\mathcal{C}_j$. Finally, let $f^k(x, y) = \lim_{j\to\infty} f_j^k(x, y)$. For each $x \in \mathcal{X}$, this limit exists and is finite for φ-almost all y. In principle, the exceptional null set depends on k and x.

So far, irreducibility has not been used. Now let

$$V_x(A) = \sum_{j=1}^{\infty} P^j(x, A), \tag{1}$$

which is the chance of visiting $A \in \mathcal{B}$ starting from $x \in \mathcal{X}$. Then V_x is a σ-finite measure on $\mathcal{B}$. Irreducibility means that $\varphi << V_x$ for each x; this takes a moment to verify. In particular, $0 < dV_x/d\varphi \leq \infty$ a.e. with respect to V_x, and hence a.e. with respect to φ. Thus, for each x, for φ-almost all y,

$$\sum_{j=1}^{\infty} f^j(x,y) > 0. \tag{2}$$

Fubini's theorem now implies that for φ^3-almost all triples (x,y,z) in $\mathcal{X}^3$,

$$\sum_{j=1}^{\infty}\sum_{k=1}^{\infty} f^j(x,y)f^k(y,z) > 0. \tag{3}$$

Consequently, there exist positive integers j and k with

$$\varphi^3\{\,(x,y,z) : f^j(x,y) > 0 \text{ and } f^k(y,z) > 0\,\} > 0. \tag{4}$$

Hence there is a positive δ' with

$$\varphi^3\{(x,y,z) : f^j(x,y) \geq \delta' \text{ and } f^k(y,z) \geq \delta'\} > 0. \tag{5}$$

We claim there are measurable sets G_1, G_2, G_3 having positive φ-measure, with

$$\varphi^3\{f^j(x,y) \geq \delta' \text{ and } f^k(y,z) \geq \delta' \,\big|\, G_1 \times G_2 \times G_3\} > 1 - \epsilon. \tag{6}$$

To see this, let $\{\mathcal{C}_n\}$ be a refining sequence of partitions that generates $\mathcal{B}$. Each $\mathcal{C}_n$ gives rise to a partition $\mathcal{C}_n^3$ of $\mathcal{X}^3$, comprising sets of the form $C_1 \times C_2 \times C_3$ with all $C_i \in \mathcal{C}_n$. As is easily verified, $\{\mathcal{C}_n^3\}$ is a refining sequence of partitions that generates $\mathcal{B}^3$. Then (6) is immediate from Lemma 3.

We claim there are measurable subsets A, B of $\mathcal{X}$ and a positive real number δ_1 such that

$$\varphi(A) > 0, \quad \varphi(B) > 0, \quad \text{and} \quad f^{j+k}(x,z) \geq \delta_1 > 0 \quad \text{for all } x \in A \text{ and } z \in B. \tag{7}$$

This will follow from Lemma 2, with G_i for $\mathcal{X}_i$ and $\varphi(\bullet \,|\, G_i)$ for φ_i. For C, take the set of all pairs (x,y) with $x \in G_1$, $y \in G_2$, and $f^j(x,y) \geq \delta'$; for D, take the set of all pairs (y,z) with $y \in G_2$, $z \in G_3$, and $f^k(y,z) \geq \delta'$. Define A and B as in Lemma 2. More explicitly, $x \in A$ iff $x \in G_1$ and $\varphi\{C_{x\bullet} \,|\, G_2\} > 1 - \sqrt{\epsilon}$; $z \in B$ iff $z \in G_3$ and $\varphi\{D_{\bullet z} \,|\, G_2\} > 1 - \sqrt{\epsilon}$. According to Lemma 2, both A and B have positive φ-measure: indeed, $\varphi(A) \geq (1-\sqrt{\epsilon})\varphi(G_1)$, and likewise for B. For $x \in A$ and $z \in B$,

$$\varphi(C_{x\bullet} \cap D_{\bullet z} \cap G_2) \geq (1 - 2\sqrt{\epsilon})\varphi(G_2).$$

Moreover, if $x \in A$, $z \in B$, and $y \in C_{x\bullet} \cap D_{\bullet z}$, then $f^j(x,y) \geq \delta'$ and $f^k(y,z) \geq \delta'$, so that

$$f^{j+k}(x,z) \geq \int_{C_{x\bullet}\cap D_{\bullet z}\cap G_2} f^j(x,y)f^k(y,z)\,\varphi(dy) \geq \delta'^2(1-2\sqrt{\epsilon})\varphi(G_2).$$

This completes the proof of (7), with $\delta_1 = \delta'^2(1-2\sqrt{\epsilon})\varphi(G_2)$.

Relationship (7) has the lower bound on a rectangle; the irreducibility condition is used again, to get the bound on a square. In more detail, let $C = C_{i\delta_2}$ be the set of all $x \in B$ with $P^i(x, A) \geq \delta_2$. We claim there is a positive integer i and a positive δ_2 for which

$$\varphi(C) > 0. \tag{8}$$

To verify (8), recall V from (1). By (7), $\varphi(A) > 0$. By irreducibility, $V_x(A) > 0$ for all $x \in B$. Since $\varphi(B) > 0$, relation (8) follows.

Displays (7) and (8) prove the proposition. Indeed, if $x \in C$ and D is a measurable subset of C, then

$$\begin{aligned}
P^{i+j+k}(x, D) &= \int_{\mathcal{X}} P^{j+k}(y, D) P^i(x, dy) \\
&\geq \int_{y \in A} \int_{z \in D} f^{j+k}(y, z)\varphi(dz) P^i(x, dy) \\
&\geq \delta_1 \int_{y \in A} \int_{z \in D} \varphi(dz) P^i(x, dy) \\
&= \delta_1 P^i(x, A)\varphi(D) \\
&\geq \delta_1\delta_2\varphi(D);
\end{aligned}$$

the third line holds because $y \in A$ and $z \in D \subset C \subset B$; the last line is because $x \in C$. This proves the proposition, with $n = i+j+k$ and $\delta = \delta_1\delta_2$. The set C is called a "C-set." Let k be the least n for which there is a positive δ such that $P^n(x, D) \geq \delta\phi(D)$ for all $D \subset C$: then C is said to have "order" k. Compare pp. 7–10 in Orey (1971).

Irreducibility and recurrence are two different ideas when the state space is infinite. For instance, take the random walk with drift on the integers. This process is irreducible—it can get from any integer to any other integer—but transient. Proposition 2 below makes the connection, when there is stationarity. The preliminary Lemma 4 is a well-known result from ergodic theory. To state it, let $(\Omega, \mathcal{F}, \mathcal{P})$ be a probability triple, and let T be a measure-preserving transformation: that is, T is a measurable mapping of Ω into itself, and $\mathcal{P}T^{-1} = \mathcal{P}$. If $F \in \mathcal{F}$, let τ_F be the time of the first visit to F: that is, $\tau_F(\omega)$ is the least $n = 1, 2, \ldots$ if any with $T^n(\omega) \in F$, and $\tau_F(\omega) = \infty$ if none.

Lemma 4. If $\mathcal{P}(F) > 0$ then $\mathcal{P}\{\tau_F < \infty | F\} = 1$.

Proof. Let $F_0 = F - (T^{-1}F \cup T^{-2}F \cup \cdots)$. Starting from $F_0 \subset F$, you never return to F. Suppose by way of contradiction that $\mathcal{P}\{F_0\} > 0$. Now $F_0, T^{-1}F_0, \ldots$ are pairwise disjoint sets with the same positive probability, which is impossible. QED

Corollary 3. Suppose $X_0, X_1, \ldots$ is a stationary stochastic process, and $\mathcal{P}\{X_0 \in F\} > 0$. Let $Y = (Y_0, Y_1, \ldots)$ be an independent copy of X. Given that $X_0 \in F$ and $Y_0 \in F$, there is with conditional probability 1 an $n > 0$ such that $X_n \in F$ and $Y_n \in F$: in other words, X and Y return to F at the same time.

Remark. Suppose $\mathcal{P}(F) > 0$. Let $\mathcal{P}_F = \mathcal{P}\{\bullet \mid F\}$. Define T_F almost surely on F as $T_F = T^{\tau_F}$. Then T_F is measure-preserving relative to $\mathcal{P}_F$.

Proposition 2. Let $P(x, \bullet)$ be a kernel on a separable measurable space $(\mathcal{X}, \mathcal{B})$. Suppose that π is a stationary probability, and $\pi(F) > 0$. Suppose further that

$$\pi\{\, x : V_x(F) > 0 \,\} = 1,$$

with V_x as in (1). Let $\tilde{F}$ be the set of x such that a chain starting from x and moving according to P is almost sure to hit F infinitely often. Then $\pi(\tilde{F}) = 1$.

Proof. Clearly, $\mathcal{X} = \bigcup_{n,m} F_{n,m}$, where $F_{n,m} = \{\, x : P^n(x, F) > 1/m \,\}$. Therefore, it suffices to prove that $P_x\{\text{ hit } F \text{ i.o.}\} = 1$ for π-almost all $x \in F_{n,m}$. In view of Lemma 4, for π-almost all $x \in F_{m,n}$, a chain that starts from x and moves according to P will return to $F_{m,n}$ i.o. a.s. On each return, the chain has chance at least $1/m$ to hit F, and the conditional form of the Borel-Cantelli lemma completes the proof. Of course, some of the $F_{m,n}$ may be empty or π-null; for such $F_{m,n}$, the argument is vacuous.

Proposition 3. Let X and Y be two independent chains, with starting measures α and β, respectively, and transition kernel P. Suppose P is φ-irreducible and C is a C-set of order 1 as in Proposition 1. Suppose that $X_n \in C$ and $Y_n \in C$ for infinitely many n, almost surely. (Both processes are to visit C at the same time.) Then $\|\alpha P^n - \beta P^n\| \to 0$ as $n \to \infty$.

Proof. Let τ_1 be the first time that both chains hit C. We can redefine X and Y so that with probability δ their next move is from φ; hence the redefined processes agree from $\tau_1 + 1$ forwards; with the remaining probability $1 - \delta$ the processes continue to move independently. There will come another time at which both chains are simultaneously in C, and then another bit of merging can be arranged; and so forth. In short, given any $\epsilon > 0$, we can construct chains X_n^ϵ and Y_n^ϵ which start from α and β respectively, such that

$$\text{Prob}\{X_n^\epsilon = Y_n^\epsilon \text{ for all sufficiently large } n\} > 1 - \epsilon.$$

Hence

$$\liminf_{n\to\infty} \text{Prob}\{X_n^\epsilon = Y_n^\epsilon\} > 1 - \epsilon,$$

which completes the proof.

If P is a kernel on $(\mathcal{X}, \mathcal{B})$, let $P \otimes P$ be the obvious kernel on $(\mathcal{X}^2, \mathcal{B}^2)$;

$$[P \otimes P](x, y, A) = [P(x, \bullet) \times P(y, \bullet)](A)$$

for each product measurable A. Even if P is φ-irreducible, then P^2 can be φ-reducible—if P is periodic. On the other hand, if P is strongly φ-irreducible, then $P \otimes P$ should be φ^2-irreducible. We will not need this fact, except for sets of the form $A \times A$, but it does follow from Lemma 5 below, whose straightforward proof is omitted.

Lemma 5. Let P be a kernel on $(\mathcal{X}, \mathcal{B})$; π and φ are probabilities on $(\mathcal{X}, \mathcal{B})$; k is a positive integer.

(a) $(P \otimes P)^k = P^k \otimes P^k$.
(b) If π is stationary under P, then π^2 is stationary under $P \otimes P$.
(c) If P is strongly φ-irreducible, then P^k is strongly φ-irreducible.
(d) If P is strongly φ-irreducible, $A \in \mathcal{B}$, $\varphi(A) > 0$, and $(x, y) \in \mathcal{X}^2$, there is a finite $n = n_{xyA}$ such that $(P \otimes P)^m(A \times A) > 0$ for all $m \geq n$.
(e) If C is a C-set of order k for P relative to φ, then $C \times C$ is a C-set of order k for $P \otimes P$ relative to φ^2.

We are now ready to prove Theorem 1. Let C be a C-set, which exists by Proposition 1. Let k be the order of C, and let $Q = P^k \otimes P^k$. According to Lemma 5, $C \times C$ is a C-set of order 1 for Q, and π^2 is stationary for Q. Let $D \subset \mathcal{X}^2$ be the set of pairs x, y such that, starting from x, y and moving according to Q, a chain visits $C \times C$ i.o. a.s. We claim

$$\pi^2(D) = 1. \tag{9}$$

Indeed, $Q^n(x, y, C \times C) > 0$ for all sufficiently large n, by Lemma 5; Proposition 2 completes the proof of (9). Proposition 3 now shows that

$$\|P^{nk}(x, \bullet) - P^{nk}(y, \bullet)\| \to 0 \tag{10}$$

as $n \to \infty$ for all $(x, y) \in D$.

Let D_0 be the set of $x \in \mathcal{X}$ such that the x-section of D has π-measure 1. By Fubini's theorem, $D_0 \in \mathcal{B}$ and $\pi(D_1) = 1$. If $x \in D_0$, then (10) holds for π-almost all y: hence $\|P^{nk}(x, \bullet) - \pi\| \to 0$ as $n \to \infty$ for all $x \in D_0$. Since π is invariant, $\pi P(D_0) = \pi(D_0) = 1$. So $D_1 = \{\, x : x \in D_0 \text{ and } P(x, D_0) = 1 \,\}$ is measurable and has π-measure 1. For $x \in D_1$, $\|P^{nk+1}(x, \bullet) - \pi\| \to 0$, because

$$P^{nk+1}(x, \bullet) - \pi = \int_{y \in D_1} [P^{nk}(y, \bullet) - \pi]\, P(x, dy),$$

so

$$\|P^{nk+1}(x, \bullet) - \pi\| \leq \int_{y \in D_1} \|P^{nk}(y, \bullet) - \pi\|\, P(x, dy).$$

Proceeding in this way, we construct a decreasing sequence of measurable sets D_j with $\pi(D_j) = 1$ and $\|P^{nk+j}(x, \bullet) - \pi\| \to 0$ as $n \to \infty$ for all $x \in D_j$.

Let $\mathcal{X}_1 = \bigcap_j D_j$. Clearly, $\pi(\mathcal{X}_1) = 1$. If $x \in \mathcal{X}_1$, it is easily seen that $P(x, D_j) = 1$ for all j, so $P(x, \mathcal{X}_1) = 1$. Furthermore, $\|P^n(x, \bullet) - \pi\| \to 0$ as $n \to \infty$. Finally, the process starting from x visits C i.o., almost surely: that is so for all $x \in D_0$, and $\mathcal{X}_1 \subset D_0$. Only recurrence remains to be proved. To do that, fix a measurable set $A \subset \mathcal{X}_1$ with $\pi(A) > 0$, and a point $x \in \mathcal{X}_1$. Let X_n be a Markov process starting from $X_0 = x$ and moving according to P. Let $\breve{X}_m$ be this process, but only at the times when it is in C. The law of $\breve{X}_m$ converges in variation distance to $\pi(\bullet \mid C)$, for instance by Theorem 2.

By the irreducibility assumption, $C = \bigcup_{i,j} C_{i,j}$, where

$$C_{i,j} = \{\, y : y \in C \text{ and } P^i(y, A) > 1/j \,\}.$$

Therefore, we can find a $\delta > 0$ and a finite disjoint sequence of sets $C_1, C_2, \ldots \subset C$ such that (i) $\pi(\cup_k C_k \,|\, C) > 1 - \delta$, and (ii) for each k there is an $i = i_k$ with $P^i(y, A) > \delta$ for all $y \in C_k$. Now $\lim_m \text{Prob}\{\, \breve{X}_m \in \cup_k C_k \,\} > 1 - \delta$, so $\text{Prob}\{\, \breve{X}_m \in \cup_k C_k \text{ i.o.}\} > 1 - \delta$. By the conditional form of the Borel-Cantelli lemma, $X_n \in A$ i.o., with probability greater than $1 - \delta$. Letting $\delta \to 0$ completes the proof. There is (at least) one irritating complication: we have have not verified that C is a subset of $\mathcal{X}_1$. On the other hand, starting from $x \in \mathcal{X}_1$, the process stays in $\mathcal{X}_1$ almost surely: thus, X_n and $\breve{X}_m$ can be taken as staying in $\mathcal{X}_1$, if desired.

6. Literature Review

The most accessible references on the Doeblin theory are perhaps Orey (1973) and Harris (1956); the latter demonstrates the existence of a σ-finite invariant measure μ for a φ-recurrent kernel, with $\mu >> \varphi$. Other references are Asmussen (1987), Meyn and Tweedie (1993), and Revuz (1984); Lindvall (1992) discusses the coupling method. Doeblin (1940) should be mentioned; Cohn (1993) gives an overview of the history. Eaton (1992) derives convergence theorems from reversibility; Lamperti (1960) gives martingale recurrence conditions. Pitman (1974) has quite sharp results for chains with discrete state space. Athreya, Doss and Sethuraman (1996) prove convergence to stationarity in the presence of a stationary measure; they assume the existence of a C-set in condition (1.5)—relative to a nonstationary auxiliary measure. Their Theorem 5 is like our Theorem 1; they attribute the result to Tierney (1994); Tierney's proof depends on work of Nummelin (1984) and Revuz (1984), so the arrangement of ideas here may be a bit easier to follow. The exceptional null set in Theorem 1 can be eliminated by assuming, for instance, recurrence relative to the stationary probability. Given the existence of a C-set, a regenerative event can be constructed, and the renewal theorem can be used to prove convergence; see, for instance, Athreya and Ney (1978). Their Definition (2.2) makes A a C-set, and their Theorem (4.1) is a bit stronger than Theorem 1 here. The argument is developed further in Athreya, McDonald and Ney (1978a). Athreya, McDonald and Ney (1978b) prove the renewal theorem by a coupling argument like Doeblin's. Also see Nummelin (1994).

References

Asmussen, S. (1987). *Applied Probability and Queues.* Wiley, New York.

Athreya, K.B., Doss, H., and Sethuraman, H. (1996). On the convergence of the Markov chain simulation method. *Annals of Statistics* vol. 24 pp.69–100.

Athreya, K.B., McDonald, D., and Ney, P. (1978a). Limit theorems for semi-Markov processes and renewal theory for Markov chains. *Ann. Probab.* vol. 6 pp.788–797.

Athreya, K.B., McDonald, D., and Ney, P. (1978b). Coupling and the renewal theorem. *Amer. Math. Monthly* vol. 85 pp.809–814.

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

Cohn, H., ed. (1993). *Doeblin and Modern Probability: Proceedings of the Doeblin Conference.* American Mathematical Society, vol. 149.

Diaconis, P. and Freedman, D. (1997). On the hit and run process. Technical report no. 497, Department of Statistics, U.C. Berkeley.

Doeblin, W. (1940). Eléments d'une théorie générale des chaînes simple constantes de Markoff. *Ann. Sci. Ecole Norm. Sup.* 57 61–111.

Doob, J.L. (1953). *Stochastic Processes.* Wiley, New York.

Lindvall, T. (1992). *Lectures on the coupling method.* New York: Wiley.

Meyn, S.P. and Tweedie, R.L. (1993). *Markov Chains and Stochastic Stability.* Springer, London.

Nummelin, E. (1994). On distributionally regenerative processes. *Journal of Theoretical Probability* 7 739–756.

Nummelin, E. (1984). *General Irreducible Markov Chains and Non-Negative Operators.* Cambridge University Press.

Orey, S. (1971). *Lecture notes on limit theorems for Markov chain transition probabilities.* Mathematical Studies, 34. Van Nostrand Reinhold Co., New York.

Pitman, J. (1974). Uniform rates of convergence for Markov chain transition probabilities. *Z. Wahrscheinlichkeitstheorie* 29 193–227.

Revuz, D. (1984). *Markov Chains.* North-Holland, Amsterdam.

Tierney, L. (1994). Markov chains for exploring posterior distributions (with discussion). *Ann. Statist.* vol. 22 pp.1701–1762.

Technical Report no. 501
Department of Statistics
University of California
Berkeley, CA 94270